J Mol Evol (1999) 49:108–121

JOURNAL OF MOLECULAR EVOLUTION

A Probable Mixed-Function Supraoperon in *Pseudomonas* Exhibits Gene Organization Features of Both Intergenomic Conservation and Gene Shuffling

Gary Xie,* Carol A. Bonner, Roy A. Jensen

Department of Microbiology and Cell Science, Building 981, University of Florida, Gainesville, FL 32611-0700, USA

Received: 15 September 1998 / Accepted: 18 February 1999

Abstract. Sequencing of an 8182-bp chromosomal region in *Pseudomonas stutzeri* revealed the major portion of an apparent mixed-function supraoperon (defined as a nested organization of transcriptional units encoding gene products which function in more than one biochemical pathway). A nearly identical supraoperon organization was apparent in the unpublished *Pseudomonas aeruginosa* genome database, where the complete *Pseudomonas* supraoperon was deduced. The *ser*C-(*pdx*F)–*aro*Q_p · *phe*A–*his*H_b–*tyr*A_c–*aro*F–*cmk*–*rps*A supraoperon encodes 3-phosphoserine aminotransferase, a bidomain chorismate mutase/prephenate dehydratase, imidazole acetol-phosphate aminotransferase, cyclohexadienyl dehydrogenase, 5-enolpyruvylshikimate 3-phosphate synthase, cytidylate kinase, and ribosomal protein S1. The member genes were identified by homology analysis, enzyme assay, and/or functional complementation. Although SerC(PdxF) and $His H_b$ exercise their primary functions in serine, pyridoxine, and histidine biosynthesis, they also have critical catalytic roles in provision of the sidechain amino groups of tryptophan, phenylalanine, and tyrosine. The likelihood of supraoperon-wide translational coupling is suggested by the highly compressed intergenic spacing (including overlapping stop and start codons), as well as by possible hairpin structures in mRNA which may sequester some of the ribosome-binding sites and thus provide a mechanism for translational coupling. A comparison of the organization of the supraoperon genes in other organisms represented in the database revealed unmistakable conservation of the linkage of these genes across wide phylogenetic boundaries, albeit with considerable gene shuffling. At least remnants and shuffled portions of the entire supraoperon are distributed throughout the Gram-negative bacteria with the *his*H_b–*tyr*A–*aro*F gene block being conserved as distantly as the gram-positive bacteria. Such conservation of mixed-function genes may reflect the selective value of still-unknown global relationships of protein–protein interaction or regulation.

Key words: Genomics — Gene organization — Supraoperon — *Pseudomonas*

Introduction

The evolutionary significance of gene organization within genomes and between genomes has been impossible to evaluate in any substantial way for lack of sufficient data. These modern times of whole-genome sequencing now provide an unprecedented opportunity to evaluate gene organization. To what extent does a given cluster of genes remain linked together across what phylogenetic distance? What relationships exist between genes that might account for bringing them together and keeping them together?

The rapidly emerging information in the database already yields the perhaps surprising result that gene order

**Present address:* Los Alamos National Laboratories, Los Alamos, NM 87545, USA
Correspondence to: R.A. Jensen; *e-mail:* rjensen@micro.ifas.ufl.edu

is not generally conserved (Mushegian and Koonin 1996). This may not be so unexpected for genes whose neighbors do not share obvious functional relationships, but even gene members of multigene operons in organisms as close as *Escherichia coli* and *Haemophilus influenzae* have sometimes become dispersed (Watanabe et al. 1997). Of the short conserved gene strings that are shared by *E. coli* and *H. influenzae*, about 50% include genes not known to have an operon organization in *E. coli* (Tatusov et al. 1996). The few instances of strong conservation usually involve physically associated gene products (Mushegian and Koonin 1996), and a clear correlation with RNA-level regulation has also been observed (Siefert et al. 1997).

Against this background we have found that gene members of a mixed-function supraoperon in *Pseudomonas stutzeri* and *P. aeruginosa* exhibit a highly conserved pattern of linkage among the gram-negative bacteria that extends in part as far away as the gram-positive bacteria, albeit with considerable gene shuffling. Supraoperons are defined as nested transcriptional units encoding gene products for a single metabolic pathway. Mixed-function supraoperons are nested transcriptional units encoding gene products for more than one biochemical pathway.

As one well-documented example of a mixed-function supraoperon, *Bacillus subtilis* has positioned a classic *trp* operon inside a previously unrecognized larger transcriptional unit (Henner and Yanofsky 1993). The upstream *B. subtilis aro*GBH operon overlaps the *trp* promoter, the *aro*H stop codon is within the *aro*GBH terminator, and the *aro*GBH terminator is a synonymous with the *trp* attenuator. A third operon (*his*H_b–*tyr*A_p–*aro*F) downstream is nested within the *trp* operon. Readthrough *trp* transcripts can proceed through the *his*H_b–*tyr*A_p–*aro*F operon, or transcripts for *his*H_b–*tyr*A_p–*aro*F can be initiated within the distal *trp*-operon gene. This exemplies a complex system of regulation whereby a given gene may obey multiple transcriptional commands that reflect its membership in a hierarchical array of nested control units.

Some well-characterized examples of mixed-function supraoperons in *E. coli* can also be cited. (i) The *ami*B–*mut*L–*mia*A–*hfq*–*hfl*X–*hfl*K–*hfl*C system governs functions which include cell-wall hydrolysis (*ami*B), DNA repair (*mut*L), tRNA modification (*mia*A), and proteolysis (*hlf*X–*hlf*K–*hfl*C). The supraoperon includes two additional genes of unknown function in the upstream region (Tsui et al. 1994a, b). The overall mechanisms of transcriptional control have been shown to be extremely complicated. These include the use of multiple internal promoters, *rho*-dependent and *rho*-independent intraoperon attenuation, and RNA processing. (ii) The *aro*K–*aro*B–*urf*–*dam*–*rpe*–*gph*–*trp*S system governs functions which include shikimate kinase (*aro*K), dehydroquinate synthase (*aro*B), DNA adenine methyltransferase (*dam*), D-ribulose-5-P epimerase (*rpe*), 2-phosphoglycolate phosphatase (*gph*), and tryptophanyl-tRNA synthetase (*trp*S) (Lyngstadass et al. 1995). (iii) The *ser*C(*pdx*F)–*aro*F system links the 3-phosphoserine aminotransferase step of serine biosynthesis and a transamination step of pyridoxine biosynthesis with the 5-enolpyruvylshikimate 3-phosphate (EPSP) synthase step in the common early pathway of aromatic amino acid biosynthesis (Duncan and Coggins 1986; Lam and Winkler 1990; Man et al. 1997).

The *Pseudomonas* supraoperon described in this paper exhibits a distinct relationship with two of the foregoing well-studied supraoperon systems, namely, with the 12-gene *B. subtilis* supraoperon and with the 2-gene *E. coli ser*C(*pdx*F)–*aro*F supraoperon. The latter *E. coli* supraoperon is immediately adjacent to an operon (*yca*L–*cmk*–*rps*A) which may in fact belong to the *ser*C–*aro*F supraoperon, judging from the comparative gene organization seen in *Pseudomonas* and other gram-negative bacteria.

Materials and Methods

Gene and Gene-Product Acronyms. The contemporary erratic naming of genes in different organisms is an increasingly awkward problem when extensive multiorganism comparisons are attempted. A universal nomenclature is inevitable. Most of the acronyms used in this paper have already been established by Gu et al. (1997) and Subramaniam et al. (1998). Thus, genes encoding enzymes in the common-pathway portion of aromatic biosynthesis are named as a logical progression in order of reaction sequence (i.e., *aro*A → *aro*G). Genes that specify fusion proteins having multiple catalytic domains corresponding to single-gene counterparts elsewhere are named with the use of bullets to separate individual domain names as suggested by Crawford (1989). Thus, $AroQ_p$· specifies the N-terminal catalytic domain of chorismate mutase located on the bidomain $AroQ_p$·PheA, and ·PheA specifies the carboxy-terminal prephenate dehydratase domain. Broad-specificity and narrow-specificity classes of imidazole acetol-phosphate (IAP) aminotransferase (Jensen and Gu 1996) are distinguished by subscript denotations. Thus, $HisH_b$ proteins can utilize aromatic amino acids in addition to IAP, in contrast to $HisH_n$ enzymes, which are restricted to IAP utilization.

