REFINING AND MINING THE PHYLOGENY OF GLYCOSIDE HYDROLASE FAMILY 74 VIA STRUCTURE-FUNCTION ANALYSIS

G. Arnal, University of British Columbia, Canada
J. Asohan, University of British Columbia, Canada
B. Henrissat, Aix-Marseille Université, France
P. Stogios, University of Toronto, Canada
T. Skarina, University of Toronto, Canada
A. Savchenko, University of Toronto, Canada
H. Brumer, University of British Columbia, Canada

Sustained interest in the use of carbohydrates from plant cell walls, coupled with the advancement of high-throughput (meta)genomic sequencing, has led to the discovery of an overwhelming number of predicted carbohydrate-active enzymes (CAZymes) in the last decade. The CAZy database provides a powerful framework for the study of CAZymes, including Glycoside Hydrolases (GHs), by enabling the prediction of key enzyme features such as 3-D fold, catalytic residues, catalytic mechanism, and – with certain limitations – substrate specificity. Refined phylogenetic analyses contribute to increasing the accuracy of predictions by further clustering proteins into sub-families (1, 2). However, reliable prediction of substrate specificity for newly discovered GHs remains a challenge due to a general lack of in-depth biochemical and structural characterization across the existing phylogenetic diversity.

Glycoside Hydrolase family 74 (GH74) comprises endo-glucanases, many of which have predominant activity toward xyloglucan, a highly branched plant cell wall matrix glycan. To better delineate overall substrate specificity, backbone cleavage position, and endo-dissociative vs. endo-processive hydrolytic modes, a broad-based structure-function analysis of GH74 guided by molecular phylogeny was performed. Seven sub-families were discerned, which grouped nearly 40% of the current >300 GH74 sequences in the public CAZy database. Thirty one GH74 members were targeted for further investigation based on their phylogenetic position and unique primary structural features identified during manual curation. The biochemical characterization of 18 recombinant GH74s revealed key sequence features governing xyloglucan backbone cleavage sites and highlighted clear phylogenetic differences between endo-dissociative and endo-processive enzymes. Commensurate with previous studies (3), site-directed mutagenesis of key active-site tryptophan residues defined their essential contributions to processivity on the soluble polysaccharide substrate. Six new GH74 tertiary structures (apo and/or in complex with xylogluco-oligosaccharides) were determined that further resolved the contribution of active-site loops in modulating the size of oligosaccharide products released by individual subfamily members. Refining the correlation between phylogeny and enzyme structure-function properties in GH74 significantly enhances the prediction of catalytic ability, highlights key steps in the evolution of function in the family, and ultimately informs applications in biomass conversion.

1. Stam MR, Danchin EGJ, Rancurel C, Coutinho PM, Henrissat B. 2006. Dividing the large glycoside hydrolase family 13 into subfamilies: towards improved functional annotations of alpha-amylase-related proteins. Protein Engineering Design & Selection 19:555-562.
2. Aspeborg H, Coutinho PM, Wang Y, Brumer H, Henrissat B. 2012. Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). Bmc Evolutionary Biology 12.
3. Matsuzawa T, Saito Y, Yaoi K. 2014. Key amino acid residues for the endo-processive activity of GH74 xyloglucanase. FEBS Lett 588:1731-8.