

## ИЕРАРХИЧЕСКАЯ КЛАССИФИКАЦИЯ ГЛИКОЗИЛ-ГИДРОЛАЗ

### Обзор

© 2011 г. Д.Г. Наумов

*ГНИИ генетики и селекции промышленных микроорганизмов,  
117545 Москва, Первый Дорожный проезд, д.1;  
электронная почта: daniil\_naumoff@yahoo.com*

Поступила в редакцию 21.01.11

В обзоре рассматриваются структурные и функциональные особенности гликозил-гидролаз — обширной группы ферментов, представленных почти во всех живых организмах. В международной классификации действующих на углеводы ферментов (CAZy) гликозил-гидролазы на основе сходства аминокислотных последовательностей каталитических доменов объединены в 120 семейств. На более высоком иерархическом уровне часть семейств сгруппирована в 14 кланов. Ферменты одного клана не только обладают эволюционным родством кодирующих их генов, но и имеют одинаковые важнейшие функциональные характеристики: строение активного центра, стереохимию расщепляемых гликозидных связей и молекулярный механизм катализируемой реакции. К настоящему времени в литературе накоплены обширные данные о родстве между семействами гликозил-гидролаз, которые относятся к различным кланам и/или не включены в состав ни одного из кланов, а также по филогенетическим взаимоотношениям между белками внутри отдельных семейств. Путем суммирования этих данных в обзоре предложена многоуровневая иерархическая классификация гликозил-гидролаз и их гомологов. Показано, что почти все разнообразие каталитических доменов этих ферментов может быть сведено к шести основным фолдам — крупным группам белков, имеющих однотипную пространственную структуру и предполагаемую общность эволюционного происхождения.

**КЛЮЧЕВЫЕ СЛОВА:** гликозил-гидролаза, CAZy, иерархическая классификация белков, эволюция белков, семейство белков, клан, суперсемейство, подсемейство, поиск гомологов, TIM-бочонок, SCOP, Pfam, PSI-BLAST.

«The establishing of a system for classification and nomenclature can be seen as a clear indication of the «coming of age» of a branch of science.»  
Alan J. Barrett, 1994 [1]

Гликозил-гидролазы (гликозидазы или карбогидразы; К.Ф. 3.2.1) — обширная группа ферментов, катализирующих гликолитическое расщепление *O*-гликозидной связи. Их принято относить к катаболическим ферментам углеводного обмена; в течение многих десятилетий они являются объектами разнообразных биохимических исследований [2–10]. Гены гликозил-гидролаз обнаружены в геномах почти всех живых организмов, исключение составляют часть архей и некоторые паразитические одноклеточные эукариоты [11, 12]. Ряд вирусов также кодирует гликозидазы. На гликозидазы и их близкие гомологи приходится около 1% всех известных генов [12]. Доля этих генов в геноме существенно варьирует в зависимости от таксономического положения организма [11]. Набор гликозил-

гидролаз у микроорганизмов в значительной мере зависит от их экологической ниши: гены гликозидаз часто подвержены дупликациям (могут быть представлены в виде нескольких паралога [7, 13–17]), элиминации (набор генов может различаться даже у штаммов одного вида [18]) и горизонтальному переносу (филогенетические деревья гликозидаз, как правило, сильно отличаются от деревьев организмов-хозяев [14–16, 19, 20]). Эволюционно близкородственные гликозидазы часто различаются по субстратной специфичности [9, 11, 21, 22], а среди гомологов гликозидаз обнаружены ферменты с самыми различными энзиматическими активностями, а также белки, не являющиеся ферментами [11, 23]. Минимальные изменения в первичной структуре гликозил-гидролаз способны из-

менить их субстратную специфичность [24]. Многие гликозил-гидролазы обладают сложной доменной структурой, а разные домены одного белка часто имеют независимую эволюционную историю. При этом в одном белке могут содержаться два или более гомологичных [25–27] или негомологичных [7, 28, 29] каталитических домена.

Гликозил-гидролазы и их гены стали модельными объектами для различных областей биологии: исследования дрожжевых лизатов, содержащих инвертазу (К.Ф. 3.2.1.26), когда-то заложили основы энзимологии [30]; своего рода эталоном регуляции экспрессии генов у бактерий стал *lac*-оперон из *Escherichia coli* [31], содержащий в-галактозидазный (К.Ф. 3.2.1.23) ген *lacZ*, который нашел широкое применение при конструировании векторов для клонирования; первым белком с экспериментально определенной третичной структурой (PDB, 1HEW) оказался лизоцим (К.Ф. 3.2.1.17) из куриного яйца [32].

Двумя самыми распространенными органическими веществами в живых организмах являются целлюлоза и хитин [2, 33], что делает углеводы доминирующим по массе компонентом в живой материи на Земле. Некоторые гликозидные связи, в частности, связи между двумя глюкозными остатками, являются химически наиболее стабильными связями между мономерами в природных биополимерах [34]. Расщепляющие такие связи гликозил-гидролазы ускоряют соответствующую реакцию на 17 порядков в сравнении со спонтанным гидролизом, что делает их одними из наиболее эффективных из известных нам катализаторов [34].

### ТРАДИЦИОННАЯ (ЭНЗИМАТИЧЕСКАЯ) КЛАССИФИКАЦИЯ ГЛИКОЗИЛ-ГИДРОЛАЗ

Гликозил-гидролазы в широком смысле слова (К.Ф. 3.2) объединяют *O*-, *N*- и *S*-гликозил-гидролазы [4, 8, 35]. Однако *N*-гликозил-гидролазы (К.Ф. 3.2.2) являются, главным образом, ферментами метаболизма нуклеиновых кислот и в настоящем обзоре не рассматриваются. Единственно известная *S*-гликозил-гидролаза (К.Ф. 3.2.3) – тиоглюкозидаза (К.Ф. 3.2.3.1) – в 2001 г. была реклассифицирована как *O*-гликозил-гидролаза с энзиматическим номером (К.Ф. 3.2.1.147) [36].

Большое разнообразие известных энзиматических активностей гликозил-гидролаз связано с огромным числом их природных субстратов – ди-, олиго- и полисахаридов, а также их произ-

водных. Количество теоретически возможных структур углеводов воистину астрономическое. Например, низкомолекулярный восстанавливающий (редуцирующий) гексасахарид  $C_{36}H_{62}O_{31}$  может иметь более  $10^{12}$  разветвленных и линейных изомеров [37]. Традиционная (биохимическая) номенклатура гликозидаз основана исключительно на их субстратной специфичности и лишь иногда (в случае амилаз – К.Ф. 3.2.1.1 и К.Ф. 3.2.1.2) учитывает молекулярный механизм катализируемой реакции [35]. При этом рассматривается только основной, наиболее предпочтительный субстрат того или иного фермента. Преимуществом этой номенклатуры является сравнительная простота и длительный опыт применения, что обеспечило ее стандартизацию. К настоящему времени зарегистрированы энзиматические активности от (К.Ф. 3.2.1.1) до (К.Ф. 3.2.1.165), при этом регистрация части номеров по разным причинам была аннулирована [36].