Bacterial Strains, Media, and Growth Conditions. The bacterial strains and plasmids utilized or constructed in this investigation are shown in Table 1. *P. stutzeri* was cultured as described by Carlson et al. (1983). *E. coli* strains were grown in M9 minimal medium (Miller 1972) or Luria–Bertani (LB) medium (Davis et al. 1980a) at 37°C. Amino acids (50 μg/ml) and thiamine hydrochloride (17 μg/ml) were added as growth supplements when required. Antibiotics, when appropriate to provide selective pressure for maintenance of plasmids, were added to the medium at standard concentrations (Maniatis et al. 1989). Agar was added at 20 g/L for preparation of solid medium. X-Gal (5-bromo-4-chloro-3-indoxyl β-D-galactopyranoside) was included at concentrations recommended by the supplier (Promega) when recombinants were screened in pUC18 or pUC19 plasmids that were transformed into *E. coli* DH5α. Growth rates were determined by measuring the turbidity of exponentially growing cultures at 600 nm in a Perkin–Elmer Model 35 spectrophotometer. T4 DNA ligase, DNA-modifying enzymes (New England Biolabs or Promega) were used as recommended by the suppliers. Antibiotics, thiamine, amino acids, and other biochemicals were obtained from Sigma Chemical Company. LB me-

Table 1. Bacterial strains and plasmids

Strain or plasmid	Genotype or description	Source or reference
E. coli		
BL21 (DE3)	F⁻ *ompT hsd*S_B ($r_B^-m_B^-$) *gal* dcm; with DE3, a λ prophage carrying the T_7 RNA polymerase gene	Novagen
DH5α	F⁻ ϕ80d*lac*ZΔM15 Δ(*lac*ZYA–*arg*F) U169 *deo*R *rec*A1 *end*A1 *sup*44gλ⁻ *hsd*R17 (r_K^-, m_K^+) *thi*−1 *gyr*A96 *rel*A1	GIBCO/BRL
AT2471	*thi*−1 *tyr*A4 *rel*A1λ⁻ *spo*T	CGSC[a] 4510
UTH780	*his*C780 *mal*A1 (λ^r) *xyl*−5 *rps*L145λ⁻	CGSC 5954
NU1113	VJS433 *ser*C(*pdx*F)::<mini-Mu dI195	Lam and Winkler (1990)
JP2255	*aro*F363 *phe*A361 *phe*0352 *tyr*A382 *thi*−1 *str*R712 *lac*Y1 *xy1*−15	Baldin and Davidson (1981)
DL39	λ⁻ *asp*C13 *fnr*−25 *ilv*E12 *tyr*B507	LeMaster and Richards (1987)
P. stutzeri		
JM300	Prototroph	Carlson et al. (1983)
Cosmids		
pHC79	Ap^r Tc^r cos	Hohn and Collins (1980)
pJF958	Phe$^+$ clone from pHC79:*Pst*I gene bank of JM300	Fischer et al. (1991)
pJF9586	Phe$^+$ 4.35-kb *Pst*I fragment subcloned from pJF958 into pHC79	Fischer et al. (1991)
Plasmids		
pUC18	Ap^r, *lac*I′PO′	Yanisch-Perron et al. (1985)
pET24b(+)	T_7*lac* promoter, *lac*I$^+$ Km^r His · Tag sequence, T_7 · Tag (11aa) sequence	Novagen
pETtyrA_c	pET24b (+) carrying *tyr*A_C translation fusion at ATG start site and C-terminal His · Tag sequence	This study
pJF1956	788-bp *Sph*I–*Eco*RI fragment subcloned from pJF1954 into pGEM-4Z	Fischer et al. (1991)
pJF1972	775-bp *Sph*I–*Pst*I fragment subcloned from pJF9586 into pGEM-5Zf (+)	Fischer et al. (1991)
pJX7200	7200-bp *Pst*I fragment subcloned from pJF958 into pUC18	This study
pJX4600	4800-bp *Sma*I–*Eco*RI fragment subcloned from pJF958 into pUC18	This study
pJX0500	530-bp *Pst*I fragment subcloned from pJF958 into pUC18	This study
pJX0001	1650-bp *Sph*I fragment subcloned from pJX4600 into pUC18	This study
pJX0300	161-bp *Pst*I fragment subcloned from pJF958 into pUC18	This study
pJX0002	1040-bp *Sph*I fragment subcloned from pJF958 into pUC18	This study
pJX0400	360-bp *Pst*I fragment subcloned from pJF958 into pUC18	This study
pJX1000	1060-bp *Eco*RI fragment subcloned from pJF9586 into pUC18	This study
pJX1670	1670-bp *Eco*RI–*Pst*I fragment subcloned from pJF9586 into pUC18	This study
pJX0003	1320-bp *Ban*II–*Sph*I fragment subcloned from pJF9586 into pUC18	This study
pJX1021	710-bp *Eco*RI–*Ban*II fragment subcloned from pJX1000 into pUC18	This study
pJX0420	450-bp *Ban*II–*Eco*RI fragment subcloned from pJX1320 into pUC18	This study
pJX0800	870-bp *Eco*RI–*Sph*I fragment subcloned from pJX1320 into pUC18	This study
pJX0500	500-bp *Sph*I–*Sma*I fragment subcloned from pJX1670 into pUC18	This study

[a] *Escherichia coli* Genetic Stock Center, Yale University.

dium and agar were purchased from Difco. Inorganic chemicals (analytical grade) were from Fisher Scientific.

Enzyme Assays. Extract preparation and assays for 3-phosphoserine aminotransferase (Duncan and Coggins 1986), chorismate mutase (Fischer et al. 1991), prephenate dehydratase (Fischer et al. 1991), imidazole acetol phosphate/aromatic aminotransferase (Gu et al. 1995), and cyclohexadienyl dehydrogenase (Zhao et al. 1993) were carried out as referenced. Protein determinations were made according to the method of Bradford (1976).

DNA Manipulation and Genetic Procedures. Standard molecular biology procedures were performed as described by Maniatis et al. (1989), unless otherwise indicated. Bacterial chromosomal DNA was purified by the method of Yuan and Lin (1982). DNA for small-scale plasmid preparations was prepared according to the protocol of Davis et al. (1980b). *E. coli* strains were transformed by the use of a $CaCl_2$ method (Dagert and Ehrlich 1979). Restriction enzymes, ligase, and calf intestine alkaline phosphatase were purchased from New England Biolabs or Promega and were used according to manufacturer instructions. DNA fragments were purified from agarose gel with a Geneclean kit (Bio101). Subcloning was conducted by standard methods (Maniatis et al. 1989).

DNA Sequencing and Data Analysis. Fischer et al. (1991) previously identified *aro*Q_p·*phe*A (encoding chorismate mutase and prephenate dehydratase) in *P. stutzeri.* To determine the DNA sequence upstream and downstream of *aro*Q_p·*phe*A, subclones were purified by the method recommended in User Bulletin 18 offered by Applied Biosystems, Inc. Double-stranded plasmid DNA was sequenced in both directions by the DNA sequencing facility at the Department of Microbiology and Cell Science, University of Florida. The nucleotide and deduced amino acid sequences were analyzed using the updated version (Version 8.0, 1994) of the sequence analysis software package offered by the Genetics Computer Group, Inc. (GCG) (Devereux et al. 1984) and using the WWW BLAST 2.0 (www.ncbi.nlm.nih.gov) program offered by the National Center for Biotechnology Information (Altschul et al. 1997).

Analysis of Raw DNA Sequence Data. Raw DNA sequence available from the *P. aeruginosa* database (www.pseudomonas.com) was screened using the built-in BLAST service. DNA sequences from GenBank and protein sequences from Swiss-Prot and PIR were used as

query entries. WS_FTP was used to fetch raw sequences. The WWW BLAST 2.0 (Altschul et al. 1997) and WWW ORF Finder offered by the National Center for Biotechnology Information were used to locate open reading frames and to confirm the similarity search result of the raw sequence. The Codon Preference program in the GCG package was used to locate and correct likely frameshift errors that were implicated by observation of suspicious sequence blocks in multiple alignments (PILEUP). The *P. aeruginosa* codon usage table was retrieved from ftp.ebi.ac.uk. The Seqedt program was used to correct sequence errors.

Analysis of Probable Ribosome Binding Sites. The 13-nucleotide (nt) sequence of *P. aeruginosa* 16S rRNA at the 3′ end (3′AUUCCUCCACUAG-5′) was compared with the region upstream of any given translation start site for the best base-pair (bp) match, assuming an ideal spacing of 8 nt between C1530 of 16S tRNA and the A nt of the ATG start codon.

Nucleotide Sequence Accession Number. The nucleotide sequence reported in this paper has been assigned GenBank accession number AF038578.

Results

DNA Sequence Analysis. Results obtained by Fischer et al. (1991), who characterized AroQ$_p$·PheA in detail at the enzymological level, revealed the intriguing observation that in *P. stutzeri* a truncated but probable *ser*C homologue was located upstream of *aro*Q$_p$·*phe*A in a position whereby the *ser*C stop codon and the *aro*Q$_p$·*phe*A start codon overlapped (**TGA**TG). We were interested in examining the entire flanking regions upstream and downstream of *aro*Q$_p$·*phe*A to determine any additional genes which might be functionally related. While this project was under way, the genomic sequencing of the closely related *Pseudomonas aeruginosa* was nearing completion. This created an opportunity to extend the analysis to *P. aeruginosa*, an organism for which the pathway of aromatic amino acid biosynthesis has been extensively studied at the enzymological level.

Approximately 1.5 kb of DNA sequence at one end (left in Fig. 1) and 8.0 kb at the other end (right in Fig. 1) of the *P. stutzeri* insert in pJF958 was determined. Overlapping DNA fragments were sequenced in both directions (strategy not shown). Computer analysis of this sequence revealed five open reading frames (ORFs) that flanked the known *aro*Q$_p$·*phe*A gene. These all showed strong conformity with the *P. stutzeri* codon usage table. The predicted amino acid sequences from each ORF were compared to the translated (all six reading frames) DNA sequences in GenBank. This analysis showed that the six ORFs had significant homology with the predicted amino acid sequences of *gyr*A, *ser*C, *aro*Q$_p$·*phe*A, *his*H$_b$, *tyr*A$_c$, and *aro*F genes from other organisms. Upstream at the other end of the insert, the sequence analysis revealed homologues of established *P. aeruginosa* genes: *pfe*S (activation of ferric enterobactin receptor) and the C-terminal portion of *pfe*A (ferric enterobactin receptor). The G+C content of these ORFs falls into the range of 60.6–66.5% for the *P. stutzeri* genome (Palleroni 1984).

