Ten years of *CAZypedia*: A living encyclopedia of carbohydrate-active enzymes

The *CAZypedia* Consortium*

*A list of contributors at the time of publication is provided in the Acknowledgements. All past and future *CAZypedia* Editors and Authors are invited to cite this article in reference to their invaluable contributions to this community resource.

Dedication: CAZypedia is dedicated to Emeritus Professor Bruce Stone (1928-2008†), whose enthusiasm to create a comprehensive encyclopedia of carbohydrate-active enzymes was essential to the genesis of this resource.

Abstract

CAZypedia was initiated in 2007 to create a comprehensive, living encyclopedia of the carbohydrate-active enzymes (CAZymes) and associated carbohydrate-binding modules involved in the synthesis, modification, and degradation of complex carbohydrates. *CAZypedia* is closely connected with the actively-curated CAZy database, which provides a sequence-based foundation for the biochemical, mechanistic, and structural characterization of these diverse proteins. Now celebrating its 10th anniversary online, *CAZypedia* is a successful example of dynamic, community-driven, and expert-based biocuration. *CAZypedia* is an open-access resource available at URL http://www.cazypedia.org.

Background

The Carbohydrate-Active Enzymes (CAZymes) classification groups catalytic and substrate-binding modules of proteins responsible for the assembly and breakdown of complex carbohydrates into sequence-based families. Since the original definition of 35 glycoside hydrolase (GH) families in 1991 (Henrissat, B. 1991), the CAZy database[1] continues to grow and currently (August 2017) encompasses 104 glycosyltransferase (GT) families, 145 GH families, 27 polysaccharide lyase (PL) families, 16 carbohydrate esterase (CE) families, 13 auxiliary activity (AA) families, and 81 carbohydrate binding module (CBM) families (Levasseur, A., Drula, E., et al. 2013, Lombard, V., Ramulu, H.G., et al. 2014). As a result of vigorous biocuration [as defined by Bourne and McEntyre (Bourne, P.E. and McEntyre, J. 2006)] and tireless technical development in response to an ever-increasing rate of gene sequencing, the CAZy database has become the *de facto* framework that unites protein sequence, biochemical, and structural data among the tremendous diversity of CAZymes in nature [see (Davies, G.J. and Sinnott, M.L. 2008) for an accessible primer and review].

The CAZy database is arranged in a conventional format, with individual family pages consisting of tables of protein names, GenBank and/or UniProt sequence accession codes, EC numbers (when activity has been experimentally defined), and Protein Data Bank accession codes (when a structure has been solved). Each family page contains a compact header that summarizes key information on substrate specificity, catalytic mechanism, three-dimensional protein fold, and carbohydrate ligand complexes.

[1] Available at URL http://www.cazy.org/

Additionally, individual genome pages provide a convenient census of all CAZyme families in individual organisms (Lombard, V., Ramulu, H.G., et al. 2014). In keeping with its primary function to list individual family members, family pages in the CAZy Database are efficiently minimalistic. *CAZypedia* arose from the idea that a more detailed and directly accessible summary of the key research on individual CAZy families would be of significant value to glycoscience researchers, particularly highlighting the primacy of key research discoveries in a family, and supporting the activities of all scientists interested in CAZymes.

Genesis

CAZypedia's roots can be traced to renowned polysaccharide biochemist Professor Bruce Stone (1928-2008† (Whelan, W. 2009)) who proposed the idea of a comprehensive encyclopedia of the CAZymes. Bruce initially raised this idea informally at the 23rd International Carbohydrate Symposium (ICS; Whistler, Canada; July 2006) among a select group of glycoscientists, including Harry Brumer, Anthony Clarke, Gideon Davies, Harry Gilbert, Bernard Henrissat, Antoni Planas, Birte Svensson, David Vocadlo, Spencer Williams, Stephen Withers, and others. Bruce's original vision was to produce a traditional printed book or series, comprising chapters written by specific experts on individual families. It was recognized early-on that that the sheer number of families at that time (>100 GH families alone), combined with rapid advancements in the field, would make the timely completion of a printed work with lasting value a Sisyphean task.

Further *ad hoc* discussions about the best way to bring Bruce's vision to fruition continued through subsequent months, culminating at a second, larger group discussion at the 7th Carbohydrate Bioengineering Meeting (CBM7; Braunschweig, Germany; April 2007). Among those in attendance were (again) Bruce Stone, Harry Brumer, Anthony Clarke, Harry Gilbert, Antoni Planas, and Birte Svensson, as well as Vincent Bulone, Marco Moracci, Warren Wakarchuk, Tony Warren, Lisa Willis, and others. Here, there was general agreement that only an online, internet-based format would have sufficient flexibility and immediacy to match the rapid advances being made in CAZymology. Inspired by the growing impact of *Wikipedia* as a community-based publishing model of encyclopedic information, the idea to use a *wiki*[2] approach to develop an online "Encyclopedia of Carbohydrate-Active Enzymes" was adopted. Hence, *CAZypedia* was born in May 2007 when Harry Brumer, then of the Kungliga Tekniska Högskolan in Stockholm, established *CAZypedia* using the MediaWiki software.[3]

Content

Content creation for *CAZypedia* was focussed initially on the GH Families, due to a particularly long and rich history of biochemical and structural characterization of these enzymes (Davies, G. and Henrissat, B. 1995, Sinnott, M.L. 1990). An original set of pages covering families GH1, GH2, GH10, and GH11 by Stephen Withers, together with GH27 and GH36 by Harry Brumer, were produced and refined with editorial input from Bernard Henrissat through the summer of 2007. In this process, a streamlined page format was devised (**Figure 1**), comprising individual sections on "Substrate specificities", "Kinetics and mechanism", "Catalytic residues", and "Three-dimensional structures", which present a concise

---

[2] See definition at URL https://en.wikipedia.org/wiki/Wiki
[3] Freely available at URL https://www.mediawiki.org/

summary of common features of each family. A "Family Firsts" section provides a brief, itemized list of references to seminal publications that define the key mechanistic and structural features of the family: the first reaction stereochemistry determination, catalytic residue identification, and three-dimensional structure solution. An overarching goal in page design was to provide a rapid entry into the key primary literature on each family (which is not directly available in the CAZy Database), through an abbreviated and consistent format. *CAZypedia* pages may be beneficially embellished with figures, although this is optional.