Многие гликозидазы имеют широкую субстратную специфичность [3, 21], что затрудняет их классификацию. Часто гликозидазы наряду с гликозил-гидролазной (К.Ф. 3.2.1) проявляют еще и трансгликозидазную активность (К.Ф. 2.4.1) [2–4, 21, 38–40]. Это формально позволяет рассматривать их как представителей другого класса ферментов – трансфераз. Для целого ряда гликозилтрансфераз (точнее трансгликозидаз) обнаружено существенное структурное сходство (на уровне первичных или третичных структур) с некоторыми гликозил-гидролазами [11]. Однако эти данные, свидетельствующие об общности эволюционного происхождения и сходстве механизма катализа у ферментов этих двух групп, не принимаются во внимание биохимической классификацией [35, 36].

Гликозидазы катализируют гидролиз гликозидной связи по двум различным механизмам (*retaining* и *invertig*), приводящим либо к сохранению (путем двойного обращения), либо к обращению аномерной конфигурации субстрата в процессе реакции [4–6, 19, 41–44]. Каждый фермент обладает одним из них вне зависимости от выбора субстрата, а энзиматический гидролиз гликозидной связи всегда является стереоспецифическим. Однако известны случаи, когда гидролиз некоторых низкомолекулярных синтетических субстратов – производных или аналогов углеводов – идет по иному механизму, чем гидролиз природного субстрата, или гидролизу могут подвергаться синтетические субстраты (например, гликозил-фториды) обеих ( $\alpha$  и  $\beta$ ) аномерных конфигураций [3–5, 39, 44, 45]. Знание стереохимии реакции гидролиза позволяет предсказать способность фермента к катализу

побочной реакции – переносу гликозидных остатков между молекулами субстрата (трансгликозилирование). Такая способность свойственна только гликозидазам, действующим с сохранением аномерной конфигурации субстрата у продукта реакции [3–7, 19, 46].

Огромный экспериментальный материал накоплен в результате исследований механизма действия и специфичности различных гликозил-гидролаз. Показано, что абсолютной специфичностью гликозидазы практически во всех случаях обладают лишь по отношению к двум элементам структуры субстрата – к конфигурации расщепляемой гликозидной связи ( $\alpha$  или  $\beta$ ) и размеру окисного цикла отщепляемого моносахаридного остатка (шестичленный пиранозный или пятичленный фуранозный циклы). В остальном для них характерна большая широта специфичности [5].

Для гликозидаз, действующих на пиранозные остатки, было предложено [4, 46–48] наряду с механизмом реакции (приводящим либо к сохранению, либо к обращению аномерной конфигурации) учитывать еще аксиальную (как правило,  $\alpha$ ) или экваториальную (как правило,  $\beta$ ) направленность гидролизуемой гликозидной связи, что позволяет выделить четыре класса ферментов. Аналогичная система классификации применима и к фуранозидазам. Наряду с этим, гликозидазы подразделяются на *syn*- и *anti*-протонаторы в зависимости от пространственного расположения каталитических остатков фермента по отношению к субстрату [49]. Стереохимия реакции гидролиза обуславливается взаимным расположением каталитически важных групп в активном центре фермента. Как правило, энзиматический гидролиз гликозидной связи нуждается в наличии двух каталитически важных аминокислотных остатков в активном центре. В подавляющем большинстве случаев их роль выполняют карбоксилсодержащие остатки (Asp и/или Glu). Карбоксильные группы этих двух остатков обычно находятся на расстоянии около 5–6 или 9–10 ангстремов в случае гликозидаз, действующих с сохранением и обращением аномерной конфигурации субстрата соответственно. В последнем случае большее расстояние обусловлено необходимостью одновременного нахождения в процессе реакции гидролиза субстрата и молекулы воды в активном центре фермента [3, 4, 41–44, 49–52].

Гликозидазы, гидролизующие олиго- и полисахариды, на основании характера расщепления ими субстрата могут быть подразделены на две группы – ферменты экзо- и эндо-типа. Несмотря на наличие некоторых переходных случа-

ев, в целом, подобное разделение весьма удачно [5, 7, 46].

### КЛАССИФИКАЦИЯ CAZy: СЕМЕЙСТВА И КЛАНЫ ГЛИКОЗИЛ-ГИДРОЛАЗ

Накопление данных о первичной структуре гликозил-гидролаз позволило разработать принципиально новую систему классификации этих ферментов, основанную на сравнении аминокислотных последовательностей их каталитических доменов. Впервые этот подход был использован при анализе  $\beta$ -1,4-гликозидаз, что позволило сгруппировать их в восемь семейств [53–55]. Затем были изучены все известные тогда последовательности гликозил-гидролаз (более 300), в итоге выявлено 36 семейств белковых гомологов [56]. За прошедшие с тех пор два десятилетия произошло, по крайней мере, стократное возрастание числа гликозидаз с известными аминокислотными последовательностями (табл. 1). Большая часть новых последовательностей оказалась принадлежащей ранее известным семействам, а остальные образовали ряд новых семейств [11, 29, 46, 47, 57–59]. С 1998 г. классификация гликозил-гидролаз была распространена на ферменты с трансгликозидазной активностью. В настоящее время известно 120 семейств гликозидаз: GH1–GH125. При этом семейства GH21, GH40, GH41, GH60 и GH69 были аннулированы, поскольку более тщательные исследования показали, что относящиеся к ним ферменты не обладают гликозидазными активностями [11, 47, 60]. Текущая версия классификации доступна в интернете на сайте CAZy [11].

Критерием для отнесения какого-то белка, точнее белкового домена, к тому или иному семейству гликозил-гидролаз является наличие в последовательности этого белка протяженного участка (не менее 100 аминокислотных остатков), гомологичного последовательности, по крайней мере, одного ранее известного представителя семейства [56]. Применение этого критерия привело к включению в состав семейств гликозил-гидролаз целого ряда ферментов, не обладающих гликозидазными активностями [23], а также большого числа «гипотетических» белков с неизвестными энзиматическими свойствами [11]. Доля неохарактеризованных белков среди членов семейств гликозидаз постоянно возрастает, что связано с быстрыми темпами секвенирования новых геномов. Семейства гликозидаз с большим числом известных представителей было предложено подразделять на под-

семейства (subfamilies, subtypes) на основании уровня сходства аминокислотных последовательностей [53, 61, 62].