The P. stutzeri gyr*A Gene.* An ORF of 2814 nt encoding *gyr*A is located upstream of *ser*C. The calculated mass of the *gyr*A polypeptide of 938 amino acids was 103,113 Da, with a p*I* of 4.74. DNA gyrase, a type-II DNA topoisomerase, is essential for bacterial viability. It catalyzes ATP-dependent negative supercoiling of DNA and is involved in DNA replication and transcription (Reece and Maxwell 1991). The enzyme exhibits an $\alpha_2\beta_2$ construction, the subunits being encoded by *gyr*A and *gyr*B, respectively. The A subunit is responsible for the double-stranded breakage and reunion of DNA, while the B subunit mediates energy transduction via ATP hydrolysis.

The derived amino acid sequence of the *P. stutzeri* JM300 *gyr*A protein shares an overall amino acid identity of 85% with the *P. aeruginosa Gyr*A protein. The amino acids around the active-site tyrosine (amino acid 122 in *E. coli*) are particularly well conserved (data not shown). Between nt −140 and nt −200 is a hairpin loop which may be an indication of supercoiling-dependent regulation of *gyr*A transcription, as has been shown for other promoters (Horwitz and Loeb 1988). Located 268 bases downstream of the UGA stop codon is a stem-loop structure which is a likely factor-independent transcription termination signal.

The P. stutzeri ser*C (*pdx*F) Gene.* Upstream of *P. stutzeri aro*Q$_p$·*phe*A the deduced amino acid sequence indicated significant homology to an enzyme involved in the biosynthesis of serine, termed 3-phosphoserine aminotransferase. This enzyme catalyzes the transamination of 3-phosphohydroxypyruvate to yield 3-phosphoserine. The original *aro*Q$_p$·*phe*A clone (pJF9586) isolated contained an intact *ser*C gene, since a 4.8-kb *Eco*RI–*Sma*I fragment cloned into the *Eco*RI–*Sma*I sites (within the multiple cloning region) of pUC18 (pJX4600) complemented the *ser*C mutation in *E. coli* NU1113. Plasmids purified from the ampicillin- and kanamycin-resistant transformants had the same size and banding pattern (supercoiled, circular, and linear) as pJX4600 on agarose gel, and a second round of transformation of *E. coli* NU1113 with these purified plasmids again conferred serine prototrophy. A transformant was purified by three rounds of single-colony isolation and used to inoculate cultures for crude extract preparation. In contrast to the control extract which lacked detectable activity, a specific activity of 0.97 μmol min^{-1} mg^{-1} was obtained for 3-phosphoserine aminotransferase.

It was of interest to determine whether *P. stutzeri ser*C was synonymous with *pdx*F, as in *E. coli* (Lam and Winkler 1990). If so, *E. coli* NU1113 transformants which were independent of the serine requirement should also be free of the pyridoxine requirement. *P. stutzeri*

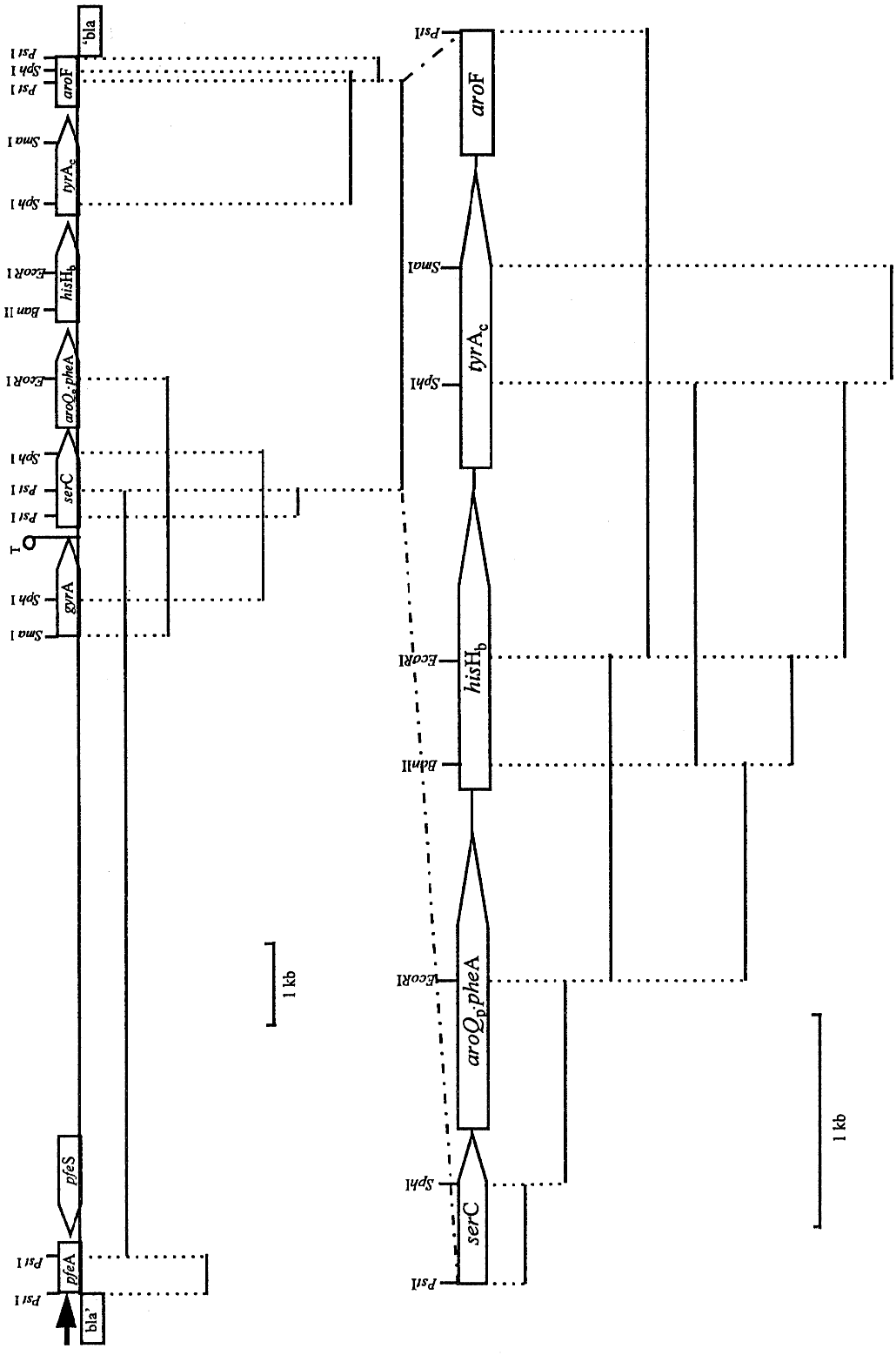

Fig. 1. Physical localization of the *P. stutzeri pfeA–pfeS/gyrA–serC–aroQ*$_p$*.pheA–hisH*$_b$*–tyrA*$_c$*–aroF* region. Linear maps of the clone and subclones are shown. Restriction sites are labeled in the cloned *P. stutzeri* DNA. The orientations of pJF958 relative to the *bla* gene of pHC79 are depicted at the *top.*

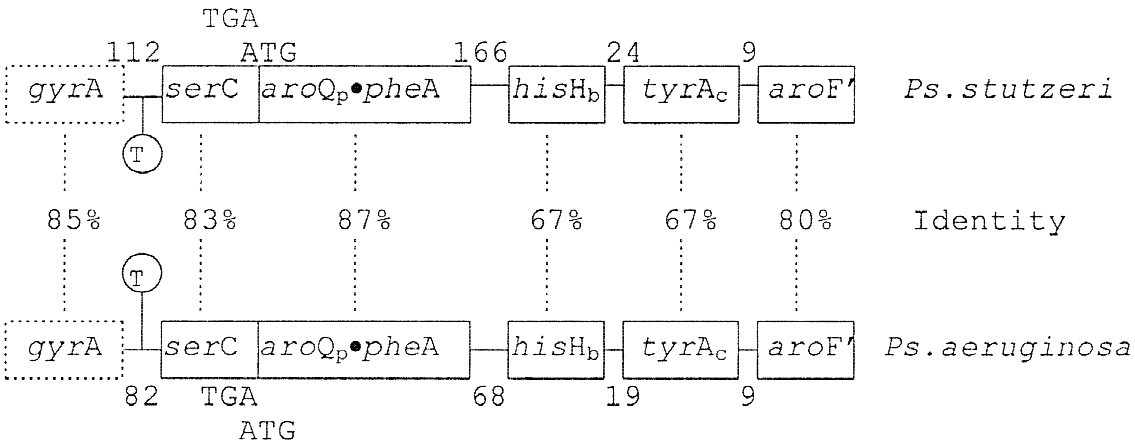

Fig. 2. Schematic diagram of the identical gene arrangements present in *P. aeruginosa* and *P. stutzeri.* The coding regions (*open boxes*) for the *gyr*A–*ser*C–*aro*Q_p-*phe*A–*his*H_b–*tyr*A_c–*aro*F genes are labeled. The percentages of amino acid identities between homologous proteins are given. The intergenic distances are not shown proportionally, but base-pair numbers are indicated. The likely terminator structure, designated T and *circled,* between *gyr*A and *ser*C is also indicated.