As part of an explicit design intent, *CAZypedia* pages do not necessarily strive to provide comprehensive reviews of all the available literature on individual families, although it should be noted that there is formally no prescribed page length. The reasons for this are largely practical. Initially, pages can be composed rapidly by focussing on the key defining literature. Compilation of a comprehensive corpus of the published work on a family, which is in many cases extensive when all individual biochemical characterization studies are considered, is therefore not required. This focus also helps to future-proof pages in a rapidly evolving field: first achievements will always remain historically significant, regardless of the number of subsequent publications on a family. For the same reason, pages explicitly avoid enumeration of time-sensitive data, such as the number of sequences or structures for individual families, which can otherwise be gleaned from the continually updated CAZy database (individual *CAZypedia* and CAZy database pages are cross-linked for this purpose). Thus, *CAZypedia* pages are designed to be perpetually accurate, regardless of the frequency of future updates from page authors. The appellation 'Curator Approved' is given to each newly minted family page once all sections contain a basic coverage of the seminal literature (see also "Technical aspects" section below).

From the initial seed of six GH families, *CAZypedia* has grown to include over 100 individual Curator Approved GH family pages, produced by a similar number of expert contributors from the CAZyme/glycoscience community. Indeed, July 2014 marked a watershed in CAZypedia's history, with the completion of the Glycoside Hydrolase Family 12 page by Gerlind Sulzenbacher as the 100th Curator Approved GH page. Pages on other groups of CAZymes (i.e., Glycosyltransferases (Coutinho, P.M., Deleury, E., et al. 2003), Polysaccharide Lyases (Lombard, V., Bernard, T., et al. 2010), and Auxiliary Activity redox enzymes (Levasseur, A., Drula, E., et al. 2013)) and non-catalytic Carbohydrate Binding Modules (Boraston, A.B., Bolam, D.N., et al. 2004) continue to be incorporated through growing community engagement. Notable *CAZypedia* firsts include the completion of the GT42 page by Warren Wakarchuk in April 2010, the PL2 page by Wade Abbott in September 2013, the AA9 lytic polysaccharide mono-oxygenase page by Paul Harris in September 2013, and the CBM32 page by Elizabeth Ficko-Blean and Alisdair Boraston in May 2013. *CAZypedia's* History page[4] serves as a repository for these and future major milestones, while the News page[5] covers recent Curator Approved pages and other newsworthy items.

---

[4] Available at URL https://www.cazypedia.org/index.php/CAZypedia:History, accessed via the 'About *CAZypedia*' menu.
[5] Available at URL https://www.cazypedia.org/index.php/News and via *CAZypedia's* Main Page.

In recognition of the complex nature of carbohydrate chemistry and CAZymes, *Cazypedia* also incorporates a Lexicon that provides a definition of key terms, explanation of specialist nomenclature, and tutorial reviews of concepts that are relevant to individual family pages. The Lexicon provides a touchstone for new readers to support their understanding and interpretation of individual families, and is linked using hyperlinks from within the text of family pages. The Lexicon and category pages for each major CAZyme class are conveniently accessed under the Content menu, prominently displayed on the left side of all *CAZypedia* pages (**Figure 1**).

At its 10th anniversary online, *CAZypedia* currently comprises 106 GH, 10 CBM, 6 PL, 2 AA, 2 GT, and 22 Lexicon pages with Curator Approved status. The MediaWiki software upon which *CAZypedia* relies tracks usage statistics, which are available through the Special Pages menu item. These statistics reveal over 12 million total page views, and over one hundred thousand views for several of the most popular GH and Lexicon pages. More conservative estimates of activity provided by Google Analytics indicate that *CAZypedia* access has increased to thousands of international users per week since data recording on that utility began in the autumn of 2009 (**Figure 2**). Regardless of the absolute values, these data highlight the sustained and growing value of *CAZypedia* to specialists and non-specialists alike.

Editorial framework

During the birth of *CAZypedia*, there was significant concern about the potential pitfalls of applying directly the *Wikipedia* model, which allows author anonymity and lacks formal editorial oversight, to the publication of a rigorous scientific encyclopedia. Thus, although *CAZypedia* adopts many of the general principles and rules of *Wikipedia*, *CAZypedia* draws on best-practice authoring and editing principles of peer-reviewed, wiki-based encyclopedia such as *Citizendium*[6] and *Scholarpedia*[7]. *CAZypedia* strives to be a dynamic, community-based resource, which at the same time balances the need for careful content curation. A full description of *CAZypedia's* editorial policies is available on the "About" page[8]; however, a few points deserve special comment.

The editorial organization of *CAZypedia* is designed with a minimum of bureaucratic and administrative overhead, because it is entirely volunteer-based and has no direct funding support. *CAZypedia* generally adopts *Wikipedia's* "Simplified Ruleset"[9], particularly the concepts of using a *neutral point-of-view*, writing *verifiable text*, including *only peer-reviewed information* (no original research), being *civil and well-behaved*, and *not infringing copyright*. As a culmination of these principles, *CAZypedia* reports on – but does not engage in critique of – the published literature, and supports all statements of fact with primary citations. Not least, *Wikipedia's* extensive "What *Wikipedia* is not" page[10] can be translated to "What *CAZypedia* is not" essentially point-by-point.

