ISSN 0096-3925, Moscow University Biological Sciences Bulletin, 2010, Vol. 65, No. 4, pp. 211–212. © Allerton Press, Inc., 2010.
Original Russian Text © K.V. Lopatovskaya, K.Yu. Gorbunov, L.Yu. Rusin, A.V. Seliverstov, V.A. Lyubetsky, 2010, published in Vestnik Moskovskogo Universiteta. Biologiya, 2010, No. 4, pp. 92–94.

The Evolution of Proline Synthesis Transcriptional Regulation in Gammaproteobacteria

K. V. Lopatovskaya, K. Yu. Gorbunov, L. Yu. Rusin, A. V. Seliverstov, and V. A. Lyubetsky

e-mail: kristina@iitp.ru

Received April 20, 2010

Abstract—We report a conserved motif in the upstream region of the *proA* and *proB* genes, which are widely found among γ-proteobacteria, particularly in the genera *Pseudomonas*, *Marinobacter*, and *Shewanella*. The conserved protein–DNA binding sites are 8 bp long. Some genes have double sites for cooperative factor binding. The phylogenetic profiles of binding sites and all protein factors were compared. We identify the tetR family protein, an ortholog of the NP_249058 protein from *P. aeruginosa* PAO1, as a transcription factor. Our algorithm was applied to construct the species tree with predicted regulation of the *pro* genes. The phylogeny of *pro* genes, transcription factor phylogeny, and a phylogeny of their binding sites were mapped onto the species tree with our algorithm. The obtained scenario displays important HGT and related events.

Keywords: proline synthesis, TetR protein family, transcription regulation, γ-proteobacteria, evolutionary scenario, horizontal gene transfer.

DOI: 10.3103/S0096392510040255

INTRODUCTION

The transcription of proline synthesis genes was experimentally shown to depend on proline concentration in *Pseudomonas aeruginosa* [1, 2]. The *ProA* and *proB* genes encode two enzymes (γ-glutamyl kinase and γ-glutamyl phosphate reductase). Preliminary analysis of a small number of representatives of *Pseudomonas* and *Shewanella* has revealed a conserved motif upstream of these genes [3]. We searched for this motif upstream of *proA* and *proB* in all proteobacteria.

EXPERIMENTAL

All data were taken from GenBank (www.ncbi.nlm.nih.gov). To find the binding factor associated with a given motif, we developed a program that compares the phylogenies of sites and all bacterial proteins [4, 5]. Briefly, the program operates as follows. To perform a comparison, a motif's phylogenetic profile and the profile of each protein are represented as vectors. The phylogenetic profile of the motif contains +1 in the *i*-th position if the appropriate site is present in the *i*-th genome or −1 if it is not. Instead of defining discrete presence or absence of the site in the *i*-th position, we can specify the site quality. Let the profile of each protein in the *i*-th position contain the sum of similarity indicies of corresponding amino acids (according to the BLOSUM62 matrix) from the alignment of this protein with the nearest homologue from the *i*-th genome. The measure of the *vectors similarity* is calculated as a cosine of the angle between the vectors. The quality of a 8 bp site was calculated as the *distance* between the consensus motif and site. For this purpose, the consensus and the site were represented as vectors in 32-dimensional space, and the distance between them was computed as the sum of magnitudes of their coordinate differences. The consensus vector contains four nucleotide frequencies in each of the eight site positions. Analogously, the frequencies in the site vector are one and three zeroes in each site position.

RESULTS AND DISCUSSION

We identified a conserved motif upstream of the *proA* and *proB* genes in many γ-proteobacteria; it is most widely represented in the genera *Pseudomonas*, *Marinobacter*, and *Shewanella*. (Extensive illustrative material is available on the Proline website of the Laboratory of mathematical methods and models in bioinformatics of the A.A. Kharkevich Institute for Information Transmission Problems of the RAS [4]). Figure 1 [4] shows a tree of species with predicted regulation based on protein-DNA interaction. *Pro* genes usually form the *proBA* operon, but in *Pseudomonas* they are transcribed independently; we did not find transcription regulation of *proB* outside the operon. The corresponding multiple alignment is shown in Fig. 2 [4]. Conserved protein–DNA binding sites are 8 bp long and usually duplicated at a distance from 10 to 16 bp. Similarly, a weakly conserved region adjoins the conserved 3'-end of the site in all studied species (Fig. 3 [4]).

