ISSN 0026-8933, Molecular Biology, 2007, Vol. 41, No. 3, pp. 513–515. © Pleiades Publishing, Inc., 2007.
Original Russian Text © A.V. Seliverstov, V.A. Lyubetsky, 2007, published in Molekulyarnaya Biologiya, 2007, Vol. 41, No. 3, pp. 572–574.

SHORT COMMUNICATIONS

UDC 577.053

Regulation of Proline Biosynthesis in Proteobacteria

A. V. Seliverstov and V. A. Lyubetsky

Institute of Information Transmission Problems, Russian Academy of Sciences, Moscow, 127994 Russia; e-mail: slvstv@iitp.ru

Received June 7, 2006
Accepted for publication August 28, 2006

DOI: 10.1134/S002689330703017X

Key words: proline biosynthesis, signal search, gene expressio

Bacterial biosynthesis of proline from glutamic acid involves, consecutively, three enzymes: γ-glutamyl kinase, γ-glutamyl phosphate reductase, and 1-pyrroline-5-carboxylate reductase. In *Pseudomonas aeruginosa*, the production of the first two enzymes depends on the proline concentration, while that of the third enzyme does not [1]. The first two enzymes are encoded by *pro*A and *pro*B, which usually form an operon in proteobacteria. *Pseudomonas pro*A and *pro*B are far apart on the chromosome and are transcribed in opposite directions. In *Bacillus subtilis*, transcription of the *pro*BA operon is regulated using the T box [2], which is involved in a general mechanism regulating the genes for aminoacyl-tRNA synthases, amino acid-metabolizing proteins, and amino acid transporters and occurs mostly in Gram-positive bacteria [3, 4]. Since such a regulation is rare in Gram-negative bacteria, it is of interest to study the protein–DNA regulation. In *Escherichia coli*, the above three enzymes are inhibited by proline and there are no data on protein–DNA regulation.

In this work, we performed a mass search for a regulatory protein-binding site in the upstream regions of the three proline biosynthesis genes. Bacterial genome sequences were extracted from GenBank. Conserved regions were sought using a program based on an original algorithm, which finds a clique in a multidivisible graph [5].

First, we examined the 5'-leader regions sized up to 200 bp of operons containing *pro*A or *pro*B in all completely sequenced genomes of γ- and α-proteobacteria. When a leader region was greater, we analyzed only the first 200 bp upstream of the first gene of the operon. A mass search of the leader regions revealed conserved sequences, which presumably contained regulatory sites of the corresponding operons. Conserved sequences of 16 bp with the consensus MTACCAAYNNNYYTSY were found in γ-proteobacteria of the genera *Pseudomonas* and *Shewanella.* A site located upstream of *pro*A, coding for γ-glutaminase, was found in *P. aeruginosa* PAO1, *P. putida* KT2440, *P. syringae* pv. *tomato* str. DC3000, *Shewanella oneidensis* MR-1, *S. amazonensis* SB2B, *Shewanella* sp. PV-4, *S. frigidimarina* NCIMB 400, *S. baltica* OS155, *S. putrefaciens* CN-32, *Shewanella* sp. MR-4, *Shewanella* sp. MR-7, and *Shewanella* sp. ANA-3. A site located upstream of *pro*B, coding for γ-glutamyl phosphate reductase, was found in *P. aeruginosa* PAO1 and *P. fluorescens* Pf-5. A corresponding multiple sequence alignment is shown in the figure. Note that, in 7 out of 15 cases, the site was repeated in the same leader region with a small distance between the repeats, suggesting a cooperative binding of the regulator. The algorithm did not find conserved sites upstream of the operons containing *pro*A or *pro*B in the other genomes examined. In addition, the algorithm did not find conserved hairpins (inverted repeats in the DNA sequence) in the leader regions of the genes in question.

It is possible to assume that the above conserved DNA region coincides with or contains a binding site for a protein that regulates transcription as dependent on the proline concentration. In *Pseudomonas* and *Shewanella*, sites regulating transcription of the γ-glutamyl kinase gene were predicted upstream of TATAA, which is characteristic of the (–10) promoter box.

Such a site was not found upstream of the operons under study in *Azotobacter vinelandii*, although this organism is evolutionarily close to the bacteria possessing the predicted site. The site was detected only upstream of the γ-glutamyl kinase gene in *P. putida*