---

[6] Available at URL http://en.citizendium.org/
[7] Available at URL http://www.scholarpedia.org
[8] See URL https://www.cazypedia.org/index.php/CAZypedia:About, accessed via 'Introduction to *CAZypedia*' under the 'About *CAZypedia*' menu
[9] Available at URL http://en.wikipedia.org/wiki/Wikipedia:Simplified_Ruleset
[10] Available at URL http://en.wikipedia.org/wiki/Wikipedia:What_Wikipedia_is_not

Following the *Citizendium* model, transparency is achieved through the use of contributors' real names in *CAZypedia*. Additionally, individual biographical pages enable readers to evaluate directly each contributor's expertise in the field. To maintain editorial quality control, every Family and Lexicon page in *CAZypedia* is overseen by a *Responsible Curator*, who is primarily responsible for overall content. Responsible Curators are selected by a panel of *Senior Curators* based on established expertise and a willingness to participate in the active maintenance of specific pages. In turn, Responsible Curators are tasked with recruiting and managing *Authors* to participate in content creation; Responsible Curators may also contribute directly to composing page content.

In the spirit of a community-driven resource, individuals are encouraged to self-nominate to become Responsible Curators or Authors. In general, individuals at any career stage are welcomed to participate as Authors, including keen undergraduates, post-graduate students, and post-doctoral scientists. Indeed, the current list of contributors (see below) includes many junior scientists (or scientists who were at least junior at the time of their first contribution). Ultimately, the quality of entries in *CAZypedia*, like *Wikipedia*, relies upon the keen eye of readers at-large to identify errors and omissions. All users who spot such oversights are encouraged to contact the Responsible Curator for that page, so that a correction can be made.

CAZypedia is an open access publication, *i.e.* it is freely available online for anyone to read, study, and otherwise use for scholarly pursuits. However, the Authors and Curators of *CAZypedia* assert their copyright for the sole purpose of preventing outright duplication and uncontrolled modification of the content, which could undermine the expert-based nature of this resource. Although we strongly advocate that readers should cite the primary research literature directly, individual *CAZypedia* pages may also be cited when practical, analogous to a book chapter or review article. Citation details are provided in the footer and via the "Tools" menu on each page (**Figure 1**).

Technical aspects

Wiki-wiki

As introduced above, *CAZypedia* runs on MediaWiki, the free, open source PHP software originally developed for *Wikipedia.* This choice was based on the demonstrated robustness and scalability of MediaWiki, as well as the availability of diverse software *Extensions* to add functionality. As Wikipedia is unlikely to disappear anytime soon, so too is MediaWiki's active community of developers likely to persist well into the future, thereby ensuring continued maintenance of the software running *CAZypedia*. A full technical and functional description of MediaWiki is beyond the scope of this article; interested readers should visit MediaWiki.org for more details.

For the content contributor and user, the most important practical aspect of the use of MediaWiki is that *CAZypedia* is a *wiki*: edits are displayed instantaneously when saved and do not require approval before appearing online. This enables dynamic development of page content driven by individual Authors. In the initial stages of development, pages are clearly marked as "Under Construction", with a warning that content is under revision and may be subject to major changes. Once vetted by the Responsible Curator, a page may be upgraded to "Curator Approved" status to indicate that it is factually accurate and essentially complete. However, "completeness" is not absolute: as a wiki, *CAZypedia* is a living document, so further development of page content is forever possible.

Creating content for *CAZypedia* is relatively intuitive. Once a new Author has been provided with a login, page editing can be conducted within a modern web browser using a simplified markup language. A boilerplate pre-populates the page with the major template features, and Authors can view the code of other pages to get ideas of ways to insert features like hyperlinks, references, and figures. A "Getting Started Guide", along with concise pages that provide help with editing, references, and adding images provide guidance to assist the novice. Here, too, the use of MediaWiki as software platform is a considerable benefit, due to vast extant help resources on editing. Finally, assistance is always at hand from *CAZypedia* Curators, who are able to activity monitor edits via the global "Recent Changes" and individual "History" pages.

BiblioPlus

MediaWiki functionality can be enhanced through Extensions, and *CAZypedia* utilizes several, including those for user administration, defining page boilerplate content, and integrating Google Analytics. Among these, BiblioPlus[11] deserves special mention as the MediaWiki extension that drives bibliographic referencing. BiblioPlus is the result of a significant effort by *CAZypedia* contributor Karen Eddy to correct compatibility issues arising in the original Biblio extension by Martin Jambon and others.[12]

Like its predecessor, BiblioPlus performs automated retrieval and formatting of citations from PubMed and the ISBN databases in MediaWiki pages. Similar to other reference formatting software, BiblioPlus automatically numbers in-text citations and generates a reference section, which is included at the bottom of a page. Notably, the reference section contains hyperlinks to original sources, specifically PubMed or the ISBNdb, HubMed, and DOI hyperlinks. BiblioPlus was specifically re-coded to utilize the modern NCBI Entrez Programming Utilities (E-utilities) interface (Anonymous 2010). A full description of features and usage instructions is available on the BiblioPlus Mediawiki Extension page.[11] It should be noted that BiblioPlus is freely available and will work together with any modern MediaWiki implementation, so that it may be broadly deployed in any wiki, scientific or otherwise.

The next 10 years: *CAZypedia* needs you!