With the developed methodology, we infer that the corresponding transcription factor is a protein from the TetR family, an ortholog to NP_249058 from *P. aeruginosa* PAO1 (see Table 2 [4]). This protein contains a hypothetical DNA-binding domain on it's N-terminus, which is typical for the TetR family (Pfam identifier PF00440, the 11–57 a.a. region has 1.5×10^{-13} value of expected difference from the consensus)[6]. Orthologs of this protein with a relatively high expected value are found in all species with predicted regulation; rare exceptions relate to poorly conserved sites. However, the protein has paralog NP_253746 in the same species, *P. aeruginosa*; close homologues of the protein and its paralog were found in *Vibrio* and other genera, including those that do not contain the motif we predicted (Table 2) [4].

In the latter case the homologues might be recognized by other factors. The algorithm described in [7] was used to reconcile species tree S with tree G of the *pro* genes (Fig. 4 [4]), tree G of transcription factor domains (Fig. 5 [4]), and tree F of factor binding sites; the corresponding forest is shown in Fig. 6 [4]. The length of the site and the promoter region is about 52 bp, which makes them more evolutionary labile compared to their corresponding factors and regulated genes. In our model, a gain is explained by a transfer from outgroup. A scenario of co-evolution of genes, regulatory factors and their binding sites is given in Fig. 1, 4–6 [4]. In this scenario, the most important are horizontal transfers and related events. The concept of the "tube" was introduced in [8] to represent an edge of the species tree that contains evolutionary events. We detect five gene transfers, six factor and two conserved site transfers (the small length of sites and their neighborhood containing the promoter often hampers distinguishing between the site gain and transfer).

Using original methodology we describe the co-evolution of *pro* genes, their regulatory factors and binding sites along the species tree. To the authors' knowledge, studies of the component co-evolution of the protein–DNA interaction based regulatory system are not presented elsewhere.

REFERENCES

1. Krishna, R.V., Beilstein, P., and Leisinger, T., Biosynthesis of Proline in *Pseudomonas aeruginosa*. Partial Purification and Characterization of γ-Glutamyl Kinase, *Biochem. J.*, 1979, vol. 181, pp. 215–222.
2. Krishna, R.V., Beilstein, P., and Leisinger, T., Biosynthesis of Proline in *Pseudomonas aeruginosa*. Properties of gamma-Glutamyl Phosphate Reductase and 1-Pyrroline-5-Carboxylate Reductase, *Biochem. J.*, 1979, vol. 181, pp. 223–230.
3. Seliverstov, A.V. and Lyubetsky, V.A., Regulation of Proline Biosynthesis in Proteobacteria, *Mol. Biol.*, 2007, vol. 41, no. 3, pp. 572–574 [*Mol. Biol.* (Engl. Transl.), 2007, vol. 41, no. 3, pp. 513–515].
4. Official Site Proline (URL: http://lab6.iitp.ru/ru/proline/ 10.2.10).
5. Leont'ev, L.A. and Lyubetskii, V.A., Algorithm for Protein Determination in Accordance with Specified Phylogenetic Profile, *Inform. Prots. (www.jip.ru)*, 2006, vol. 6, no. 1, pp. 24–32.
6. Kisker, C., Hinrichs, W., Tovar, K., and Hillen, W., The Complex Formed between Tet Repressor and Tetracycline–Mg^{2+} Reveals Mechanism of Antibiotic Resistance, *J. Mol. Biol.*, 1995, vol. 247, no. 2, pp. 260–280.
7. Gorbunov, K.Yu. and Lyubetsky, V.A., Reconstructing the Evolution of Genes along the Species Tree, *Mol. Biol.,* 2009, vol. 43, no. 5, pp. 946–958 [*Mol. Biol.* (Engl. Transl.), 2009, vol. 43, no. 5, pp. 881–893].