В настоящее время в большинстве семейств гликозил-гидролаз, по крайней мере, для одного представителя известен молекулярный механизм или сохранением аномальной конфигурации субстрата [9, 11, 41, 43, 46, 49, 58, 63]. Во всех известных случаях ферменты одного семейства имеют одинаковый механизм [41, 43, 63], исключением являются семейства GH23 и GH97, объединяющие «сохраняющие» и «обращающие» гликозил-гидролазы [64]. Консервативной в пределах каждого семейства гликозидаз также является аномальная конфигурация расщепляемой гликозидной связи ( $\alpha$  или  $\beta$ ), точнее говоря, ее аксиальная или экваториальная ориентация [43, 46]. Единственно известное исключение представляет семейство GH4 гликозидаз: в его состав входят как ферменты расщепляющие  $\alpha$ -гликозидную связь, так и ферменты расщепляющие  $\beta$ -гликозидную связь [11, 52, 65].

Более высокая консервативность третичных структур белков в сравнении с первичными позволяет предполагать возможное сходство пространственных структур некоторых гликозидаз, относящихся к разным семействам. Известен целый ряд подобных случаев, которые, прежде всего, касаются семейств, содержащих белки с экспериментально определенными третичными структурами. В ряде других случаев о предполагаемом сходстве трехмерных структур гликозидаз можно судить на основании наличия отдаленного, но статистически достоверного сходства аминокислотных последовательностей этих белков. Семейства, к которым относятся гликозидазы, имеющие или предположительно имеющие родственные третичные структуры, было предложено объединять в суперсемейства [61]. Число суперсемейств постепенно возрастало. Это было обусловлено тремя причинами. Во-первых, постоянно увеличивается число гликозидаз с известными пространственными структурами. Во-вторых, в базах данных появляются новые последовательности, обладающие статистически достоверным сходством с представителями сразу двух семейств (тут действует правило «друзья моих друзей — мои друзья» [61]). В-третьих, вследствие повышения чувствительности методов компьютерного анализа последовательностей выявляются сходства между гликозидазами разных семейств.

Общепринятая классификация гликозил-гидролаз (CAZy), основанная на гомологии последовательностей, принимает во внимание сходство между белками только тех семейств,

которые характеризуются одинаковым механизмом гидролиза гликозидной связи [11, 46, 58, 66]. По аналогии с классификацией протеаз [1, 67] такие семейства объединены в кланы, которых к настоящему времени известно четырнадцать (GH-A—GH-N, табл. 2). Полагают, что все гликозидазы, относящиеся к семействам одного клана, имеют общее эволюционное происхождение, сходную трехмерную структуру, консервативное расположение каталитических остатков (т.е. в гомологичных положениях) и одинаковый механизм гидролиза гликозидных связей [42, 46, 47, 58–60, 63, 66].

Хотя в пределах одного семейства гликозидаз, как правило, не бывает ферментов, обладающих разными механизмами гидролиза гликозидной связи (кроме семейств GH23 и GH97), экспериментально было показано, что механизм гидролиза может быть изменен на обратный путем замены лишь одного аминокислот-

Таблица 1. Развитие классификации гликозил-гидролаз

Год (ссылка)	Число белков	Число семейств (в т.ч. в составе кланов)	Число кланов
1989 [55]	21	6 (0)	—
1991 [56]	322	36 (0)	—
1993 [57]	482	45 (0)	—
1995 [41, 61]	...	52 (17)	5
1996 [58]	~950	57 (17)	5
1997 [46, 60]	~1500	62 (19)	5
1998 [11]	~2200	70 (26)	8
1999 [11]	>2800	77 (...)	10
2000 [11]	...	82 (33)	10
2001 [11]	...	85 (37)	11
2002 [11]	~8000	87 (40)	12
2003 [11]	...	91 (43)	13
2004 [11]	...	95 (46)	14
2005 [11]	>20 000	101 (46)	14
2006 [11]	~30 000	106 (46)	14
2007 [11]	>30 000	110 (46)	14
2008 [11]	>40 000	114 (49)	14
2009 [11]	>59 000	115 (50)	14
2010 [11]	~82 000	118 (50)	14
2011 [11]	~86 000	123 (50)	14

Примечание. За 1989 г. указаны данные по классификации только  $\beta$ -1,4-гликозидаз [55]. За 2011 г. статистика дана по состоянию на 15 февраля 2011 г. Трехточие означает отсутствие данных, а прочерк — отсутствие классификации.

Таблица 2. Клань гликозил-гидролаз

Клан	Семейства (GH)	Аномерная конфигурация	Трехмерная структура	Ссылка
GH-A	1, 2, 5, 10, 17, 26, 30, 35, 39, 42, 50, 51, 53, 59, 72, 79, 86, 113	сохраняется (экв.)	( $\beta/\alpha$ ) <sub>8</sub> -бочонок	7, 10, 11, 25, 40, 46, 48, 57, 59, 66, 88, 94, 95, 96, 99, 150, 151
GH-B	7, 16	сохраняется (экв.)	$\beta$ -сэндвич	152
GH-C	11, 12	сохраняется (экв.)	$\beta$ -сэндвич	19, 153
GH-D	27, 31, 36	сохраняется (акс.)	( $\beta/\alpha$ ) <sub>8</sub> -бочонок	16, 56, 61, 75, 79, 103–106, 154, 155
GH-E	33, 34, 83, 93	сохраняется (экв.)	6-лопастной $\beta$ -пропеллер	11, 156
GH-F	43, 62	меняется (экв.)	5-лопастной $\beta$ -пропеллер	78, 130–132
GH-G	37, 63	меняется (акс.)	( $\alpha/\alpha$ ) <sub>6</sub> -бочонок	99
GH-H	13, 70, 77	сохраняется (акс.)	( $\beta/\alpha$ ) <sub>8</sub> -бочонок	79, 108, 111, 157–162
GH-I	24, 46, 80	меняется (экв.)	$\alpha + \beta$ -лизозим	11, 140–142
GH-J	32, 68	сохраняется ( $\beta$ -фуранозид)	5-лопастной $\beta$ -пропеллер	57, 78, 104, 129–132, 163, 164
GH-K	18, 20, 85	сохраняется (экв.)	( $\beta/\alpha$ ) <sub>8</sub> -бочонок	11, 48, 88, 95
GH-L	15, 65, 125	меняется (акс.)	( $\alpha/\alpha$ ) <sub>6</sub> -бочонок	11, 49, 99
GH-M	8, 48	меняется (экв.)	( $\alpha/\alpha$ ) <sub>6</sub> -бочонок	99
GH-N	28, 49	меняется (акс.)	( $\beta$ ) <sub>3</sub> -соленоид	95, 99, 144, 145

Примечание. В графе «Аномерная конфигурация» указан тип гликозидной связи субстрата (экваториальный или аксиальный) и его изменение в процессе реакции гидролиза.

ного остатка. Замена у гликозидаз, гидролизующих субстрат с сохранением его аномерной конфигурации, нуклеофильного карбоксилсодержащего остатка (Asp/Glu) активного центра на ненуклеофильный (Ala или Gly) приводит к инактивации фермента. Однако при добавлении в реакционную смесь молекул нуклеофила (азид, ацетата или формиата) механизм гидролиза 2,4-динитрофенильных гликозидов меняется на обратный в сравнении с ферментом дикого типа: продуктом такой реакции является гликозил-азид (или ацетат, или формиат) с обращенной аномерной конфигурацией [44, 51, 68, 69].