The continued success of *CAZypedia* will remain entirely dependent on the diligence and commitment of experts and keen junior scientists to voluntarily contribute to the maintenance and growth of this reference work. The job of building *CAZypedia* is by no means complete, and as a living encyclopedia, it never will be – especially as research continues to reveal new CAZyme families, tertiary structures, and mechanistic details (Abe, K., Nakajima, M., et al. 2017, Campos, B.M., Liberato, M.V., et al. 2016, Munoz-Munoz, J., Cartmell, A., et al. 2017a, Munoz-Munoz, J., Cartmell, A., et al. 2017b, Ndeh, D., Rogowski, A., et al. 2017, Venditto, I., Luis, A.S., et al. 2016). Currently, many pages remain to be written and existing pages would benefit from regular updates as new data come to hand, which requires the identification of experts willing to assume the responsibility for page creation and maintenance.

---

[11] Freely available at URL https://www.mediawiki.org/wiki/Extension:BiblioPlus

[12] See URL https://www.mediawiki.org/wiki/Extension:Biblio

Thus, the *CAZypedia* Consortium openly invites all interested glycoscientists, regardless of career stage (including keen undergraduate and postgraduate students, post-doctoral researchers, industrial scientists, and professors) to peruse the 'Unassigned Pages' lists for each CAZyme class and see if they might be able to help. The growth of *CAZypedia* will depend exclusively on the generous and selfless contributions of the existing and new generations of CAZypedians. We invite you to join us! Contact information is available at URL http://www.cazypedia.org/.

Acknowledgements

We thank Stephen MacDonald and Vince Tingey (Michael Smith Laboratories, University of British Columbia), and Eric Björkvall (School of Biotechnology, Kungliga Tekniska Högskolan), for invaluable IT support. Dr. Karen Eddy (Brumer group, MSL, UBC) is thanked for developing the BiblioPlus extension[13]. *CAZypedia* is the result of many hours of effort by the following group of current contributors:[14]

Wade Abbott, Agriculture and Agri-Food Canada, Canada

Orly Alber, Weizmann Institute of Science, Israel

Ed Bayer, Weizmann Institute of Science, Israel

Jean-Guy Berrin, Institut National de la Recherche Agronomique, France

Alisdair Boraston, University of Victoria, Canada

Harry Brumer, University of British Columbia, Canada

Ryszard Brzezinski, Université de Sherbrooke, Canada

Anthony Clarke, University of Guelph, Canada

Beatrice Cobucci-Ponzano, National Research Council of Italy, Italy

Darrell Cockburn, Penn State University, United States of America

Pedro Coutinho, Aix Marseille Université, France

Mirjam Czjzek, Centre National de la Recherche Scientifique, France

Bareket Dassa, Weizmann Institute of Science, Israel

Gideon John Davies, University of York, United Kingdom

Vincent Eijsink, Norwegian University of Life Sciences, Norway

Jens Eklöf, University of British Columbia, Canada

Alfons Felice, Universität für Bodenkultur, Austria

Elizabeth Ficko-Blean, Centre National de la Recherche Scientifique, France

Geoff Fincher, University of Adelaide, Australia

Thierry Fontaine, Institut Pasteur, France

Zui Fujimoto, National Agriculture and Food Research Organisation, Japan

Kiyotaka Fujita, Kagoshima University, Japan

Shinya Fushinobu, University of Tokyo, Japan

Harry Gilbert, Newcastle University, United Kingdom

Tracey Gloster, University of St. Andrews, United Kingdom

Ethan Goddard-Borger, Walter and Eliza Hall Institute of Medical Research, Australia

---

[13] Freely available for download at URL https://www.mediawiki.org/wiki/Extension:BiblioPlus

[14] A continually updated list is available at URL http://www.cazypedia.org/index.php/Category:Contributors, accessed via the 'About CAZypedia' menu.