*ser*C was indeed able to complement the pyridoxine requirement of the *E. coli ser*C (*pdx*F) auxotroph. (It was noted by accident that NU1113 transformants excreted indole, apparently due to anomalous induction of tryptophanase induced by the exogenous L-tryptophan provided. The parent strain grows more slowly than normal, and the probable consequence of elevated cAMP levels may prevent normal catabolite repression of tryptophanase.) It remains possible that a separate gene encodes PdxF in *P. stutzeri* and that the complementation results could be explained by a combination of a high copy number and a low substrate specificity of SerC for the PdxF-reaction substrates. If so, the putative PdxF must not be a homologue of SerC judging from the absence of SerC homologues in the *P. aeruginosa* database. At a minimum, the complementation results demonstrate the evolutionary potential of SerC to catalyze the PdxF reaction.

The restriction sites shown in Fig. 1 were utilized to construct a set of subclones from pJF9586 in pUC18 for nucleotide sequencing. The *ser*C gene, 1098 bp in length, begins at codon ATG and stops at codon TGA. *P. stutzeri* SerC shows 57% identity to *Yersinia enterocolitica* SerC and also significant homology to the SerC proteins from other organisms, including those from eubacteria, fungi, and mammals. A strong ribosome-binding site was apparent. *P. stutzeri ser*C encodes a deduced protein having a p*I* of 4.74 and a molecular mass of 40,238 Da.

The P. stutzeri hisH_b *Gene.* Fischer et al. (1991) identified a partial ORF of 293 bp located directly downstream of *phe*A in *P. stutzeri.* Amino acid sequence comparison with other sequences in the database indicated that the downstream gene might encode imidazole acetol-phosphate (IAP) aminotransferase. The original *P. stutzeri* clone (pJF9586) isolated (Fischer et al. 1991) was indeed able to complement *E. coli* UTH780, an IAP aminotransferase mutant. Cosmids purified from the tetracycline-resistant transformants had the same size and banding pattern (supercoiled, circular, and linear) as pJF9586 on agarose gel, and a second transformation of *E. coli* UTH780 with these purified cosmids again conferred histidine prototrophy.

The deduced amino acid sequence yields a protein of 370 residues with a molecular mass of 39,699 Da and a p*I* of 6.53. *P. stutzeri* $HisH_b$ exhibits the highest amino acid identity (50.8%) with $HisH_b$ from the closely related *P. aeruginosa,* and these genes cluster with the broad-specificity class of IAP aminotransferase ($HisH_b$) within the subfamily-Iβ division of the family- I aminotransferases defined by Jensen and Gu (1996). The most divergent of the currently recognized members of family Iβ exhibit amino acid identities of about 25%.

The expectation that *P. stutzeri his*H_b encodes a broad-specificity aminotransferase capable of aromatic aminotransferase activity in addition to its histidine-pathway activity was tested by functional complementation. *E. coli* DL39 carries deficient genes for *ilv*E, *tyr*B, and *asp*C. As a result, it requires isoleucine, valine, leucine, tyrosine, phenylalanine, and aspartate for growth. This combination of mutations is required in *E. coli* for a clean background of aromatic auxotrophy owing to the broad-substrate specificities of *E. coli asp*C and *ilv*E (Gelfand and Steinberg 1977). *P. stutzeri his*H_b was indeed able to complement the requirement of *E. coli* DL39 for L-phenylalanine and L-tyrosine (but not for L-aspartate or for branched-chain amino acids).

Two independent transformant clones were purified and used for enzymatic verification of the expected *in vitro* properties of the $HisH_b$ class of aminotransferase. Transformants grown under the foregoing selective conditions were used for preparation of extracts which were subjected to DEAE-cellulose chromatography as de-

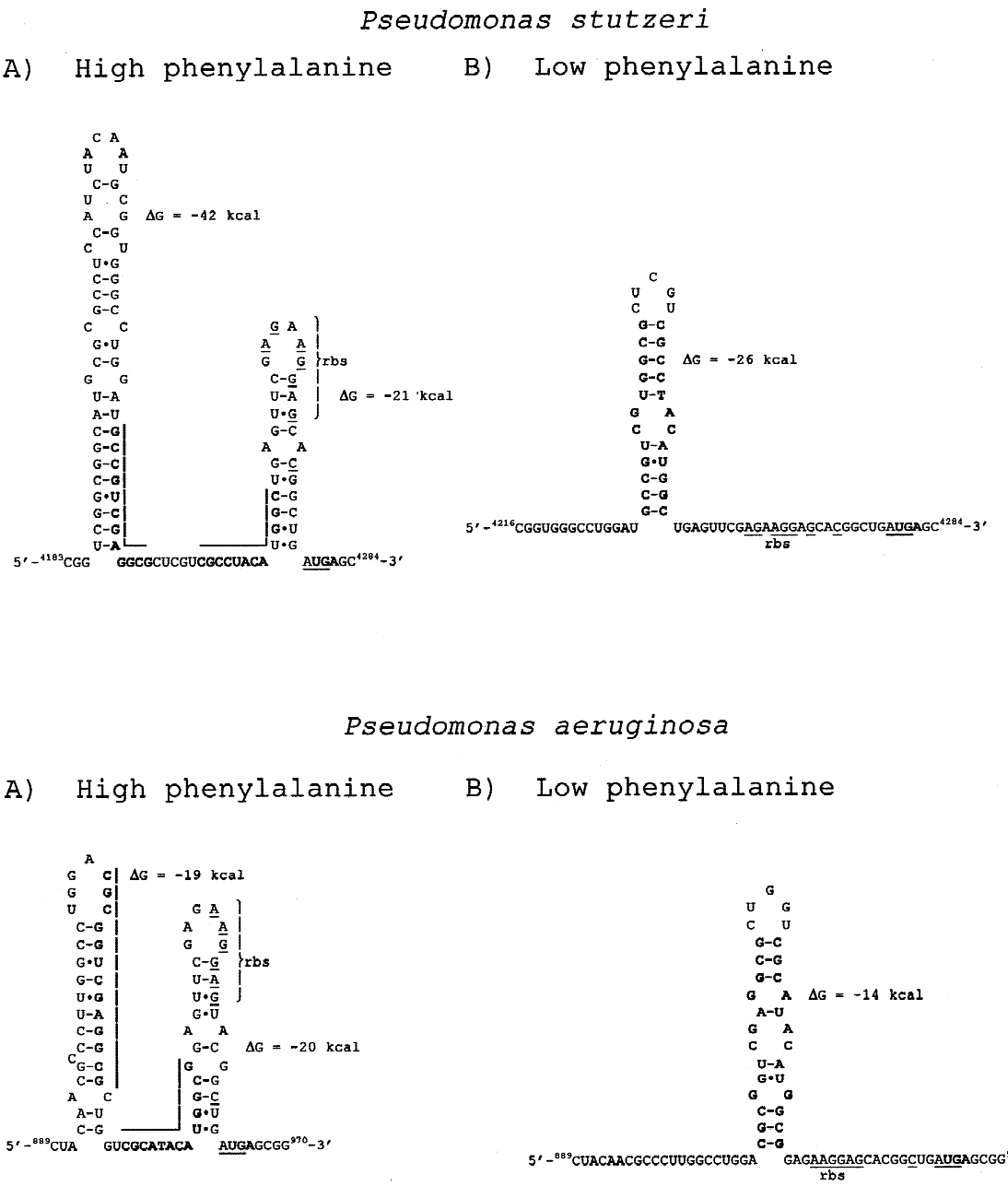

Fig. 3. Alternative stem-loop structures between *ser*C and *aro*Q_p·*phe*A genes of *P. stutzeri* and *P. aeruginosa.* Configurations in which the ribosome binding site (rbs) could be either sequestered or unmasked are shown in panels **A** or **B,** respectively.

scribed by Gu et al. (1995). A single, coincident peak of activity was obtained using L-phenylalanine, L-tyrosine, or histidinol phosphate as substrates in combination with α-ketoglutarate (2.9, 2.8, and 1.1 μmol min^{-1} mg^{-1}, respectively). Activity with oxaloacetate was negligible. Precise kinetic data were not obtained, but it was apparent that the K_m for histidinol phosphate was well over an order of magnitude greater than with L-phenylalanine or L-tyrosine. *P. stutzeri* $HisH_b$ thus has catalytic properties similar to those of *Zymomonas mobilis* (Gu et al. 1995).

DNA Sequence and Identification of the P. stutzeri tyrA_c *Gene.* The ATG start codon of *tyr*A_c is separated from the TGA stop codon of *his*H_b by only 24 bp. *tyr*A_c is 885 bp in length and encodes cyclohexadienyl dehydrogenase. $TyrA_c$ is sensitive to feedback inhibition by L-tyrosine or 4-hydroxyphenylpyruvate. *P. stutzeri tyr*A_c was found to complement the tyrosine auxotrophy of *E. coli* AT2471. Extensive characterization of the physical and catalytic properties of the purified enzyme have recently been completed (Xie and Jensen, submitted). K_m values for L-arogenate, prephenate, and NAD^+ were 0.56, 0.07, and 0.14 m*M,* respectively.