Замена остатка Thr-26 в активном центре лизоцима бактериофага T4 (семейство GH24 гликозидаз) на нуклеофильный остаток (Glu или His) привела к изменению механизма катализируемой реакции. Она стала идти путем двойного замещения, приводящего к конечному сохранению  $\beta$ -аномерной конфигурации субстрата у продукта [52, 70, 71].

Все клань гликозил-гидролаз содержат, по крайней мере, одно семейство, включающее в себя гликозидазы с известными пространственными структурами (табл. 2). Обособленность каждого из этих кланов, а также принадлежность к ним конкретных семейств, как правило, не вызывает сомнений. При этом большое число семейств остается не включенными в состав кланов, а в течение последних семи лет не было выделено ни одного нового клана (табл. 1). Таким образом, классификация CAZy в настоящее время является двухуровневой (семейство и клан) [11]. Лишь в случае семейств GH13 и GH30 выделен уровень подсемейства. Уже более двух лет назад (публикация он-лайн от 5 октября 2008 г. [12]) появилось сообщение о разбивании семейств GH1, GH2 и GH5 на подсемейства, однако это так и не было реализовано на сайте базы данных [11]. Более того, в нерцензуемых (non-peer-reviewed) публикациях авторов CAZy уже в 2005–2006 гг. сообщалось о вы-

делении подсемейств в 19 семействах гликозил-гидролаз [72, 73]. Однако лишь в случае семейств GH1, GH7, GH13 и GH15 были опубликованы данные, позволяющие читателям хотя бы в какой-то степени различать представителей разных подсемейств [73].

### ИЕРАРХИЧЕСКАЯ КЛАССИФИКАЦИЯ ГЛИКОЗИЛ-ГИДРОЛАЗ

В 1988 г. было впервые продемонстрировано [74] сходство аминокислотных последовательностей гликозил-гидролаз, затем отнесенных к семействам (GH13 и GH31), входящим в различные кланы (GH-N и GH-D соответственно). Родство семейств GH13 и GH31 впоследствии было подтверждено [60, 75–77] и даже обсуждалась целесообразность объединения их в один клан [60]. В 1990 г., еще до создания универсальной классификации всех гликозил-гидролаз [56], в составе четырех, из восьми выделенных к тому времени семейств, было предложено различать подсемейства [53]. В 2000 г. нами впервые была предложена иерархическая (подсемейство/семейство/клан/суперсемейство) классификация гликозил-гидролаз на примере  $\beta$ -фруктозидазного суперсемейства [78]. В 2004 г. аналогичная классификация была предложена для  $\alpha$ -галактозидазного и  $\alpha$ -глюкозидазного суперсемейств [79], а в 2006 г. она была распространена на все гликозил-гидролазы с каталитическим доменом в виде  $(\beta/\alpha)_8$ -бочонка с одновременным увеличением числа уровней в классификации до семи [80]. Верхними уровнями в этой классификации являются «фолд» и «группа близких фолдов», в которых от каталитических доменов наряду с предполагаемой общностью эволюционного происхождения требуется только однотипность пространственных структур. При этом уровень «фолда» в предложенной нами классификации примерно соответствует уровню «клана» в классификации Pfam [81] и уровню «суперсемейства» в классификации SCOP [82]. Недавно была опубликована работа [83], в которой 50 семейств каталитических доменов гликозил-гидролаз было предложено объединить в семь C-семейств (C1–C7).

**Гликозидазы с каталитическим доменом в виде  $(\beta/\alpha)_8$ -бочонка.** Домен с этим типом пространственной структуры (PDB, 1TIM) впервые был обнаружен [84] у триозофосфатизомеразы (К.Ф. 5.3.1.1) и поэтому часто именуется TIM-бочонком. Этот тип структуры является одним из самых распространенных у белков [85–87], а у ферментов – это наиболее часто встречающийся тип структуры каталитических доменов

[86]. У гликозил-гидролаз на его долю приходится около половины известных семейств [11, 88]. Среди всех семейств функционально охарактеризованных белков с трехмерной структурой в виде  $(\beta/\alpha)_8$ -бочонка подавляющее большинство содержат ферменты [89, 90]. При этом представители около половины семейств проявляют гидролазные активности, в основном – O-гликозил-гидролазные [88, 89]. Следует отметить, что большинство известных науке гликозил-гидролазных активностей удалось обнаружить у ферментов с каталитическим доменом в виде  $(\beta/\alpha)_8$ -бочонка [88].

Считается, что все домены с этим типом 3D структуры обладают общностью происхождения [17, 90]: ген, который кодирует их общего эволюционного предшественника, предположительно возник путем дубликации гена, кодировавшего  $(\beta/\alpha)_4$ -структуру [85, 86, 91, 92], или даже путем двух последовательных дубликаций гена, кодировавшего  $(\beta/\alpha)_2$ -структуру [93]. Это позволяет рассматривать все  $(\beta/\alpha)_8$ -домены гликозил-гидролаз как эволюционно родственные, хотя они, по всей видимости, не образуют монофилетической группы. Наиболее дивергентными являются три семейства гликозил-гидролаз (GH6, GH38 и GH57), имеющих необычную «незавершенную»  $(\beta/\alpha)_7$ -структуру. В предложенной нами иерархической классификации [80] эти три семейства отнесены к отдельному «фолду», в то время как большинство остальных семейств сгруппированы в четыре основных «типа» в рамках единого «фолда». При этом домены разных семейств одного «типа» при итеративном скрининге с помощью программы PSI-BLAST находят (или ожидается, что будут находить – где это еще не проверялось) друг друга в большинстве комбинаций. В классификации SCOP гликозидазы с каталитическими доменами в виде  $(\beta/\alpha)_8$ - и  $(\beta/\alpha)_7$ -бочонков отнесены соответственно к фолдам «TIM beta/alpha-barrel» (суперсемейство «(Trans)glycosidases») и «7-stranded beta/alpha barrel» (суперсемейства «Glycosyl hydrolases family 6, cellulases» и «Glycoside hydrolase/deacetylase») [82], а в классификации Pfam – к кланам CL0058 и CL0158 [81]. В классификации C-семейств они объединены вместе в семейство C3, в пределах которого выделено 27 субгрупп [83]. В литературе высказывались сомнения в общности эволюционного происхождения  $(\beta/\alpha)_8$ - и  $(\beta/\alpha)_7$ -гликозидаз [89]. В классификации CAZy [11] часть семейств  $(\beta/\alpha)_8$ -гликозидаз объединено в четыре клана (GH-A, GH-D, GH-N и GH-K), при этом все они содержат только ферменты, которые сохраняют аномальную конфигурацию субстрата в процессе катализируемой ими реакции (табл. 2). В работе На-

гано и соавт. 2001 г. [48], девять GH-семейств (в том числе, представители трех различных кланов) гликозидаз с экспериментально определенными на тот момент пространственными структурами ( $\beta/\alpha$ )<sub>8</sub>-типа были объединены в шесть функциональных субгрупп (F1–F6) и на более высоком иерархическом уровне – в четыре структурные группы (S1–S4).