Ian Greig, Simon Fraser University, Canada

Jan-Hendrik Hehemann, Max Planck Institute for Marine Microbiology, Germany

Glyn Hemsworth, University of Leeds, United Kingdom

Bernard Henrissat, Centre National de la Recherche Scientifique, France

Masafumi Hidaka, University of Tokyo, Japan

Ramon Hurtado-Guerrero, University of Zaragoza, Spain

Kiyohiko Igarashi, University of Tokyo, Japan

Takuya Ishida, University of Tokyo, Japan

Stefan Janecek, Slovak Academy of Sciences, Slovakia

Seino Jongkees, University of Tokyo, Japan

Nathalie Juge, Quadram Institute, United Kingdom

Satoshi Kaneko, University of the Ryukyus, Japan

Takane Katayama, Ishikawa Prefectural University, Japan

Motomitsu Kitaoka, National Agriculture and Food Research Organisation, Japan

Naotake Konno, Utsunomiya University, Japan

Daniel Kracher, Universität für Bodenkultur, Austria

Anna Kulminskaya, Petersburg Nuclear Physics Institute, Russia

Alicia Lammerts van Bueren, University of Groningen, Netherlands

Sine Larsen, University of Copenhagen, Denmark

Junho Lee, University of British Columbia, Canada

Markus Linder, Aalto University, Finland

Leila LoLeggio, University of Copenhagen, Denmark

Roland Ludwig, Universität für Bodenkultur, Austria

Ana Luis, Universtiy of Lisbon, Portugal

Mirko Maksimainen, University of Oulu, Finland

Brian Mark, University of Manitoba, Canada

Richard McLean, University of Lethbridge, Canada

Gurvan Michel, Centre National de la Recherche Scientifique, France

Cedric Montanier, Institut National de la Recherche Agronomique, France

Marco Moracci, National Research Council of Italy, Italy

Haruhide Mori, Hokkaido University, Japan

Hiroyuki Nakai, Niigata University, Japan

Wim Nerinckx, Ghent University, Belgium

Takayuki Ohnuma, Kinki University, Japan

Richard Pickersgill, Queen Mary University of London, United Kingdom

Kathleen Piens, Sveriges Lantbruksuniversitet, Sweden

Tirso Pons, National Centre for Biotechnology, Spain

Etienne Rebuffet, Centre National de la Recherche Scientifique, France

Peter Reilly, Iowa State University, United States of America

Magali Remaud-Simeon, Institut National des Sciences Appliquées, France

Brian Rempel, University of British Columbia, Canada

Kyle Robinson, University of British Columbia, Canada
David Rose, University of Waterloo, Canada
Juha Rouvinen, University of Eastern Finland, Finland
Wataru Saburi, Hokkaido University, Japan
Yuichi Sakamoto, Iwate Biotechnology Research Center, Japan
Mats Sandgren, Sveriges Lantbruksuniversitet, Sweden
Fathima Shaikh, University of British Columbia, Canada
Yuval Shoham, Technion, Israel
Franz St. John, United States Department of Agriculture, United States of America
Jerry Stahlberg, Sveriges Lantbruksuniversitet, Sweden
Michael Suits, Wilfrid Laurier University, Canada
Gerlind Sulzenbacher, Centre National de la Recherche Scientifique, France
Tomomi Sumida, RIKEN, Japan
Ryuichiro Suzuki, Akita Prefectural University, Japan
Birte Svensson, Danmarks Tekniske Universitet, Denmark
Toki Taira, University of the Ryukyus, Japan
Ed Taylor, University of Lincoln, United Kingdom
Takashi Tonozuka, Tokyo University of Agriculture and Technology, Japan
Breeanna Urbanowicz, University of Georgia, United States of America
Gustav Vaaje-Kolstad, Norwegian University of Life Sciences, Norway
Wim Van den Ende, Katholieke Universiteit Leuven, Belgium
Annabelle Varrot, Centre National de la Recherche Scientifique, France
Maxime Versluys, Katholieke Universiteit Leuven, Belgium
Florence Vincent, Centre National de la Recherche Scientifique, France
David Vocadlo, Simon Fraser University, Canada
Warren Wakarchuk, Ryerson University, Canada
Tom Wennekes, Universiteit Utrecht, Netherlands
Rohan Williams, University of Melbourne, Australia
Spencer Williams, University of Melbourne, Australia
David Wilson†, Cornell University, United States of America
Stephen Withers, University of British Columbia, Canada
Katsuro Yaoi, National Institute of Advanced Industrial Science and Technology, Japan
Vivian Yip, University of British Columbia, Canada
Ran Zhang, University of British Columbia, Canada

This Letter was composed by H. Brumer (brumer@msl.ubc.ca) and S.J. Williams (sjwill@unimelb.edu.au), with input from B. Svensson, B. Henrissat, G.J. Davies, H.J. Gilbert, A.J. Clarke, W.W. Wakarchuk, D.W. Abbott, D.J. Vocadlo, E. Ficko-Blean, A.B. Boraston, A. Planas, and S. Fushinobu.

Page charges for this article were supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant to H. Brumer.

References

Abe K, Nakajima M, Yamashita T, Matsunaga H, Kamisuki S, Nihira T, Takahashi Y, Sugimoto N, Miyanaga A, Nakai H*, et al.* 2017. Biochemical and structural analyses of a bacterial endo-beta-1,2-glucanase reveal a new glycoside hydrolase family. *Journal of Biological Chemistry*, 292:7487-7506.

Anonymous. 2010. *Entrez Programming Utilities Help* National Center for Biotechnology Information:Bethesda, MD, USA. Available from: https://www.ncbi.nlm.nih.gov/books/NBK25501/.

Boraston AB, Bolam DN, Gilbert HJ, Davies GJ. 2004. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochemical Journal*, 382:769-781.

Bourne PE, McEntyre J. 2006. Biocurators: Contributors to the world of science. *PLoS Computational Biology*, 2:1185-1185.

Campos BM, Liberato MV, Alvarez TM, Zanphorlin LM, Ematsu GC, Barud H, Polikarpov I, Ruller R, Gilbert HJ, Zeri ACD*, et al.* 2016. A Novel Carbohydrate-binding Module from Sugar Cane Soil Metagenome Featuring Unique Structural and Carbohydrate Affinity Properties. *Journal of Biological Chemistry*, 291:23734-23743.

Coutinho PM, Deleury E, Davies GJ, Henrissat B. 2003. An evolving hierarchical family classification for glycosyltransferases. *Journal of Molecular Biology*, 328:307-317.

Davies G, Henrissat B. 1995. Structures and Mechanisms of Glycosyl Hydrolases. *Structure*, 3:853-859.

Davies GJ, Sinnott ML. 2008. Sorting the Diverse. *The Biochemist*, 30:26-32.

Henrissat B. 1991. A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochemical Journal*, 280:309-316.

Levasseur A, Drula E, Lombard V, Coutinho PM, Henrissat B. 2013. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnology for Biofuels*, 6:14.

Lombard V, Bernard T, Rancurel C, Brumer H, Coutinho PM, Henrissat B. 2010. A hierarchical classification of polysaccharide lyases for glycogenomics. *Biochemical Journal*, 432:437-444.

Lombard V, Ramulu HG, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res*, 42:D490-D495.

Munoz-Munoz J, Cartmell A, Terrapon N, Basle A, Henrissat B, Gilbert HJ. 2017a. An evolutionarily distinct family of polysaccharide lyases removes rhamnose capping of complex arabinogalactan proteins. *The Journal of biological chemistry*, 292:13271-13283.

Munoz-Munoz J, Cartmell A, Terrapon N, Henrissat B, Gilbert HJ. 2017b. Unusual active site location and catalytic apparatus in a glycoside hydrolase family. *Proceedings of the National Academy of Sciences of the United States of America*, 114:4936-4941.

