

SEQUENCE ANALYSIS AND CLASSIFICATION OF α -GALACTOSIDASES

Daniil G. Naumoff

State Institute for Genetics and Selection of Industrial Microorganisms, Moscow 113545, Russia

E-mail: daniil_naumoff@yahoo.com

Poster prize (<35): Yes

Glycosyl hydrolases are a widespread group of enzymes that hydrolyze the glycosidic bonds between two carbohydrate residues. Currently, several thousand sequences of the proteins are known. They are grouped into 83 families on the basis of sequence similarity (<http://afmb.cnrs-mrs.fr/CAZY/>). Four of these families (GH4, GH27, GH36, and GH57) include enzymes with α -galactosidase activity [E.C. 3.2.1.22].

GH4 and GH57 families encompass several types of glycosyl hydrolases. The α -galactosidase activity has been demonstrated only for three enzymes of GH4 family (from *Escherichia coli*, *Thermotoga maritima*, and *T. neapolitana*) and two enzymes of GH57 family (from *Pyrococcus furiosus* and *Thermococcus alcaliphilus*).

The majority of known α -galactosidases belong to GH27 and GH36 families, which compose a superfamily (clan GH-D). We have performed sequence analysis of the two families. Most of GH27 sequences have a high level of sequence similarity (more than 30% of identical residues) and make up a distinct subfamily. Four main subgroups can be distinguished in this subfamily (> 50% of identical residues). One of them includes plant and *Clostridium josui* α -galactosidases and a hypothetical protein (ORF) from *Streptomyces coelicolor*. Other two subgroups comprise glycosidases from Ascomycota yeast and Vertebrata respectively. Some fungal α -galactosidases form the fourth subgroup. One of the *Trichoderma reesei* α -galactosidases, *Arthrobacter globiformis* isomaltodextranase, and *Bacillus halodurans* hypothetical protein along with two *Arabidopsis thaliana* hypothetical proteins are the most divergent representatives of GH27 family and can be considered the only known representatives of three other subfamilies.

GH36 family is composed of two main subfamilies. Subfamily 36a includes α -galactosidases from Gram-positive bacteria, *Absidia corymbifera*, *Aspergillus niger*, *Escherichia coli*, and *Trichoderma reesei* as well as a hypothetical protein from *Yersinia pestis*. Subfamily 36b contains enzymes from Gram-negative bacteria (Proteobacteria, Thermotogales, and Thermus) and one hypothetical protein from *Streptomyces coelicolor*. Another hypothetical protein from *St. coelicolor* can be thought of as the only representative of the third subfamily.

We will present data about homology of α -galactosidases from GH27 and GH36 families with glycosidases of GH31 family and proteins of α GalT family. The α GalT family is comprised of α -galactosyltransferases and seed imbibition proteins from higher plants, *Bifidobacterium breve* α -galactosidase, and hypothetical proteins from *Sulfolobus solfataricus* and *Sul. tokodaii*. We propose to include GH31 and α GalT families into the α -galactosidase superfamily, in addition to GH27 and GH36 families.