**Тип I классических ( $\beta/\alpha$ )<sub>8</sub>-гликозидаз.** К этому типу относится наиболее многочисленный клан GH-A, который в настоящее время объединяет 18 семейств гликозил-гидролаз (табл. 2). Соответствующие ферменты используют субстраты с экваториально ориентированными гликозидными связями и сохраняют аномерную конфигурацию субстрата у продукта. Клан GH-A также известен как «суперсемейство 4/7» [94, 95]; по расположению двух каталитически важных аминокислотных остатков Glu – донора/акцептора протона и нуклеофила – в конце четвертого и седьмого  $\beta$ -слоев ( $\beta/\alpha$ )<sub>8</sub>-бочонка соответственно. К этому типу также было отнесено еще несколько семейств гликозил-гидролаз: GH14, GH29, GH44 и GH71 [80]. При этом семейство GH14 было включено на основе его высокого сходства с семействами GH5 [48], GH10 [40, 94, 95] и GH42 [81, 88], а семейства GH29, GH44 и GH71 – с семействами GH2 и GH5 [88]. Белки семейства GH44 также проявляют сходство с представителями семейства GH30 [96]. В состав этого типа также могут быть включены семейство GH98 на основании сходства с GH1 и GH5 [97] и семейство GH99 на основании сходства с GH71 (неопубликованные данные). Следует отметить, что гликозидазы семейств GH14, GH71 и GH98 меняют аномерную конфигурацию субстрата в процессе реакции гидролиза. Семейства клана GH-A (GH1, GH2, GH5, GH10 и GH17) вошли в состав функциональной субгруппы F2, а семейство GH14 – субгруппы F3. Эти две субгруппы образуют структурную группу S2 [48]. Таким образом, тип I классических ( $\beta/\alpha$ )<sub>8</sub>-гликозидаз примерно соответствует группе S2.

Для семейства GH1 было предложено три независимых разбиения на подсемейства: согласно одному из них выделено пять подсемейств [25], а согласно двум другим – по четырнадцать [62, 73]. Было показано, что семейство GH2 состоит из нескольких обособленных групп белков [40]. Семейство GH5 сначала разбили на пять подсемейств [53], а впоследствии их число достигло восьми [98]. В семействе GH30 было выделено две группы белков, состоящих из восьми подгрупп [96], которые теперь рассматриваются в классификации CAZy как восемь подсемейств [11]. Сравнение гликозидаз из семейств GH30 и

GH59 показало их высокое сходство [25], что позволяет рассматривать эти два семейства как подсемейства одного семейства GH30/59. Недавно была изменена граница между семействами GH5 и GH30 путем переноса части подсемейств из семейства GH5 в GH30 [11, 96].

Итеративный скрининг базы данных аминокислотных последовательностей с помощью белков семейства GH5 позволил проследить их эволюционные связи с представителями семейств GH13 и GH14 гликозил-гидролаз [48], скрининг с помощью белков семейства GH14 – с белками из GH35, GH42 и ряда других семейств клана GH-A [99], а скрининг с помощью белков семейства GH50 – с белками из семейства GH42 [88].

**Тип II классических ( $\beta/\alpha$ )<sub>8</sub>-гликозидаз.** К этому типу относятся  $\alpha$ -галактозидазное (клан GH-D; табл. 2) и  $\alpha$ -глюкозидазное (клан GH-H) суперсемейства, а также семейства GH66, GH97, GH101 и GH114 гликозил-гидролаз [80, 100]. Ферменты этого типа действуют на аксиально ориентированную гликозидную связь и почти все они (исключение составляет лишь семейство GH97) сохраняют аномерную конфигурацию субстрата в процессе гидролиза [11]. На основании гомологии к этому типу также может быть отнесен целый ряд семейств энзиматически неохарактеризованных белков (гипотетических гликозидаз) [75, 80, 101, 102]. Вопрос о необходимости объединения кланов GH-D и GH-H на более высоком иерархическом уровне обсуждался в работе [77]. Семейство GH13 гликозил-гидролаз было отнесено к функциональной субгруппе F1, образующей структурную группу S1 [48]. Поскольку это было единственное семейство из типа II классических ( $\beta/\alpha$ )<sub>8</sub>-гликозидаз для которого тогда уже была известна пространственная структура, то мы можем условно отождествить тип II с группой S1.

**$\alpha$ -Галактозидазное суперсемейство (клан GH-D).** Впервые название « $\alpha$ -галактозидазное суперсемейство» было предложено в 2001 г. (ранее использовали название «семейство DAG» [103]) для объединения семейств GH27 и GH36 гликозил-гидролаз [104], которые составляли клан GH-D [11] и включали в себя большинство известных тогда  $\alpha$ -галактозидаз (К.Ф. 3.2.1.22). В то время семейства GH27 и GH36 были достаточно четко обособлены между собой [103]. В пределах семейства GH36 тогда [104] удавалось выделить два подсемейства (позже названные GH36A и GH36B [105]). В 2002 г. [106] в состав этого суперсемейства было предложено включить семейства GH31 и aGalT (ныне GH36C [105]). Семейство aGalT (от англ. « $\alpha$ -galactosyl-transferase») было составлено [106] из группы

белков преимущественно растительного происхождения, для некоторых из которых были продемонстрированы  $\alpha$ -галактозилтрансферазные активности (К.Ф. 2.4.1.67 и К.Ф. 2.4.1.82). После опубликования геномной последовательности *Sulfolobus solfataricus* (GenBank, AE006641) летом 2002 г. кодируемый этой археей белок (GenPept, AAK43227) вместе с рядом других представителей семейства aGalT был включен в классификации CAZy в состав семейства GH36 [11]. Однако по уровню сходства аминокислотных последовательностей белки семейства aGalT примерно равноудалены от семейств GH27 и GH36 гликозил-гидролаз, имея с ними не более 20% идентичных аминокислотных остатков. Такое расширенное трактование семейства GH36 впоследствии обусловило попадание в его состав еще большего числа лишь отдаленно родственных белков. Например, секвенирование гена  $\alpha$ -N-ацетилгалактозаминидазы (К.Ф. 3.2.1.49)

из *Clostridium perfringens* [107] привело к включению в это семейство еще одной группы гомологов (ныне GH36D [105]). После этого в классификации CAZy [11] в состав семейства GH36 стали включать практически все белки клана GH-D, кроме искусственно обособленного, но сравнительно гомогенного семейства GH27. Уже в 2003 г. было констатировано отсутствие монофилетического статуса у белков, относимых тогда к семейству GH36, и предложено его разбиение на четыре подсемейства GH36A-GH36D [105]. В 2004 г. [79] эти подсемейства было предложено рассматривать в качестве самостоятельных семейств в составе  $\alpha$ -галактозидазного суперсемейства (клана GH-D). К настоящему времени уже выделено одиннадцать семейств (GH36A-GH36K; табл. 3) на основе бывшего семейства GH36 [16, 75, 100, 101], однако только часть из них содержит белки с экспериментально определенными энзиматическими