Ndeh D, Rogowski A, Cartmell A, Luis AS, Basle A, Gray J, Venditto I, Briggs J, Zhang XY, Labourel A*, et al.* 2017. Complex pectin metabolism by gut bacteria reveals novel catalytic functions. *Nature*, 544:65-+.

Sinnott ML. 1990. Catalytic mechanisms of enzymatic glycosyl transfer. *Chemical Reviews*, 90:1171-1202.

Venditto I, Luis AS, Rydahl M, Schuckel J, Fernandes VO, Vidal-Melgosa S, Bule P, Goyal A, Pires VMR, Dourado CG*, et al.* 2016. Complexity of the *Ruminococcus flavefaciens* cellulosome reflects an expansion in glycan recognition. *Proceedings of the National Academy of Sciences of the United States of America*, 113:7136-7141.

Whelan W. 2009. Obituary: Bruce A. Stone. *IUBMB Life*, 61:84-84.

Figures

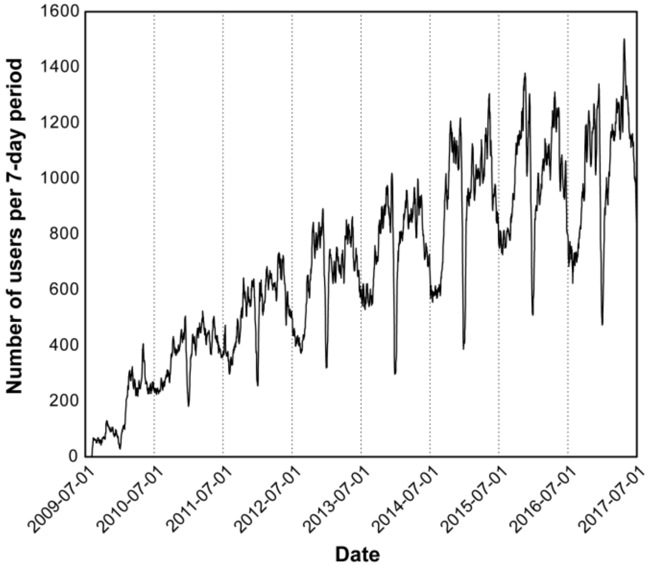
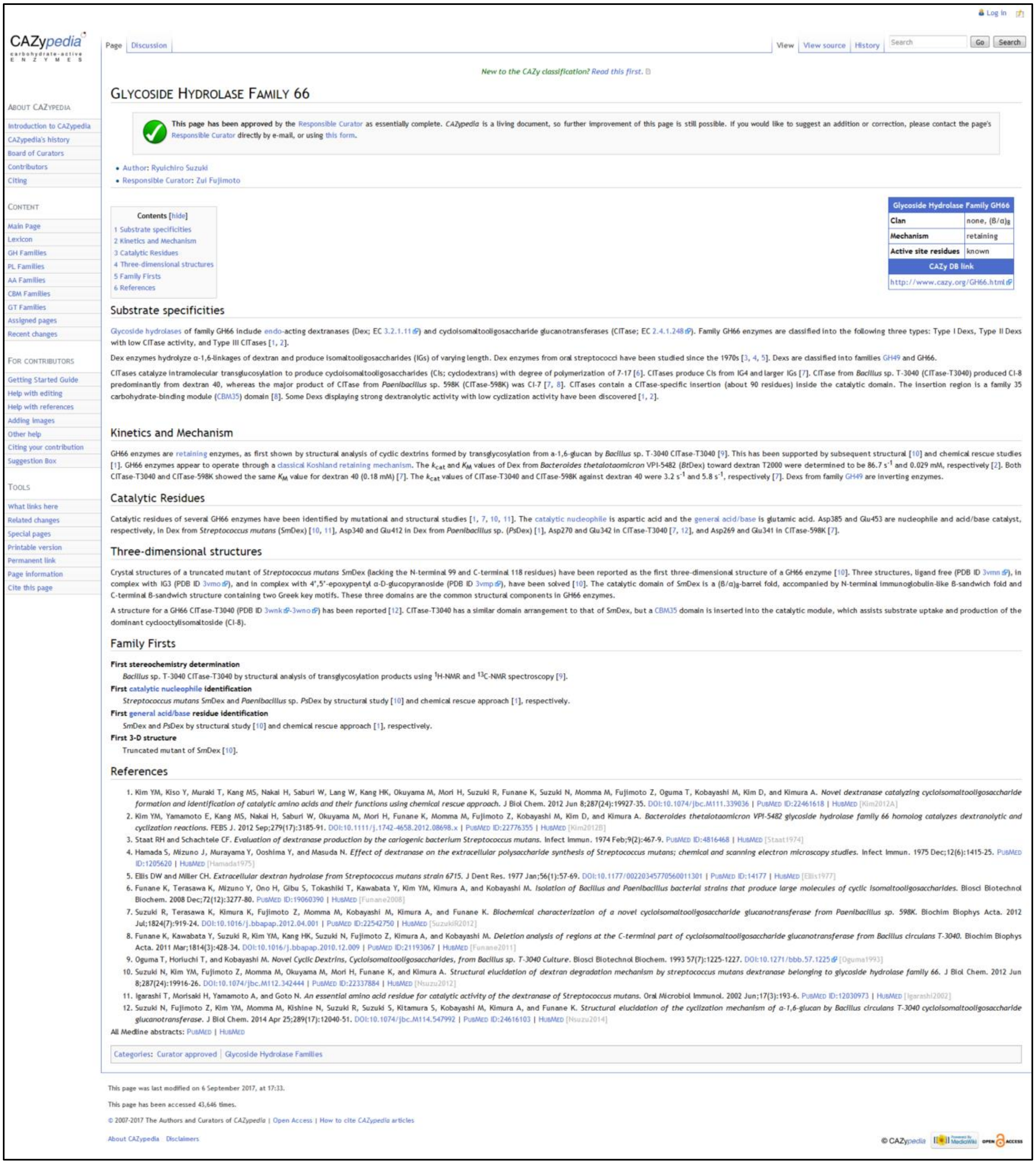


Figure 1. Layout of a typical CAZyme family page in *CAZypedia*.



