# Bioinformatics approaches to studying the relationships between structure, function and evolution of amylolytic enzymes – the case of the alpha-amylase family GH57

Janeček Štefan1,2

1 Laboratory of Protein Evolution, Institute of Molecular Biology, Slovak Academy of Sciences, Dubravska cesta 21, SK-84551 Bratislava, Slovakia; e-mail: [Stefan.Janecek@savba.sk](mailto:Stefan.Janecek@savba.sk)

2 Institute of Biology and Biotechnology, Faculty of Natural Sciences, University of SS Cyril and Methodius, Nam. J. Herdu 2, SK-91701 Trnava, Slovakia

In the sequence-based classification of carbohydrate-active enzymes, the CAZy database (<https://www.cazy.org/>), four glycoside hydrolase (GH) families – GH13, GH57, GH119 and GH126 – are known as the alpha-amylase families. The family GH57, created in 1996 and named as the second alpha-amylase family has recently been classified within the clan GH-T with family GH119. It counts more than 5,800 sequences from Bacteria and Archaea, but only 44 its members have already been experimentally characterized as various amylolytic enzymes. The catalytic domain adopts an incomplete TIM-barrel succeeded by a bundle of alpha-helices with the catalytic machinery formed by a glutamic acid and aspartic acid at the strands, respectively, beta-4 and beta-7 of the barrel domain. The first in silico analysis of the family GH57 was performed in 2004 when the five conserved sequence regions (CSRs) characteristic for the family were defined. Later, in 2012, the five well-established family GH57 CSRs were described as the so-called “sequence fingerprints” containing the unique sequence features typical for the individual enzyme specificities of the family. Subsequent 2018 bioinformatics study of ~1,600 GH57 members delivered clusters in the phylogenetic tree reflecting the eight distinct enzyme specificities. Finally, in 2025, based on a detailed in silico analysis of a dataset of ~5,000 sequences, the family GH57 has been officially divided into ten subfamilies. Importantly, each GH57 subfamily can be characterized by its sequence fingerprints, i.e. the logo of the five GH57 CSRs.