DNA Sequence and Identification of the P. stutzeri aro*F Gene.* A partial ORF of 2814 nt (3′-truncated) cod-

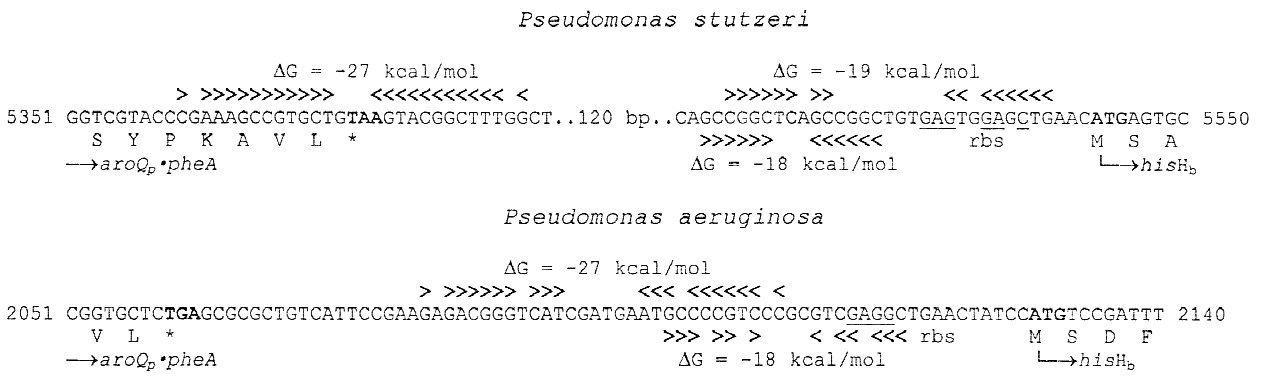

Fig. 4. The stem-loop structures between *aro*Q_p·*phe*A and *his*H_b genes of *P. stutzeri* and *P. aeruginosa.* Deduced amino acids are shown *beneath* the corresponding codons. The Shine–Dalgarno region is *underlined* and labeled (rbs). Translational start sites are indicated by *bent arrows,* and stop codons by *asterisks.* Nucleotides forming the complementary stems are marked with tandem *arrowheads.* Nucleotide numbers assigned are shown at the *far left* and *far right.*

ing for *aro*F was found immediately downstream of *tyr*A_c. The *aro*F gene encodes EPSP synthase, which is the sixth enzyme in the seven-step common pathway for aromatic amino acid biosynthesis (the prechorismate pathway). This enzyme has attracted considerable attention because it is the site of inhibition in plants of the commercially successful herbicide glyphosate (Steinrücken and Amrhein 1980). Pathogenic bacteria with mutations in *aro*F (the gene encoding EPSP synthase) are, in most cases, attenuated in mammalian infection models (Hoiseth and Stocker 1981; Bowe et al. 1989; Roberts et al. 1990), probably due to a lack of one or more of the aromatic compounds required for bacterial growth. Thus, this enzyme has also attracted attention in its use to construct live attenuated bacterial strains that would not revert to virulence in the host.

The predicted amino acid sequence of *P. stutzeri* AroF shows 80% identity with *P. aeruginosa* AroF, 56% identity with *aro*F from *Burkholderia pseudomallei,* and a high similarity to *aro*F gene products of other organisms. The sequences of the *aro*F genes are now available from a large number of bacteria, archaea, and eukaryotes (yeasts and plants).

The P. stutzeri/P. aeruginosa *Mixed-Function Supraoperons.* Our sequencing of an 8-kb DNA fragment from *P. stutzeri* and analysis of the corresponding *P. aeruginosa* genes uncovers a gene organization that suggests a mixed-function supraoperon. The compact intergenic spacing between *his*H_b–*tyr*A_c (19 bp) and between *tyr*A_c–*aro*F (9 bp), the translationally coupled *ser*C and *aro*Q_p·*phe*A genes (**TGA**TG), the absence of any obvious terminator structure between *aro*Q_p·*phe*A and *his*H_b, and the translationally coupled *aro*F and *cmk* genes (**TGA**TG) suggest the possibility of a single transcriptional unit that extends at least as far as *cmk.* In view of the complexity of other systems, additional overlapping transcripts produced by possible internal promoters and/or attenuators seem quite likely.

Figure 2 shows a comparison of the gene organization present in the partial *P. stutzeri* supraoperon with that of *P. aeruginosa.* These closely related but physiologically distinctive species exhibit a supraoperon composition and organization that can be considered nearly identical.

Discussion

Regulation of aroQ_p·phe*A in* Pseudomonas. Enteric bacteria such as *E. coli* (Hudson and Davidson 1984) and *Erwinia herbicola* (Xia et al. 1993) control *aro*Q_p·*phe*A solely by attenuation. The latter mode of attenuation depends upon a ribosome-stalling mechanism operating during translation of a phenylalanine-rich leader peptide located immediately upstream of *aro*Q_p·*phe*A. In this case, alternative stem-loop structures are favored, depending upon the rate of leader-peptide translation. *Pseudomonas stutzeri aro*Q_p·*phe*A clearly lacks an upstream leader peptide that would be needed to participate in a ribosome-stalling mechanism since it, in fact, overlaps with *ser*C. Although the gene product of *ser*C itself might conceivably act as a leader peptide, this seems unlikely since *ser*C is not phenylalanine-rich at the carboxy terminus.

In *Xanthomonas campestris* Gu et al. (1997) noted the existence of alternative stem-loop structures between *ser*C and *aro*Q_p·*phe*A—one a possible antiterminator structure and the other a *rho*-independent terminator. Since no upstream leader peptide is present, an attenuation mechanism exploiting the alternative stem-loop structures would require a presently unknown regulatory element (e.g., an RNA-binding protein). A pair of antiterminator/terminator structures, such as exemplified by *X. campestris,* was not found in the region of *ser*C/*aro*Q_p·pheA overlap in *P. stutzeri.*

However, a strong stem-loop structure was identified (Figs. 3A, left) which sequesters the ribosome-binding site of *aro*Q_p·*phe*A. This could provide a mechanism

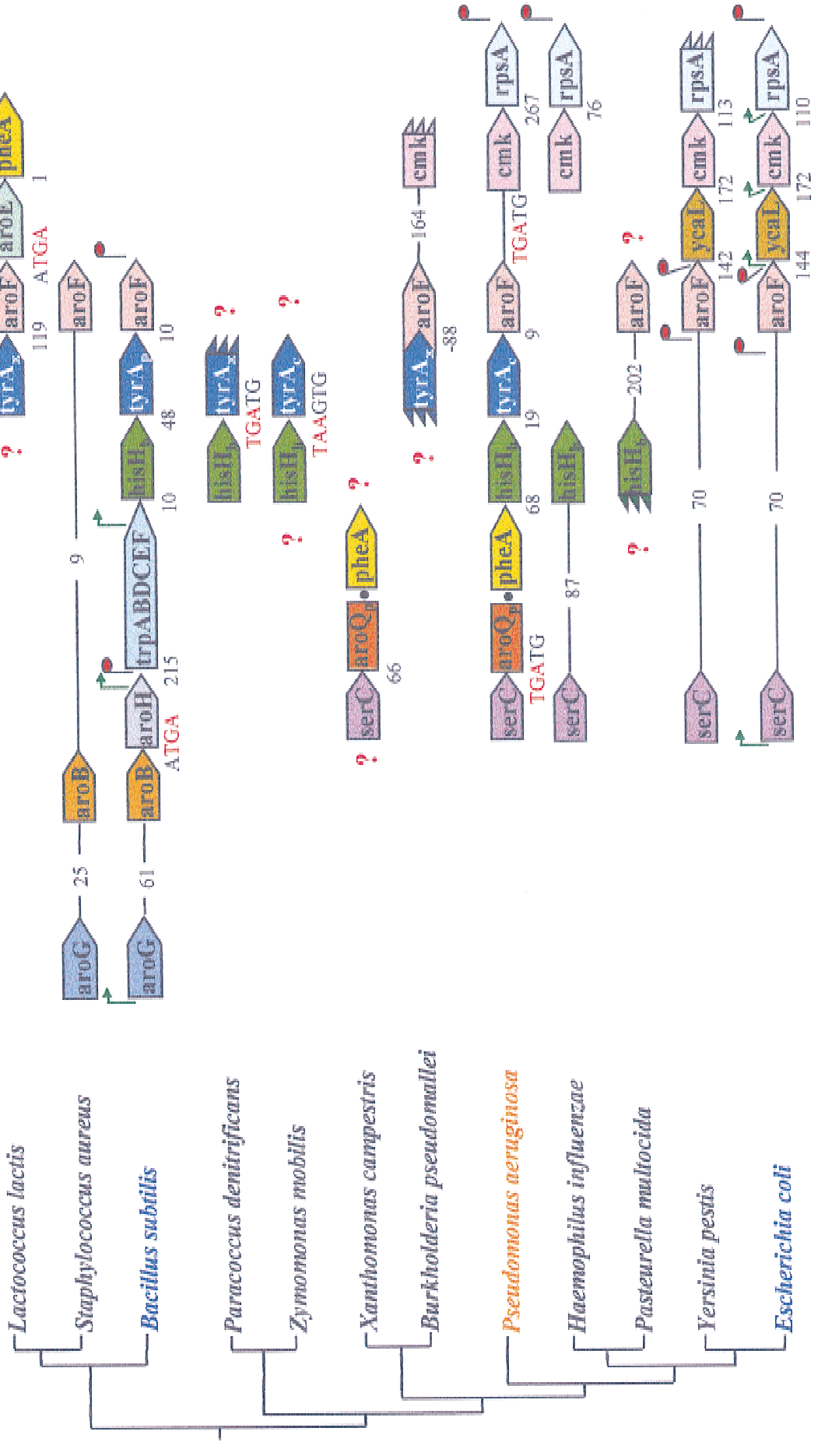
Lactococcus lactis
Staphylococcus aureus
Bacillus subtilis
Paracoccus denitrificans
Zymomonas mobilis
Xanthomonas campestris
Burkholderia pseudomallei
Pseudomonas aeruginosa
Haemophilus influenzae
Pasteurella multocida
Yersinia pestis
Escherichia coli
aroG
aroB
aroH
trpABDCEF
tyrA
aroF
aroE
pheA
hisH
tyrA
tyrA
serC
aroQ
pheA
hisH
tyrA
aroF
cmk
rpsA
serC
hisH
tyrA
aroF
cmk
rpsA
serC
aroF
ycaL
cmk
rpsA
serC
aroF
ycaL
cmk
rpsA
ATGA
TGATG
TAAGTG
TGATG
ATGA