Log in

CAZypedia
carbohydrate-active ENZYMES

Page | Discussion | View | View source | History | Search | Go | Search

About CAZypedia
Introduction to CAZypedia
CAZypedia's history
Board of Curators
Contributors
Citing

Content
Main Page
Lexicon
GH Families
PL Families
AA Families
CBM Families
GT Families
Assigned pages
Recent changes

For contributors
Getting Started Guide
Help with editing
Help with references
Adding images
Other help
Citing your contribution
Suggestion Box

Tools
What links here
Related changes
Special pages
Printable version
Permanent link
Page information
Cite this page

New to the CAZy classification? Read this first.

Glycoside Hydrolase Family 66

This page has been approved by the Responsible Curator as essentially complete. *CAZypedia* is a living document, so further improvement of this page is still possible. If you would like to suggest an addition or correction, please contact the page's Responsible Curator directly by e-mail, or using this form.

- Author: Ryuichiro Suzuki
- Responsible Curator: Zui Fujimoto

Contents [hide]

Glycoside Hydrolase Family GH66	
Clan	none, $(\beta/\alpha)_8$
Mechanism	retaining
Active site residues	known
CAZy DB link	
http://www.cazy.org/GH66.html	

Substrate specificities

Glycoside hydrolases of family GH66 include endo-acting dextranases (Dex; EC 3.2.1.11) and cycloisomaltooligosaccharide glucanotransferases (CITase; EC 2.4.1.248). Family GH66 enzymes are classified into the following three types: Type I Dexs, Type II Dexs with low CITase activity, and Type III CITases [1, 2].

Dex enzymes hydrolyze α-1,6-linkages of dextran and produce isomaltooligosaccharides (IGs) of varying length. Dex enzymes from oral streptococci have been studied since the 1970s [3, 4, 5]. Dexs are classified into families GH49 and GH66.

CITases catalyze intramolecular transglucosylation to produce cycloisomaltooligosaccharides (CIs; cyclodextrans) with degree of polymerization of 7-17 [6]. CITases produce CIs from IG4 and larger IGs [7]. CITase from *Bacillus* sp. T-3040 (CITase-T3040) produced CI-8 predominantly from dextran 40, whereas the major product of CITase from *Paenibacillus* sp. 598K (CITase-598K) was CI-7 [7, 8]. CITases contain a CITase-specific insertion (about 90 residues) inside the catalytic domain. The insertion region is a family 35 carbohydrate-binding module (CBM35) domain [8]. Some Dexs displaying strong dextranolytic activity with low cyclization activity have been discovered [1, 2].

Kinetics and Mechanism

GH66 enzymes are retaining enzymes, as first shown by structural analysis of cyclic dextrins formed by transglycosylation from a-1,6-glucan by *Bacillus* sp. T-3040 CITase-T3040 [9]. This has been supported by subsequent structural [10] and chemical rescue studies [1]. GH66 enzymes appear to operate through a classical Koshland retaining mechanism. The k_{cat} and K_M values of Dex from *Bacteroides thetaiotaomicron* VPI-5482 (*Bt*Dex) toward dextran T2000 were determined to be 86.7 s^{-1} and 0.029 mM, respectively [2]. Both CITase-T3040 and CITase-598K showed the same K_M value for dextran 40 (0.18 mM) [7]. The k_{cat} values of CITase-T3040 and CITase-598K against dextran 40 were 3.2 s^{-1} and 5.8 s^{-1}, respectively [7]. Dexs from family GH49 are inverting enzymes.

Catalytic Residues

Catalytic residues of several GH66 enzymes have been identified by mutational and structural studies [1, 7, 10, 11]. The catalytic nucleophile is aspartic acid and the general acid/base is glutamic acid. Asp385 and Glu453 are nucleophile and acid/base catalyst, respectively, in Dex from *Streptococcus mutans* (*Sm*Dex) [10, 11], Asp340 and Glu412 in Dex from *Paenibacillus* sp. (*Ps*Dex) [1], Asp270 and Glu342 in CITase-T3040 [7, 12], and Asp269 and Glu341 in CITase-598K [7].

Three-dimensional structures

Crystal structures of a truncated mutant of *Streptococcus mutans* SmDex (lacking the N-terminal 99 and C-terminal 118 residues) have been reported as the first three-dimensional structure of a GH66 enzyme [10]. Three structures, ligand free (PDB ID 3vmn), in complex with IG3 (PDB ID 3vmo), and in complex with 4',5'-epoxypentyl α-D-glucopyranoside (PDB ID 3vmp), have been solved [10]. The catalytic domain of SmDex is a $(\beta/\alpha)_8$-barrel fold, accompanied by N-terminal immunoglobulin-like β-sandwich fold and C-terminal β-sandwich structure containing two Greek key motifs. These three domains are the common structural components in GH66 enzymes.

A structure for a GH66 CITase-T3040 (PDB ID 3wnk-3wno) has been reported [12]. CITase-T3040 has a similar domain arrangement to that of SmDex, but a CBM35 domain is inserted into the catalytic module, which assists substrate uptake and production of the dominant cyclooctylisomaltoside (CI-8).

Family Firsts

First stereochemistry determination
Bacillus sp. T-3040 CITase-T3040 by structural analysis of transglycosylation products using ^{1}H-NMR and ^{13}C-NMR spectroscopy [9].
First catalytic nucleophile identification
Streptococcus mutans SmDex and *Paenibacillus* sp. *PsDex* by structural study [10] and chemical rescue approach [1], respectively.
First general acid/base residue identification
SmDex and *PsDex* by structural study [10] and chemical rescue approach [1], respectively.
First 3-D structure
Truncated mutant of *SmDex* [10].