Таблица 3. Разбиение семейства GH36 на 11 новых семейств (GH36A-GH36K)

Семейство белков	Представитель					Всего белков	
	Номер в GenPept	Организм	Длина	Фрагмент	Аннотация	GenPept	CAZy
GH36A	AAG49420.1	<i>Geobacillus stearothermophilus</i>	729	325–626	$\alpha$ -galactosidase AgaA	795	351
GH36B	CAA04514.1	<i>Thermotoga maritima</i>	552	180–482	$\alpha$ -galactosidase	104	50
GH36C	AAD02832.1	<i>Cucumis sativus</i>	784	198–568	raffinose synthase	281	127
GH36D	AAM55479.1	<i>Clostridium perfringens</i>	629	213–538	$\alpha$ -N-acetylgalactosaminidase	64	26
GH36E	EFB00895.1	<i>Victivallis vadensis</i>	654	268–556	hypothetical protein	20	1
GH36F	EDM35914.1	<i>Pedobacter</i> sp. BAL39	684	278–590	$\alpha$ -galactosidase	29	8
GH36G	ABF39669.1	<i>Candidatus Koribacter versatilis</i>	684	279–593	glycoside hydrolase, clan GH-D	9	3 + 1
GH36H	BAC74977.1	<i>Streptomyces avermitilis</i>	610	196–494	hypothetical protein	15	3
GH36I	ABW01612.1	<i>Caldivirga maquilingsensis</i>	708	214–530	hypothetical protein	15	0
GH36J	ADB29787.1	<i>Kribbella flavida</i>	569	256–521	hypothetical protein	1	0
GH36K	CBL19648.1	<i>Ruminococcus</i> sp. SR1/5	561	201–500	$\alpha$ -galactosidase	6	3

Примечание. Длина – суммарное число аминокислотных остатков в белке-предшественнике. Фрагмент – порядковые номера первого и последнего аминокислотного остатка, соответствующих предполагаемому каталитическому домену. Всего белков – количество неидентичных аминокислотных последовательностей в базе данных, содержащих домен данного семейства. GenPept – количество белков, обнаруженных в базе данных GenPept (NCBI) при помощи поиска по гомологии против «non-redundant protein sequences», проведенного 13 января 2011 г. CAZy – количество белков из предыдущей колонки, представленных в базе данных CAZy [11] по состоянию на то же число. Все они в CAZy отнесены к семейству GH36. Исключение составляет один из четырех белков семейства GH36G (GenPept, CAR57390.1), который находится в списке неклассифицированных гликозил-гидролаз (известных также как условное семейство GH0).



активностями. В пределах семейства GH27 сначала было выделено три подсемейства [16, 79], а потом их число возросло до шести [100]. В пределах семейства GH31 было предварительно выделено 38 подсемейств [75, 100], но часть из них, вероятно, будет объединена в более крупные, так как они не обладают монофилетическим статусом. Около трех лет назад в классификации CAZy семейство GH31 было включено в состав клана GH-D наряду с семействами GH27 и GH36 [11]. В классификации Pfam клану GH-D соответствует три семейства: PF01055 (GH31), PF02065 (GH27 и частично GH36) и PF05691 (GH36C) [81].

Итеративный скрининг базы данных аминокислотных последовательностей с помощью белков  $\alpha$ -галактозидазного суперсемейства позволил обнаружить их эволюционные связи с целым рядом белков из других семейств: GH5, GH13, GH66, GH97, GH101, GH114, COG1306, COG1649 и COG2115 [17, 75, 76, 79, 80, 100]. Следует отметить, что COG2115 является семейством ксилозиомераз (К.Ф. 5.3.1.5).

#### *$\alpha$ -Глюкозидазное суперсемейство (клан GH-H).*

Впервые название « $\alpha$ -глюкозидазное суперсемейство» было предложено в 2004 г. для объединения семейств GH13, GH70 и GH77 гликозил-гидролаз, образующих клан GH-H [79]. Согласно базам данных ортологичных групп белков микроорганизмов (COG) и эукариот (KOG), белки семейства GH13 в обоих случаях объединены в четыре кластера: COG0296, COG0366, COG1523, COG3280, KOG0470, KOG0471, KOG2212 и KOG3625 [108]. Семейство GH13 является одним из самых многочисленных среди семейств гликозил-гидролаз (>9000 белков [11]). К нему относятся ферменты с 22 типами энзиматических активностей (в т.ч. гидролазы, трансферазы и синтазы), а также транспортеры аминокислот [11]. Это – одно из двух семейств гликозил-гидролаз, для которых на сайте CAZy дано разбиение на подсемейства: 36 подсемейств (GH13\_1–GH13\_36) объединяют часть белков из семейства GH13 [11]. Исходно, в 2006 г. было предложено выделить 35 подсемейств [109], однако потом их число возросло до 40 [73]. Сравнение белков 36 подсемейств, представленных на сайте CAZy [11], позволило показать [108], что уровень различий между белками разных подсемейств неодинаков. Три пары семейств – GH13\_15 и GH13\_24, GH13\_20 и GH13\_21, GH13\_29 и GH13\_31 – оказались очень сходны между собой и могут рассматриваться как три подсемейства: GH13\_15/24 (или KOG2212), GH13\_20/21 и GH13\_29/31. Подсемейства GH13\_25 и GH13\_33 оказались эволюционно более далеки от остальных подсемейств

этого семейства, чем семейство GH70, что позволило рассматривать GH13\_25 (KOG3625 или GL3R2899 в базе данных Genolevures) и GH13\_33 как самостоятельные семейства в составе  $\alpha$ -глюкозидазного суперсемейства. В пределах семейства GH13 было предложено выделить еще десять подсемейств: GH13\_A–GH13\_G [108]. В качестве одиннадцатого подсемейства можно рассматривать семейство GL3C0220 из базы данных дрожжевых геномов Genolevures [14, 110]. В семействе GH77 было выделено 11 подсемейств [100], а разбиение семейства GH70 на подсемейства было признано нецелесообразным вследствие его высокой гомогенности [111].

Итеративный скрининг базы данных аминокислотных последовательностей с помощью белков семейства GH13 позволил обнаружить их эволюционные связи с целым рядом белков из других семейств: GH31, GH36D, GH70, COG1649, COG2342, GHL4 и GHL5 [101, 108]. По уровню сходства аминокислотных последовательностей [102] семейство GHL5 экспериментально неохарактеризованных белков из прокариот может быть включено в состав  $\alpha$ -глюкозидазного суперсемейства наряду с семействами GH13, GH13\_25, GH13\_33, GH70 и GH77 гликозил-гидролаз.