	Upstream of the γ-glutamyl kinase gene
P. aeruginosa	t**CTACCGACCGGCTGTT****gcggctgggcgcgccggggctggctcgc**TATAA**tc
P. putida	g**CCACCCGCCGCTGGGC******gt**ACACCTGCGCGGCAGT************TATAA**tc
P. syringae	t**CGACCATCCGCCTGTC**tg**cgctgtcgttcgtgccccacgacagc**TATAA**tc
S. oneidensis	t**CTACCAACATCCTGCC**cactta**CTACCAATTACCTAGC**ggattagg**TATAA**gg
S. amazonensis	a**ATATCCATCGGCGCCG**ttgtgcaagggcggttgct******ttggg**TATAA**gg
S. sp. PV-4	gctctaatcttttatcttacccg**CTACCATTTGCCCTGG**ggattggg**TATAA**ct
S. frigidimarina	t**ATTCCAGTATTAGCGG**ctttca**CTATGAATTATCGAGT**gaattggg**TATAA**aa
S. baltica	tatctcgagttaataactccttg**CTACCAATTACCCAGC**ggattagg**TATAA**gg
S. putrefaciens	tatctcgagttaatcactcctcg**CTACCAATTATCTAGC**ggattagg**TATAA**gg
S. sp. MR-4	t**CTACCAACATCCTGCC**cactaa**CTACCAATTGCTCTGC**ggattagg**TATAA**gg
S. sp. MR-7	t**ATACCAACATGTAGCT**cactta**CTACCAATTGCCCAAC**ggattagg**TATAA**gg
S. sp. ANA-3	t**CTACCAACATCCGGCC**cactaa**CTACCAATTGCTCTGC**ggattagg**TATAA**gg
	Upstream of the γ-glutamyl phosphate reductase gene
P. aeruginosa	g**ATACCCCGCCGGGCGG**gccggcggggctgaagagtcgaagcctgggct**TCTAA**gg
P. fluorescens	t**CGACCATCCGCCCGTG**accgctgtcc
	Upstream of the gene for hypothetical proteinAAN56706
S. oneidensis	c**AAACCAATTCGGAGAC**tt*aaa**GATAACAATCGACTAA**gctagacacacttcg

Multiple sequence alignment of the leader regions containing the putative binding site for a regulatory protein. Putative repressor-binding sites and (–10) promoter boxes are in capitals. *P, Pseudomonas; S, Shewanella.*

and *P. syringae*, only upstream of the γ-glutamyl phosphate reductase gene in *P. fluorescens*, and in the upstream regions of both *pro*A and *pro*B in *P. aeruginosa.*

In *P. aeruginosa, pro*A and *pro*B whose upstream regions contain the predicted site have experimentally been demonstrated to respond to changes in proline concentration, although the mechanism of this regulation has not been identified [1].

The amplitude of changes in gene expression is only about 40% in *P. aeruginosa* [1]. The low efficiency of regulation is possibly explained by the fact that the putative regulatory protein-binding site is not repeated in *P. aeruginosa.* It is known that repetitive sites usually provide for a greater amplitude of changes in expression owing to a cooperative binding of the regulator.

Since *Pseudomonas* and *Shewanella* are rather distant, it is important to note that the regions between the putative sites found upstream of the orthologous genes are not conserved in species of the different genera (figure).

Second, we checked whether the multiple sequence alignment of the putative sites was arbitrary, while similar studies are commonly restricted to construction of an alignment. For this purpose, we examined the 5'-leader regions sized up to 200 bp or truncated to 200 bp for all genes in *S. oneidensis.* Sequences close to the above 16-bp consensus was found only in two cases: upstream of the *pro*AB operon and upstream of a hypothetical gene coding for unknown protein AAN56706. The sequence was repeated in either case. The first repetitive sequence was already found by the algorithm at the fist step of our study. The other one was also included in the total set of positive results at this step (figure), although the situation is less clear in this case in view of the following. Protein AAN56706 has a domain homologous to a domain of *E. coli* transcriptional activator IlvY, and the second site of the repeat found upstream of the AAN56706 gene is adjacent to the putative promoter (–10) box. On the other hand, the association of this protein with proline biosynthesis is unclear. In the other *Shewanella* species, homologs of the AAN56706 gene lack a site corresponding to the consensus in their leader regions (in *S. baltica* and *Shewanella* sp. PV-4) or are immediately downstream of other genes; i.e., a proper leader region is lacking (in *S. denitrificans* and *S. frigidimarina*).

When the consensus was truncated to the most conserved sequence MTACCAAY and the same leader regions were analyzed, sites close to the sequence were found only upstream of 80 genes in *S. oneidensis.* With the exception of the above two cases, not a single site detected by this less stringent search was repeated at least twice in a leader region and, what is more important, was 10–20 bp upstream of a sequence differing from TATAA in no more than one position. Thus, a genome-wide search in *S. oneidensis* did not reveal a new site similar to the sites shown in the figure. The multiple sequence alignment (figure) and the results of the total search of one of the *Shewanella* genomes suggest the significance of the predicted sites.

ACKNOWLEDGMENTS

We are grateful to M.S. Gelfand and A.G. Vitreshchak for helpful discussion and critical comments.

This work was supported by the International Science and Technology Center (project no. 2766).

REFERENCES

1. Krishna R.V., Beilstein P., Leisinger T. 1979. Biosynthesis of proline in *Pseudomonas aeruginosa*. Properties of γ-glutamyl phosphate reductase and 1-pyrroline-5-carboxylate reductase. *Biochem. J.* **181**, 223–230.
2. Chopin A. 1998. Analysis of the *Bacillus subtilis* genome sequence reveals nine new T-box leaders. *Mol. Microbiol.* **29**, 661–669.
3. Grundy F.J., Henkin T.M. 2003. The T box and S box transcription termination control systems. *Frontiers Biosci.* **8**, d20–d31.
4. Leontiev L.A., Seliverstov A.V., Lyubetsky V.A. 2005. Algorithm for a large-scale search for T-box transcription regulation in bacteria. *Mol. Biol.* **39**, 1076–1078.
5. Seliverstov A.V., Lyubetsky V.A. 2006. Algorithm for a search for conserved nucleotide sequences. *Inform. Protsessy.* **6**, 33–36.