Fig. 5. Conserved gene organization in nature. Organisms having the gene organizations shown are placed on a dendrogram derived from 16S rRNA sequence comparisons at the *left*. Homologous genes are *color-coded*. A hypothetical protein in *Yersinia* is shown in white. Intergenic distances are not shown proportionally, but intergenic base-pair distances are indicated. (In the *B. subtilis trp* operon all of the structural genes overlap except for a single genie interface.) Genes connected by a *bar* are adjacent. Incomplete gene sequences are indicated with *ragged edges*. Flanking regions marked with *question marks* might contain genes of interest, but these regions have not yet been sequenced. In the well-studied *B. subtilis* and *E. coli* systems where transcript analysis has been done, promoters are shown with *bent green arrows*, and transcription terminators or attenuators are shown with *red flags*. In other cases, where stem-loop structures followed by a run of uridine bases in the mRNA are likely candidates for *rho*-independent terminators, these are also shown by *red flags*. The following genes (not drawn in proportion to actual size) encode the indicated enzymes: *aro*B → dehydroquinate synthase; *aro*E → shikimate kinase; *aro*F → 5-enolpyruvylshikimate-3-P synthase; *aro*G → chorismate synthase; *aro*H → the AroH homology class of chorismate mutase; *aro*Q_p· → a member of the AroQ homology class of chorismate mutase which is fused to a prephenate dehydratase domain encoded by ·*phe*A; *his*H_b → broad-specificity class of imidazoleacetol-P aminotransferase; *tyr*A_c → tyrosine-pathway dehydrogenase utilizing either prephenate or L-arogenate (cyclohexadienyl dehydrogenase); *tyr*A_p → tyrosine-pathway dehydrogenase specific for prephenate; *tyr*A_x → tyrosine-pathway dehydrogenase having unknown substrate specificity; *ser*C → 3-phosphoserine aminotransferase (synonymous with *pdx*F, an aminotransferase of pyridoxine biosynthesis); *trp*ABFCDE → the tryptophan pathway suite of enzymes (except for *trp*G); *cmk* → cytidylate kinase; *rps*A → ribosomal protein S1.

whereby *ser*C translation is required to activate *aro*Q_p·*phe*A translation by unmasking the sequestered ribosome binding site. The additional presence of an alternative stem-loop structure (Figs. 3B, right) might reflect a mechanism whereby *aro*Q_p·*phe*A is differentially regulated at the translational level. Thus, under conditions of L-serine sufficiency where *ser*C translation is minimal, *aro*Q_p·*phe*A translation may be uncoupled from *ser*C translation if a secondary mRNA structure is presented as illustrated under "low-phenylalanine" conditions of nutrition. This would require a presently unknown regulatory gene.

The plausibility that the *P. stutzeri* alternative stem-loop structures are operational regulatory elements is reinforced by the finding that similar alternative stem-loop structures—one sequestering the ribosome binding site—were found in the *P. aeruginosa* system (Fig. 3, bottom). In recent years regulatory mechanisms acting at the level of translation initiation have gained recognition as being more important than previously thought in prokaryotes (De Smit and Van Duin 1990).

Intercistronic Region Between aroQ_p·phe*A and* hisH_b. Within the supraoperon boundaries, only the intergenic junctions separating *aro*Q_p·*phe*A/*his*H_b and *cmk/rps*A are sufficiently long to contain promoters, attenuators and/or associated regulatory elements that do not overlap coding regions. Figure 4 illustrates the mRNA stem-loop regions which potentially exist in the *aro*Q_p·*phe*A/*his*H_b intergenic regions of *P. stutzeri* and *P. aeruginosa*.

In *P. stutzeri* a strong hairpin structure ($\Delta G = -27.0$ kcal/mol) overlaps the ·*phe*A stop codon. This could be an attenuator structure. If so, unknown elements of regulation upstream and downstream may act to remediate the imperfect uridine-rich segment present, as reported in other systems (Reynolds et al. 1992). Downstream at the far 3′ end of the intergenic space is another hairpin structure ($\Delta G = -19.0$ kcal/mol) which sequesters the ribosome binding site. An alternative stem-loop structure ($\Delta G = -18.0$ kcal/mol) would expose the ribosome-binding site. Perhaps unknown elements of regulation exist to dictate the stabilization of one structure or the other?

The shorter intergenic region of *P. aeruginosa* was also examined for comparable secondary-structure possibilities (Fig. 4). No hairpin resembling the far upstream stem-loop of *P. stutzeri* was found. However, a downstream hairpin structure ($\Delta G = -18.0$ kcal/mol) sequestering the ribosome-binding site was found, also accompanied by an alternative stem-loop structure ($\Delta G = -27.0$ kcal/mol).

No hairpin structures were found in the intergenic regions between *his*H_b and *tyr*A_c or between *tyr*A_c and *aro*F. Thus, the features of close intergenic spacing and possibilities of mechanisms masking ribosome-binding sites prior to the event of upstream translation may ac-

commodate coupled translation through much or all of the system under some conditions. Under other conditions, translation may be uncoupled for selective expression of some genes.

Intercistronic Region Between cmk *and* rps*A. cmk* has been shown to lie within the *rps*A operon in *E. coli*, where it is required for a normal replication rate (Fricke et al. 1995). The transcript is read from multiple promoters (Fig. 5), and it appears that the putative heat-shock protein YcaL may also lie within the *rps*A operon. YcaL, a protein with few homologues in the database, appears to be a recent addition to the supraoperon of enteric bacteria. *P. aeruginosa* possesses a strong stem-loop structure ($\Delta G = -33$ kcal/mol) between *cmk* and *rps*A which may be a site of regulation. The lack of an obvious terminator structure, the conservation of *cmk* linkage with *rps*A in other gram-negative bacteria, and the established location of *cmk* in an operon with *rps*A in *E. coli* are all consistent with the possible inclusion of *rps*A within the *Pseudomonas* supraoperon.

Intergenomic Conservation of the Pseudomonas *Supraoperon.* With the identification of the eight protein domains comprising the supraoperon shown in Fig. 5 for *P. aeruginosa* (and *P. stutzeri*), we noticed an overlapping resemblance to two established supraoperon systems reported in the literature. First, the *his*H_b–*tyr*A_c–*aro*F gene block in the *Pseudomonas* supraoperon exactly matches the *his*H_b–*tyr*A_p–*aro*F segment at the 3′ end of the *B. subtilis* supraoperon. Second, the well-known *ser*C–*aro*F supraoperon of *E. coli* contains genes present in the *Pseudomonas* supraoperon but lacks the intervening genes which, in *Pseudomonas*, encode four additional catalytic domains.

In addition, the *aro*F–*cmk*–*rps*A linkage of *Pseudomonas* also prevails in *E. coli*, except that *yca*L is inserted between *aro*F and *cmk* in *E. coli*. In *H. influenzae*, only the *ser*C–*his*H_b linkage has been retained. In *Pasteurella* the *his*H_b–*aro*F linkage has been retained, but without the intervening *tyr*A. A relatively small phylogenetic cluster roughly equivalent to the *E. coli*–*Haemophilus* portion of the dendrogram shown in Fig. 5 shares a recent evolutionary event resulting in fusion of a catalytic domain for chorismate mutase to the $TyrA_c$ domain (denoted $AroQ_t{\cdot}TyrA_c$). Dynamic supraoperon reshuffling (i.e., loss of both *aro*Q_p·*phe*A and *tyr*A_c from the supraoperon) seems to be correlated with this evolutionary event. This might be consistent with the proposed origin of *aro*Q_t· from *aro*Q_p· (Ahmad and Jensen 1988) via gene duplication, followed by recombinational events which resulted in gene translocation.

Among gram-negative bacteria, a comparison of *E. coli* and *P. aeruginosa* is reminiscent of a comparison between *B. subtilis* and *Staphylococcus* among gram-positive bacteria in that *Staphylococcus* has retained only the 5′ and 3′ ends of the *B. subtilis* supraoperon. At this point one might project that the *his*H_b–*tyr*A–*aro*F gene organization will prove to be to be widely conserved throughout the entire gram-positive/gram-negative assemblage of organisms. *ser*C–*aro*F–*cmk*–*rps*A linkage may be generally retained in gram-negative bacteria.