*Семейство GH97.* Это – уникальное семейство, объединяющее «сохраняющие»  $\alpha$ -галактозидазы (К.Ф. 3.2.1.22) и «обращающие»  $\alpha$ -глюкозидазы (К.Ф. 3.2.1.20) [11, 64, 69, 112, 113]. В пределах семейства GH97 удалось выделить пять подсемейств [17, 114]. Итеративный скрининг базы данных аминокислотных последовательностей с помощью белков семейства GH97 позволил обнаружить [17] их эволюционные связи с целым рядом белков из других семейств гликозил-гидролаз: GH13, GH20, GH27, GH31 и GH36, а также с белками из семейства COG0535 (семейство гипотетических Fe-S оксидоредуктаз). Анализ пространственных структур показал, что десятью наиболее похожими на GH97-домен структурами обладают гликозидазы из семейств: GH27, GH36, GH13, GH31, GH84 и GH5 [112].

*Семейство GH101.* Каталитические домены этого семейства имеют структуру искаженного  $(\beta/\alpha)_8$ -бочонка [102]. В пределах семейства GH101 удалось выделить шесть подсемейств [102, 115]. Итеративный скрининг базы данных аминокислотных последовательностей с помощью белков этого семейства позволил обнаружить их эволюционные связи с целым рядом белков из других семейств: GH13, GH20, GH27, GH29, GH31, GH36, GH66, GH70, GH97, COG1306, COG1649, GHL1, GHL2, GHL3,

GH14 и GH15 [101, 102, 115]. При этом эволюционно наиболее близким семейством гликозил-гидролаз оказалось семейство GH13 [102, 115]. По уровню сходства аминокислотных последовательностей семейство GH101 может быть объединено с тремя семействами GH11–GH13 экспериментально неохарактеризованных белков в одно суперсемейство.

**Семейства GH114 и COG2342.** Филогенетический анализ показал невозможность разбить семейство GH114 (или COG3868) на четко обособленные подсемейства [101, 116]. По уровню сходства аминокислотных последовательностей семейства GH114 и COG2342 могут быть объединены в одно суперсемейство. В классификации Pfam они рассматриваются как единое семейство PF03537 (DUF297) [81]. Итеративный скрининг базы данных аминокислотных последовательностей с помощью белков семейства GH114 позволил обнаружить их эволюционные связи с целым рядом белков из других семейств: GH13, GH18, GH20, GH27, GH29, GH31, GH35, GH36A, GH36B, GH36E, GH36F, GH36G, GH36H, GH36J, GH36K, GH66, COG1306, COG1649, COG2342, а также из еще 13 GH-семейств (GH13–GH15) [101]. Использование белков семейства COG2342 для итеративного скрининга базы данных позволило выявить их эволюционные связи с белками семейств GH5, GH13, GH20, GH27, GH29, GH31, GH35, GH36A, GH36B, GH36F, GH36G, GH36H, GH42, GH66, GH97, GH101, GH114, COG1306, COG1649, GH13 и GH14 [116].

**Семейство COG1649.** В пределах семейства COG1649 (DUF187) удалось выделить четыре основных подсемейства [80]. Обнаружены эволюционные связи с белками из семейств GH13, GH20, GH31 и GH36 гликозил-гидролаз [80], что позволяет его рассматривать в качестве семейства гипотетических гликозидаз.

**Семейства GH11–GH15.** Пятнадцать новых семейств GH11–GH15 гипотетических белков были обнаружены при итеративном поиске с помощью программы PSI-BLAST гомологов для гликозидаз семейств GH13 [108], GH101 [102] и GH114 [101]. Использование представителей этих 15 новых семейств, в свою очередь, позволило проследить их эволюционные связи как между собой, так и с белками из семейств: GH5, GH13, GH13\_33, GH17, GH18, GH20, GH27, GH29, GH31, GH35, GH36A, GH36B, GH36C, GH36D, GH36E, GH36F, GH36G, GH36H, GH36J, GH36K, GH39, GH42, GH53, GH66, GH97, GH101, GH107, GH112, GH114, DUF3111 (PF11308 [81]), COG1082, COG1306, COG1649 и COG2342 (неопубликованные данные).

**Тип III классических  $(\beta/\alpha)_8$ -гликозидаз.** К этому типу относятся клан GH-K (табл.2), а также семейства GH25 и GH84 гликозил-гидролаз [80]. При этом семейство GH25 было включено на основе его высокого сходства с семейством GH18 [82], а семейство GH84 – с GH20 [88]. Гликозидазы семейства GH18 относятся к функциональным субгруппам F4 и F5, образующим структурную группу S3, а ферменты семейства GH20 включены в субгруппу F6 (группа S4) [48]. Для белков семейства GH18 была предложена подробная иерархическая классификация, в основе которой лежит разбиение на три крупных подсемейства [27]. Эукариотические белки семейства GH20 образуют три подсемейства [117].

**Тип IV классических  $(\beta/\alpha)_8$ -гликозидаз.** К этому типу были отнесены семейства GH67 и GH89 гликозил-гидролаз [80] на основе данных, представленных в работе [88]. В пределах семейства GH67 было выделено три подсемейства [118].

**Другие семейства классических  $(\beta/\alpha)_8$ -гликозидаз.** Ни к одному из четырех выше упомянутых типов не отнесены семейства GH3 и GH56 гликозил-гидролаз [80], обладающие  $(\beta/\alpha)_8$ -структурой каталитических доменов. Семейство GH3 было разбито на три подсемейства, внутри одного из них было выделено четыре кластера, а в пределах трех из этих кластеров – восемь подкластеров белков [119].

Каталитические домены семейства GH112 имеют структуру «разорванного»  $(\beta/\alpha)_8$ -бочонка, структурно они наиболее близки к доменам семейства GH42 гликозил-гидролаз [120]. Обнаружено отдаленное сходство белков семейств GH29, GH14, GH16 и GH11 с гликозидазами семейства GH107 (неопубликованные данные), что позволяет предполагать, что каталитический домен семейства GH107 имеет структуру  $(\beta/\alpha)_8$ -бочонка. Предсказанная вторичная структура  $\beta$ -N-ацетилгалактозаминидазы из *Paenibacillus* sp. указывает на то, что каталитический домен этого фермента, относящийся к недавно выделенному семейству GH123, также имеет третичную структуру в виде  $(\beta/\alpha)_8$ -бочонка [121].

В рамках классификации гликозил-гидролаз имеет смысл также рассматривать семейство PF11790 энзиматически неохарактеризованных белков, которое включено в классификации Pfam в клан CL0058 и наиболее близко к семейству GH39 [81].