References

1. Kim YM, Kiso Y, Muraki T, Kang MS, Nakai H, Saburi W, Lang W, Kang HK, Okuyama M, Mori H, Suzuki R, Funane K, Suzuki N, Momma M, Fujimoto Z, Oguma T, Kobayashi M, Kim D, and Kimura A. *Novel dextranase catalyzing cycloisomaltooligosaccharide formation and identification of catalytic amino acids and their functions using chemical rescue approach.* J Biol Chem. 2012 Jun 8;287(24):19927-35. DOI:10.1074/jbc.M111.339036 | PubMed ID:22461618 | HubMed [Kim2012A]
2. Kim YM, Yamamoto E, Kang MS, Nakai H, Saburi W, Okuyama M, Mori H, Funane K, Momma M, Fujimoto Z, Kobayashi M, Kim D, and Kimura A. *Bacteroides thetaiotaomicron VPI-5482 glycoside hydrolase family 66 homolog catalyzes dextranolytic and cyclization reactions.* FEBS J. 2012 Sep;279(17):3185-91. DOI:10.1111/j.1742-4658.2012.08698.x | PubMed ID:22776355 | HubMed [Kim2012B]
3. Staat RH and Schachtele CF. *Evaluation of dextranase production by the cariogenic bacterium Streptococcus mutans.* Infect Immun. 1974 Feb;9(2):467-9. PubMed ID:4816468 | HubMed [Staat1974]
4. Hamada S, Mizuno J, Murayama Y, Ooshima Y, and Masuda N. *Effect of dextranase on the extracellular polysaccharide synthesis of Streptococcus mutans; chemical and scanning electron microscopy studies.* Infect Immun. 1975 Dec;12(6):1415-25. PubMed ID:1205620 | HubMed [Hamada1975]
5. Ellis DW and Miller CH. *Extracellular dextran hydrolase from Streptococcus mutans strain 6715.* J Dent Res. 1977 Jan;56(1):57-69. DOI:10.1177/00220345770560011301 | PubMed ID:14177 | HubMed [Ellis1977]
6. Funane K, Terasawa K, Mizuno Y, Ono H, Gibu S, Tokashiki T, Kawabata Y, Kim YM, Kimura A, and Kobayashi M. *Isolation of Bacillus and Paenibacillus bacterial strains that produce large molecules of cyclic isomaltooligosaccharides.* Biosci Biotechnol Biochem. 2008 Dec;72(12):3277-80. PubMed ID:19060390 | HubMed [Funane2008]
7. Suzuki R, Terasawa K, Kimura K, Fujimoto Z, Momma M, Kobayashi M, Kimura A, and Funane K. *Biochemical characterization of a novel cycloisomaltooligosaccharide glucanotransferase from Paenibacillus sp. 598K.* Biochim Biophys Acta. 2012 Jul;1824(7):919-24. DOI:10.1016/j.bbapap.2012.04.001 | PubMed ID:22542750 | HubMed [SuzukiR2012]
8. Funane K, Kawabata Y, Suzuki R, Kim YM, Kang HK, Suzuki N, Fujimoto Z, Kimura A, and Kobayashi M. *Deletion analysis of regions at the C-terminal part of cycloisomaltooligosaccharide glucanotransferase from Bacillus circulans T-3040.* Biochim Biophys Acta. 2011 Mar;1814(3):428-34. DOI:10.1016/j.bbapap.2010.12.009 | PubMed ID:21193067 | HubMed [Funane2011]
9. Oguma T, Horiuchi T, and Kobayashi M. *Novel Cyclic Dextrins, Cycloisomaltooligosaccharides, from Bacillus sp. T-3040 Culture.* Biosci Biotechnol Biochem. 1993 57(7):1225-1227. DOI:10.1271/bbb.57.1225 [Oguma1993]
10. Suzuki N, Kim YM, Fujimoto Z, Momma M, Okuyama M, Mori H, Funane K, and Kimura A. *Structural elucidation of dextran degradation mechanism by streptococcus mutans dextranase belonging to glycoside hydrolase family 66.* J Biol Chem. 2012 Jun 8;287(24):19916-26. DOI:10.1074/jbc.M112.342444 | PubMed ID:22337884 | HubMed [Nsuzu2012]
11. Igarashi T, Morisaki H, Yamamoto A, and Goto N. *An essential amino acid residue for catalytic activity of the dextranase of Streptococcus mutans.* Oral Microbiol Immunol. 2002 Jun;17(3):193-6. PubMed ID:12030973 | HubMed [Igarashi2002]
12. Suzuki N, Fujimoto Z, Kim YM, Momma M, Kishine N, Suzuki R, Suzuki S, Kitamura S, Kobayashi M, Kimura A, and Funane K. *Structural elucidation of the cyclization mechanism of α-1,6-glucan by Bacillus circulans T-3040 cycloisomaltooligosaccharide glucanotransferase.* J Biol Chem. 2014 Apr 25;289(17):12040-51. DOI:10.1074/jbc.M114.547992 | PubMed ID:24616103 | HubMed [Nsuzu2014]

All Medline abstracts: PubMed | HubMed

Categories: Curator approved | Glycoside Hydrolase Families

This page was last modified on 6 September 2017, at 17:33.

This page has been accessed 43,646 times.

Figure 2. ***CAZypedia* usage statistics from Google Analytics.** Access tracking with this service was initiated in August 2009. Sharp dips correspond to December holidays and broad troughs correspond to summer in the northern hemisphere.