Interlocking Metabolic Relationships. By definition, mixed-function supraoperons contain genes whose relationships with one another are not as obvious as those encoding steps of simple linear pathways. However, metabolic pathways are in fact intricately interwoven networks that tend to be simplified by reductionist textbook approaches. The putative supraoperons of *P. stutzeri* and *P. aeruginosa* each contains genes of serine, pyridoxine, histidine, and aromatic amino acid biosynthesis. Although not immediately apparent, distinct and overlapping relationships of SerC(PdxF) and $HisH_b$ with aromatic amino acid biosynthesis do exist.

In nature HisH proteins are uniquely equipped to transaminate imidazoleacetol phosphate for histidine biosynthesis, and the $HisH_n$ subclass is specific for this substrate. However, the $HisH_b$ subclass has a broadened substrate specificity which allows the additional utilization of the bulky aromatic substrates. In *B. subtilis* $HisH_b$ functions both in histidine biosynthesis and in aromatic amino acid biosynthesis (Nester and Montoya 1976).

Although the *ser*C–*aro*F relationship in *E. coli* has been rationalized mainly in terms of the input of both genes into enzymes needed ultimately for iron siderophore (enterochelin) production, this does not explain the broader retention of this linkage in organisms which do not produce enterochelin. We propose that because *ser*C is intimately tied to the mechanism for introduction of the amino group into the sidechain of tryptophan, there has been selection for a balanced expression of *ser*C with genes of phenylalanine and tyrosine biosynthesis. This includes *his*H_b (in those organisms that possess it) whose gene product introduces the amino group into the side chains of phenylalanine and tyrosine. A salvage pathway operating between serine and tryptophan is illustrated in Fig. 6. Biosynthesis of every molecule of tryptophan requires input of one serine molecule, serine being the source of the sidechain. Glyceraldehyde-3-P derived from the indoleglycerol phosphate substrate molecule is released, and this 3-carbon unit can be recycled to serine via the two enzyme steps of glycolysis shown. The generation of ATP largely offsets the metabolic expense of PRPP utilization in the second step of tryptophan biosynthesis. Thus, the transamination step carried out by SerC can be considered to have an impact for tryptophan biosynthesis that is equivalent to the impact of $HisH_b$ upon tyrosine and phenylalanine biosynthesis. Together, SerC and $HisH_b$ are responsible for introducing the

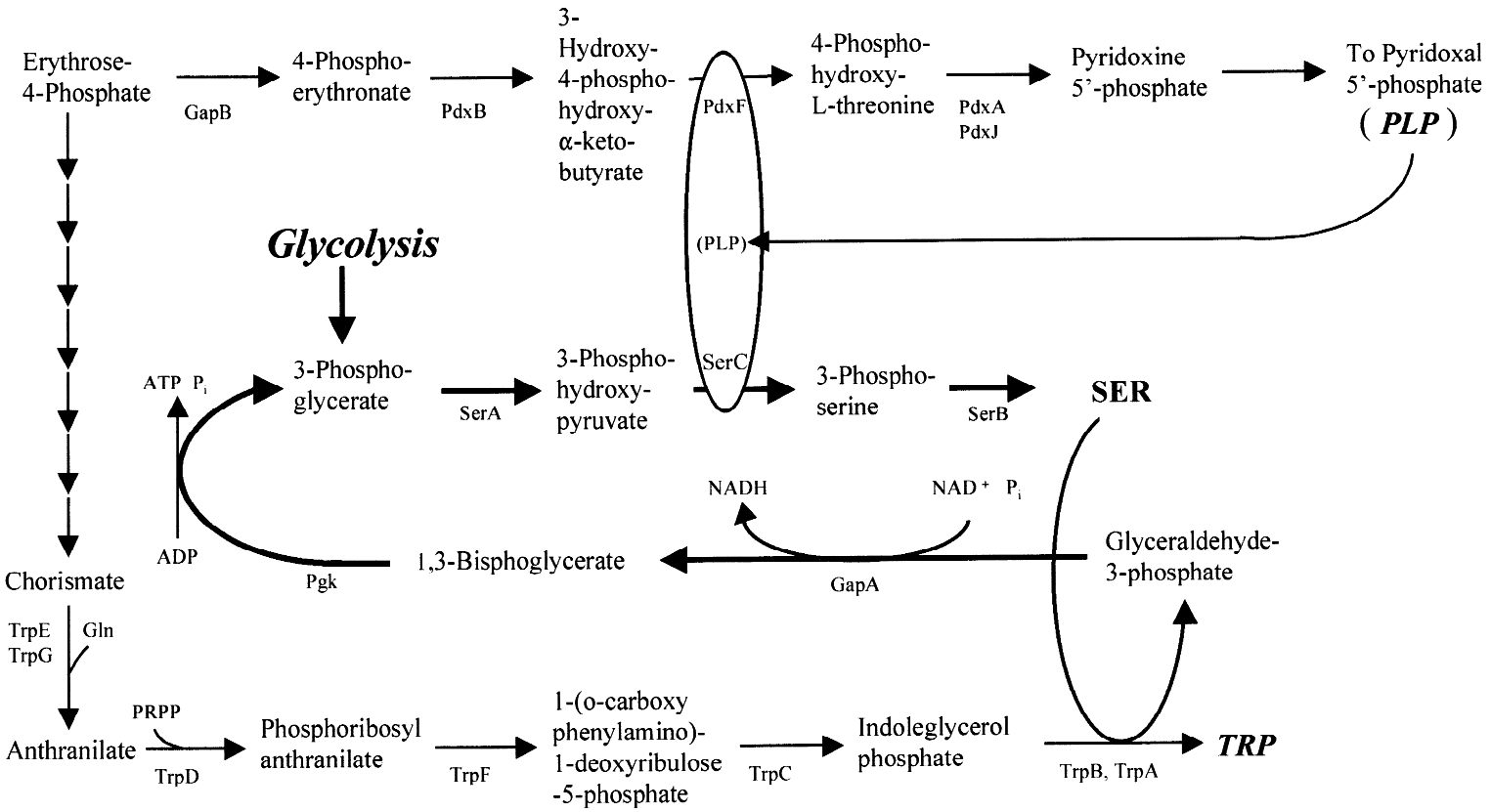

Fig. 6. The serine salvage pathway. In the last step of tryptophan biosynthesis, serine combines with indoleglycerol phosphate to produce tryptophan and glyceraldehyde-3-P. The two glycolytic enzymes, glyceraldehyde-3-P dehydrogenase (GapA) and phosphoglycerate mutase (Pgk), recycle the 3-carbon glyceraldehyde-3-P to 3-phosphoglycerate. In higher plants growing under conditions of inorganic phosphate (P_i) and adenylate insufficiency, an NADP-dependent dehydrogenase (nonphosphorylating) exists which can convert glceraldehyde-3-P directly to 3-phosphoglycerate (Plaxton 1996). TRP, L-tryptophan; Gln, L-glutamine; Glu, glutamate: TrpE and TrpG, large and small subunits of anthranilate synthase; TrpD, anthranilate phosphoribosyl transferase; TrpF, phosphoribosyl anthranilate isomerase; TrpC, indoleglycerol phosphate synthase; TrpB and TrpA, tryptophan synthetase; SerA, 3-phosphoglycerate dehydrogenase; SerB, 3-phosphoserine phosphatase; SerC, 3-phosphoserine aminotransferase; GapB, erythrose-4-P dehydrogenase; PdxB, 4-phosphoerythronate dehydrogenase; PdxA and PdxJ are thought to be involved in the formation of the pyridine ring of pyridoxine 5′-phosphate from 1-deoxy-D-xylulose and 4-phosphohydroxy-L-threonine (Hill and Spenser 1996).

amino group into the sidechains of each of the three aromatic amino acids.

Presumably, tryptophan production would normally place a fractionally small demand upon the total pool resource of serine. The location of *ser*C in the aromatic supraoperon may be a mechanism primarily for elevation of SerC levels under conditions of need for unusually high tryptophan production.

It is unclear what the basis for the relatively strong conservation of the supraoperon gene organization might be. Although a general rationale favoring selective pressure for coregulation of these genes certainly exists, the gene organization of more straightforward gene candidates for coregulation seem to have been widely dispersed with evolutionary time. For example, tryptophan-pathway genes in *P. aeruginosa* are scattered into three widely spaced groups—rather than coexisting within one operon, as they are in the relatively close relative, *E. coli* (Crawford 1989). Since the few examples of highly conserved gene organizations across wide phylogenetic boundaries are those in which physically associated gene products are made (Mushegian and Koonin 1996; Dandekar et al. 1998), perhaps the supraoperon enzymes form protein complexes. Translational coupling is suggested by compact intergenic spacing, and cotranslational folding is thought to be an essential component of biosynthetic folding of many proteins (Federov and Baldwin 1997).

Acknowledgment. This is Florida Agricultural Experiment Station journal series no. R-06848.

