FEBS Letters 586 (2012) 3843–3851

journal homepage: www.FEBSLetters.org

Lateral gene transfer between the *Bacteroidetes* and *Acidobacteria*: The case of α-L-rhamnosidases

Daniil G. Naumoff*, Svetlana N. Dedysh

Winogradsky Institute of Microbiology, Russian Academy of Sciences, Moscow, Russia

ARTICLE INFO

Article history:
Received 15 July 2012
Revised 4 September 2012
Accepted 6 September 2012
Available online 26 September 2012

Edited by Takashi Gojobori

Keywords:
Glycoside hydrolase
α-L-Rhamnosidase
GH78 family
GH106 family
Lateral gene transfer
Acidobacteria

ABSTRACT

α-L-Rhamnosidases catalyze the hydrolysis of the terminal α-L-rhamnose residues in various carbohydrates. The catalytic domains in most of these enzymes belong to the families GH78 and GH106 of glycoside hydrolases. In this study, we show that almost all genes encoding the GH78- and GH106-containing proteins from members of the poorly characterized bacterial phylum *Acidobacteria* originated from precursors belonging to the phylum *Bacteroidetes*. Members of the *Acidobacteria* and *Bacteroidetes* display similar functional capabilities and specialize on degradation of plant-derived organic matter. Several proposed lateral gene transfers between the *Acidobacteria* and *Bacteroidetes* occurred presumably during specialization of these bacteria for their environments.

1. Introduction

Glycoside hydrolases or glycosidases [EC 3.2.1] are a widespread group of enzymes, which hydrolyze the glycosidic bonds between two carbohydrates or between a carbohydrate and an aglycone moiety. Glycosidase genes are found in the vast majority of the living organisms and in some viruses. About 1% of all known genes encode glycosidases and their close homologues. The assortment of these genes in a genome varies significantly depending on the taxonomic position of the organism and its ecological niche. Glycosidase genes often undergo duplications, elimination, and lateral transfer [1]. Based on the sequence similarity of the catalytic domains, glycoside hydrolases are grouped into 125 families (GH1–GH130, except for GH21, GH40, GH41, GH60, and GH69) according to the Carbohydrate-Active Enzymes (CAZy) classification (http://www.cazy.org/; [2]). Glycosidases with the same type of enzymatic activity can often be found in several protein families and sometimes can display different tertiary structure of the catalytic domain (i.e. to be evolutionarily unrelated).

α-L-Rhamnosidases or α-L-rhamnoside rhamnohydrolases [EC 3.2.1.40] are glycoside hydrolases, which catalyze the hydrolysis of terminal, non-reducing α-L-rhamnose residues in α-L-rhamnosides including glycolipids and glycosides, such as plant pigments, flavonoid glycosides, pectic polysaccharides, and gums [3–7]. The enzyme has wide occurrence in nature and has been reported from animals, plants, yeasts, fungi, and bacteria [4–7].

Almost all currently known α-L-rhamnosidases belong to the family GH78 of glycoside hydrolases. According to the CAZy database, this family contains 470 proteins, including 13 proteins from the Archaea, 394 proteins from the Bacteria, and 59 proteins from members of the Eukaryota (Table 1 and Supplementary table). Eighteen of these proteins are characterized biochemically and display the same type of enzymatic activity.

α-L-Rhamnosidase from *Sphingomonas paucimobilis* is the only experimentally characterized representative of the family GH106, which comprises 86 bacterial proteins according to the CAZy database (Table 1 and Supplementary table). Only the families GH78 and GH106 of glycoside hydrolases are composed exclusively of enzymes with the α-L-rhamnosidase activity. Notably, the most abundant groups of proteins in both families (108 and 57 proteins in GH78 and GH106, respectively) originate from bacteria of the phylum *Bacteroidetes* (according to CAZy; Table 1). This is not surprising since members of this phylum are known as efficient degraders of various macromolecules including complex polysaccharides [8]. Genomes of these bacteria contain a significant number of genes encoding carbohydrate-processing enzymes [8–11].

Our attention, however, was focused on α-L-rhamnosidases from members of the poorly characterized bacterial phylum *Acidobacteria*. The metabolic potential and the functional role of

* Corresponding author. Address: Winogradsky Institute of Microbiology, Russian Academy of Sciences, Prospect 60-letya Octyabrya 7/2, Moscow 117312, Russia. Fax: +7 499 1356530.
E-mail address: daniil_naumoff@yahoo.com (D.G. Naumoff).

0014-5793/$36.00
http://dx.doi.org/10.1016/j.febslet.2012.09.005

Table 1
Taxonomic distribution of GH78 and GH106 proteins according to the CAZy database.