References

Ahmad S, Jensen RA (1988) New prospects for deducing the evolutionary history of metabolic pathways in prokaryotes: Aromatic biosynthesis as a case-in-point. Origins Life Evol Biosphere 18: 41–57

Altschul SF, Madden TL, Schaffer AA, Zhang JH, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Baldwin GS, Davidson BE (1981) A kinetic and structural comparison of chorismate mutase/prephenate dehydratase from mutant strains of *Escherichia coli* K-12 defective in the *pheA* gene. Arch Biochem Biophys 211:66–75

Bowe F, O'Gaora P, Maskell D, Cafferkey M, Dougan G (1989) Virulence, persistence and immunogenicity of *Yersinia enterocolitica* 0:8 *aroA* mutants. Infect Immun 57:3234–3236

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72:248–254

Carlson CA, Pierson LS, Rosen JJ, Ingraham JI (1983) *Pseudomonas stutzeri* and related species undergo natural transformation. J Bacteriol 153:93–99

Crawford IP (1989) Evolution of a biosynthetic pathway: The tryptophan paradigm. Annu Rev Microbiol 43:567–600

Dagert M, Ehrlich SD (1979) Prolonged incubation in calcium chloride improves the competence of *Escherichia coli* cells. Gene 6:23–28

Dandekar T, Snel B, Huynen MA, Bork P (1998) Conservation of gene order: A fingerprint of proteins that physically interact. TIBS 23: 324–328

Davis RW, Botstein D, Roth JR (1980a) Advanced bacterial genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, p 201

Davis RW, Thomas M, Cameron J, St John TP, Scherer S, Padgett RA (1980b) Rapid DNA isolations for enzymatic and hybridization analysis. Methods Enzymol 65:404–411

De Smit MH, Van Duin J (1990) Secondary structure of the ribosome binding site determines translational efficiency: A quantitative analysis. Proc Natl Acad Sci USA 87(19):7668–7672

Devereux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12: 387–395

Duncan K, Coggins JR (1986) The *serC-aroA* operon of *Escherichia coli.* Biochem J 234:49–57

Fedorov AN, Baldwin TO (1997) Cotranslational protein folding. J Biol Chem 272:32715–32718

Fischer RS, Zhao G, Jensen RA (1991) Cloning, sequencing, and expression of the P-protein gene (*pheA*) of *Pseudomonas stutzeri* in *Escherichia coli:* Implications for evolutionary relationships in phenylalanine biosynthesis. J Gen Microbiol 137:1293–1301

Fricke J, Neuhard J, Kelln RA, Pederson S (1995) The *cmk* gene encoding cytidine monophosphate kinase is located in the *rpsA* operon and is required for normal replication rate in *Escherichia coli.* J Bacteriol 177:517–523

Gelfand DH, Steinberg RA (1977) *Escherichia coli* mutants deficient in the aspartate and aromatic amino acid aminotransferases. J Bacteriol 130:429–440

Gu W, Zhao G, Eddy C, Jensen RA (1995) Imidazole acetol phosphate aminotransferase in *Zymomonas mobilis:* Molecular genetic, biochemical, and evolutionary analyses. J Bacteriol 177:1576–1584

Gu W, Williams DS, Aldrich HC, Xie G, Gabriel DW, Jensen RA (1997) The *AroQ* and *PheA* domains of the bifunctional P-protein from *Xanthomonas* campestris in a context of genomic comparison. Microbiol Comp Genom 2:141–158

Henner D, Yanofsky C (1993) Biosynthesis of aromatic amino acids. In: Sonenshein AL, Hoch J, Losick R (eds). *Bacillus subtilis* and other gram-positive bacteria: Biochemistry, physiology, and molecular genetics. ASM Press, Washington, DC, pp 269–280

Hill RE, Spenser ID (1996) Biosynthesis of vitamin B_6. In: Neidhardt FC (ed) *Escherichia coli* and *Salmonella:* Cellular and molecular biology, 2nd ed., Vol 1, ASM Press, Washington, DC, pp 695–703

Hohn B, Collins J (1980) A small cosmid for efficient cloning of large DNA fragments. Gene 11:291–298

Hoiseth SK, Stocker BAD (1981) Aromatic-dependent *Salmonella typhimurium* are non-virulent and are effective live vaccines. Nature 291:238–239

Horwitz MS, Loeb IA (1988) An *E. coli* promoter that regulates transcription by DNA superhelix-induced cruciform extrusion. Science 241:703–705

Hudson GS, Davidson BE (1984) Nucleotide sequence and transcription of the phenylalanine and tyrosine operons of *Escherichia coli* K-12. J Mol Biol 180:1023–1051

Jensen RA, Gu W (1996) Evolutionary recruitment of biochemically specialized subdivisions of family I within the protein superfamily of aminotransferases. J Bacteriol 178:2161–2171

Lam HM, Winkler ME (1990) Metabolic relationships between pyridoxine (vitamin B_6) and serine biosynthesis in *Escherichia coli* K-12. J Bacteriol 172:6518–6528

LeMaster DM, Richards FM (1987) NMR sequential assignment of *Escherichia coli* thioredoxin utilizing random fractional deuteriation. Biochemistry 27:142–150

Lyngstadass A, Løbner-Olesen A, Boye E (1995) Characterization of three genes in the dam-containing operon of *Escherichia coli.* Mol Gen Genet 247:546–554

Man TK, Pease AJ, Winkler ME (1997) Maximization of transcription of the *serC(pdxF)-aroA* multifunctional operon by antagonistic effects of the cyclic AMP (cAMP) receptor protein-cAMP complex

and Lrp global regulators of *Escherichia coli* K-12. J Bacteriol 179:3458–3469

Maniatis T, Fritsch EF, Sambrook J (1989) Molecular cloning: A laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Miller JH (1972) Experiments in molecular genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Mushegian AR, Koonin EV (1996) Gene order is not conserved in bacterial evolution. Trends Genet 12:289–290

Nester EW, Montoya AL (1976) An enzyme common to histidine and aromatic amino acid biosynthesis in *Bacillus subtilis.* J Bacteriol 126:699–705

Palleroni NJ (1984) Gram-negative aerobic rods and cocci Family I. Pseudomonadaceae. In: Krieg NR, Holt JG (eds) Bergeys manual of systematic bacteriology, Vol 1. Williams & Wilkins, Baltimore, pp 172–173

Plaxton WC (1996) The organization and regulation of plant glycolysis. Annu Rev Plant Physiol 47:185–214

Reece RJ, Maxwell A (1991) DNA gyrase: Structure and function. Crit Rev Biochem Mol Biol 26:335–375

Reynolds R, Bermudez-Cruz RM, Chamberlin MJ (1992) Parameters affecting transcription termination by *E. coli* RNA polymerase. J Mol Biol 224:31–51

Roberts M, Maskell D, Novotny P, Dougan G (1990) Construction and characterization *in vivo* of *Bordetella pertussis aroA* mutants. Infect Immun 58:732–739

Siefert JL, Martin KA, Abdi F, Widger WR, Fox GE (1997) Conserved gene clusters in bacterial genomes provide further support for the primacy of RNA. J Mol Evol 45:467–472

Steinrücken HC, Amrhein N (1980) The herbicide glyphosate is a potent inhibitor of 5-enolpyruvyl-shikimic acid-3-phosphate synthase. Biochem Biophys Res Commun 94:1207–1212

Subramaniam PS, Xie G, Xia T, Jensen RA (1998) Substrate ambiguity of 3-deoxy-D-*manno*-octulosonate 8-phosphate synthase from *Neisseria gonorrhoeae* in the context of its membership in a protein family containing a subset of 3-deoxy-D-arabino-heptulosonate 7-phosphate synthases. J Bacteriol 180:119–127

Tatusov RL, Mushegian AR, Bork P, et al. (1996) Metabolism and evolution of *Haemophilus influenzae* deduced from a whole-genome comparison with *Escherichia coli.* Curr Biol 6:279–291

Tsui HC, Pease AJ, Kochler T, Winkler ME (1994a) Detection and quantitation of RNA transcribed from bacterial chromosomes. In: Adolph K (ed) Methods in molecular genetics: Molecular microbiology. Academic Press, New York, pp 179–204

Tsui HC, Zhao G, Feng G, Leung HC, Winkler ME (1994b) The *mutL* repair gene of *Escherichia coli* K-12 forms a superoperon with a gene encoding a new cell-wall amidase. Mol Microbiol 11:189–202

Watanabe H, Mori H, Itoh T, Gojobori T (1997) Genome plasticity as a paradigm of eubacteria evolution. J Mol Evol 44 (Suppl 1):S57–S64

Zhao G, Xia T, Ingram L, Jensen RA (1993) An allosterically insensitive class of cyclohexadienyl dehydrogenase from *Zymomonas mobilis.* Eur J Biochem 212:157–165

Xia T, Zhao G, Jensen RA (1993) The pheA/tyrA/aroF region from Erwinia herbicola: An emerging comparative basis for analysis of gene organization and regulation in enteric bacteria. J Mol Evol 36:107–120

Yanisch-Perron C, Vieira J, Messing J (1985) Improved M13 phage cloning vectors and host strains: Nucleotide sequences of the M13mp18 and pUC 19 vectors. Gene 33:103–119

Yuan R, Lin Y (1982) The purification of restriction enzymes and DNA substrates. Their use in DNA sequencing. In: Huang PC, Kuo TT, Wu R (eds) Genetic engineering techniques: Recent developments. Academic Press, New York, pp 337–349