Taxon	GH78	GH106	Total
Acidobacteria	9	10	19
Actinobacteria	93	10	103
Archaeae	13	0	13
Bacteroidetes	108	57	165
Chloroflexi	11	0	11
Firmicutes	123	0	123
Fungi	58	0	58
Planctomycetes	6	0	6
Proteobacteria	28	5	33
Spirochaetes	5	0	5
Verrucomicrobia	3	2	5
Others	13	2	15
Total	470	86	556

Note. Taxonomic distribution of GH78 and GH106 proteins listed at the CAZy database [2] on August 29, 2012 is indicated. For more information see table in the supplements.

these bacteria in nature remain largely unexplored. Currently, this phylum is represented by only a handful of isolates most of which are versatile heterotrophs. The genomes of these organisms contain a large number of glycosidase genes, placing them in the top 5% of all bacterial genomes surveyed in the CAZy database [12]. In the *Acidobacteria,* families GH78 and GH106 are represented by 18 proteins [13], which are encoded in five completely sequenced genomes of *Acidobacterium capsulatum* ATCC 51196, *Granulicella mallensis* MP5ACTX8, *Granulicella tundricola* MP5ACTX9, *Candidatus* Solibacter usitatus Ellin6076, and *Terriglobus saanensis* SP1PR4. As revealed by our analysis, the majority of the closest homologues of these acidobacterial proteins belong to members of the *Bacteroidetes*. Here, we describe our findings and suggest a lateral gene transfer as an explanation for an unexpectedly high level of sequence similarity of the hypothetical α-L-rhamnosidases from the *Acidobacteria* and *Bacteroidetes*.

Preliminary results of this study were reported at the 8th International Conference on Bioinformatics of Genome Regulation and Structure\Systems Biology [14].

2. Methods and algorithms

Searches in the non-redundant protein database were performed on March 15, 2012 in order to obtain the updated list of proteins containing GH78 or GH106 domains. 25 and 69 isolated catalytic domains of the most divergent representatives of GH78 and GH106 families, respectively, were used as the queries (see list of the proteins in the Supplementary file). The exact borders of GH78 domain were determined based on the experimental data on the 3D structure for two family members (PDB, 2OKX and 3CIH). Approximate borders of the GH106 domain were deduced based on homology to GH42 domain (PDB, 1KWG). In particular, protein fragments from 397th to 722nd amino acid residues and from 37th to 454th residues were used in the case of hypothetical α-L-rhamnosidases from *Acidobacterium capsulatum* with GenPept accession numbers ACO31354.1 (family GH78) and ACO34217.1 (GH106), respectively.

The results of blast searches with different queries were summarized using the PSI Protein Classifier program [15]. Detection of a protein using the respective domain of a known family as a query after the first PSI-BLAST iteration with *E*-value $\leqslant 10^{-7}$ was the criterion used to classify this protein as the family member [15–17].

Phylogenetic analysis of the families GH78 and GH106 was performed using proteins selected based on the sequence similarity. The closest homologues (according to *E*-value) of the acidobacterial proteins were preferentially chosen but only one representative of a genus was usually used if several highly similar proteins were available. As a rule, incomplete sequences were removed from the analysis (for example, AEK44383.1, EEC98578.1, EEG21220.1, and EHP36561.1). In total, 701 proteins (358 and 343 representatives from GH78 and GH106 families, respectively) were used for building the multiple sequence alignments and the subsequent phylogenetic analysis (including 21 proteins from the phylum *Acidobacteria*). The protein regions corresponding to PF05592 module (see the Supplementary file) and those homologous to the full-sized β-galactosidase from *Thermus thermophilus* (PDB, 1KWG) were used for phylogenetic reconstruction of the GH78 and GH106 families, respectively.

Multiple sequence alignments were produced manually with the BioEdit program (http://www.mbio.ncsu.edu/BioEdit/bioedit.html). If long deletions were revealed in selected sequences or some fragments showed no similarity to other sequences, the corresponding nucleotide sequences were analyzed in the three reading frames. In the case of some eukaryotic sequences (GenPept: EDQ48560.1, EGR44690.1, EGU84335.1, and EHL01716.1), the exon–intron structures of the genes were refined to increase the similarity as far as possible. Position of the start and stop codons were changed in several sequences in order to continue the region of homology. In several sequences, frameshifts and stop codons in the frame were ignored. GenPept accession numbers of the edited amino acid sequences are listed in the Supplementary file and are shown in figures in lowercase letters.

After omission of the most variable sequence fragments, phylogenetic trees were constructed with the PROTPARS program (Protein Sequence Parsimony Method, MP) and the NEIGHBOR program (Neighbor-Joining Method, NJ) from the PHYLIP package (http://evolution.gs.washington.edu/phylip.html). The confidence limits for each node obtained in both kinds of trees were estimated by 100 bootstrap replicates.

3. Results

3.1. Protein collection

Our search for GH78- and GH106-related sequences yielded 1222 and 486 proteins, respectively. These sequence pools far exceed those indicated at the CAZy site (470 proteins from the family GH78 and 86 proteins from the family GH106). In total, 9 and 12 proteins from the phylum *Acidobacteria* were found in the families GH78 and GH106, respectively. Domain structure of these proteins is described in the Supplementary file.

3.2. Phylogenetic analysis of the GH78 domains

The multiple sequence alignment of 358 GH78-containing proteins (including 9 proteins from the phylum *Acidobacteria*, shown in Fig. 1) was used for the phylogenetic reconstruction of the family evolution. We obtained the neighbor-joining (NJ; Fig. 2) and the maximum parsimony (MP; data not presented) trees. Both trees display a similar topology suggesting the correct interpretation of the evolutionary events. Five acidobacterial proteins compose two stable clusters (100% of bootstrap support on NJ- and MP-trees) and the other four proteins are spread over the trees. Therefore, the acidobacterial proteins belong to six different clusters, which are discussed below.

The cluster A (Fig. 2A) contains one of the proteins from *Solibacter usitatus* (GenPept, ABJ84713.1), which is arranged among proteins from various representatives of the phylum *Bacteroidetes* (Fig. 2B). The other *Solibacter usitatus* protein (ABJ86506.1) composes the sister group for a large cluster of proteins from the