**Семейства гликозидаз с «незавершенной»  $(\beta/\alpha)_8$ -структурой.** Как уже было отмечено выше, каталитические домены гликозидаз, имеющие этот необычный вариант 3D структуры в базах данных SCOP и Pfam классифицируются независимо от доменов гликозидаз с классичес-

кими ( $\beta/\alpha$ )<sub>8</sub>-бочонками. При этом, в SCOP домены семейств GH6, GH38 и GH57 отнесены к фолду «7-stranded beta/alpha barrel» [82], а в классификации Pfam в состав клана CL0158 попали только семейства GH38 и GH57, в то время как семейство GH6 (PF01341) не включено ни в один из существующих кланов [81]. К этой группе из трех семейств вероятно можно еще добавить недавно выделенное семейство GH119, в отношении которого имеются указания о его близком родстве с GH57 [11].

С помощью программы PSI-BLAST удалось показать родство семейств GH38 и GH57, при этом каталитически важный нуклеофильный остаток находится в соответствующих белках в гомологичном положении [122]. Также имеются данные о гомологии гликозидаз из семейств GH13 и GH57 [123], что указывает на общность эволюционного происхождения ( $\beta/\alpha$ )<sub>7</sub>- и ( $\beta/\alpha$ )<sub>8</sub>-бочонков.

Семейства II и III  $\alpha$ -маннозидаз [124] можно рассматривать как два подсемейства в составе семейства GH38. Семь основных подсемейств было выделено в семействе GH57 [125]. Семейство GH6 было разбито на восемь подсемейств [126].

Было показано сходство пространственных структур гликозидаз семейства GH6 и ксилозоизомераз [40].

**Гликозидазы с каталитическим доменом в виде  $\beta$ -пропеллера.** Домены с этим типом пространственной структуры могут быть разбиты на пять вариантов в зависимости от числа лопастей в  $\beta$ -пропеллере: от четырех до восьми. В классификации SCOP эти пять вариантов рассматриваются как пять различных фолдов [82]. Считается, что все или почти все  $\beta$ -пропеллеры, вне зависимости от числа лопастей в них, являются гомологами [127]. При этом ген их общего эволюционного предшественника кодировал белок, соответствующий одной лопасти  $\beta$ -пропеллера [127, 128]. Среди каталитических доменов гликозил-гидролаз представлены  $\beta$ -пропеллеры с пятью, шестью и семью лопастями. Почти все они образуют компактную (предположительно монофилетическую) группу среди всех  $\beta$ -пропеллеров, при этом гликозидазы с равным числом лопастей в пропеллере как правило сильнее похожи друг на друга [127]. В классификации CAZy [11] они отнесены к 11 семействам, часть из которых образует три клана (GH-E, GH-F и GH-J; табл. 2). В классификации C-семейств они объединены вместе в семейство C7 [83].

*Пяти-лопастные  $\beta$ -пропеллеры (фуранозидазное суперсемейство).* Этот вариант  $\beta$ -пропеллеров очень редкий и встречается почти исключительно среди каталитических доменов гликозил-гидролаз. В зависимости от используемого меха-

низма гликозидазы этой группы относятся к кланам GH-F («обращающий» механизм) и GH-J («сохраняющий»). Исторически — это первая группа гликозил-гидролаз, для которой была предложена иерархическая классификация [78].

Проведенное почти полтора десятилетия назад сравнение аминокислотных последовательностей  $\beta$ -фруктозидаз (семейство GH32) и  $\beta$ -фруктозилтрансфераз (впоследствии отнесены к семейству GH68) выявило в них многочисленные участки локального сходства, что позволило их объединить в  $\beta$ -фруктозидазное суперсемейство [129]. Последнее нашло свое отражение в классификации CAZy [11], где эти два семейства образовали клан GH-J (табл. 2). Множественное выравнивание аминокислотных последовательностей всех белков семейства GH68 показало, что наиболее протяженным консервативным участком является фрагмент последовательности L13, содержащий донор протона из активного центра фермента [130]. Использование «усредненной» (т.е. консенсусной) последовательности фрагмента L13 аминокислотной последовательности  $\beta$ -фруктозилтрансфераз позволило с помощью программы rblast обнаружить его гомологию с участком последовательностей гликозил-гидролаз семейства GH43 [131]. К тому времени семейства GH43 и GH62 уже были объединены в клан GH-F [11], таким образом, впервые появились основания предполагать об эволюционном родстве гликозидаз двух разных кланов (GH-F и GH-J). Использование белков семейства GH43 в качестве запроса при итеративном скрининге базы данных аминокислотных последовательностей подтвердило их гомологию с гликозил-гидролазами из семейств GH32, GH62 и GH68, а также с экспериментально неохарактеризованными белками из нового семейства GHLP (впоследствии получило названия COG2152 и DUF377) [78, 130, 131]. На основании полученных данных семейства клана GH-F были включены в состав  $\beta$ -фруктозидазного суперсемейства. Учитывая то, что общим свойством ферментов всех семейств этого суперсемейства является способность отщеплять от субстрата моносахаридные остатки в фуранозидазной форме ( $\alpha$ -L-арабинофуранозидазные в случае семейств клана GH-F и  $\beta$ -D-фруктофуранозидазные в случае семейств клана GH-J), оно было переименовано в фуранозидазное суперсемейство [130]. В состав этого суперсемейства на основании гомологии также было включено семейство GHLP [78, 130, 131]. В классификации Pfam это суперсемейство соответствует клану CL0143 [81], а в классификации SCOP — суперсемейству «Arabinanase/levansucrase/invertase», в состав которого также включено семей-

ство DUF1861 [82]. Семейство DUF1861 эволюционно наиболее близко к семейству GHLP (неопубликованные данные).

В составе семейств GH32, GH43 и GH68 было выделено четыре, семь и два основных подсемейств соответственно [130, 132]. Однако два наиболее дивергентных белка семейства GH43 не были отнесены к конкретным подсемействам. Среди них — экспериментально неохарактеризованный белок из бактерии *Streptomyces coelicolor* (GenPept, CAB61805.1), который впоследствии был включен в состав недавно образованного семейства GH117 гликозил-гидролаз [11], но в классификации Pfam семейства GH43 и GH117 продолжают рассматриваться как одно (PF04616) [81].

**Шести- и семилопастные  $\beta$ -пропеллеры.** Шестилопастные  $\beta$ -пропеллеры формируют каталитические домены гликозил-гидролаз клана GH-E, объединяющего четыре семейства: GH33, GH34, GH83 и GH93 (табл. 2). Такую же структуру имеют и гликозидазы семейства GH58. Семилопастные  $\beta$ -пропеллеры обнаружены в семействе GH74. В классификации Pfam все эти семейства рассматриваются как клан CL0434 [81]. Сравнение пространственных структур изолированных лопастей  $\beta$ -пропеллеров выявляет их высокое сходство у гликозидаз семейств GH33, GH58 и GH74 [127], подтверждая обоснованность объединения шести- и семилопастных  $\beta$ -пропеллеров в одну группу. В классификации SCOP гликозил-гидролазы со структурой шести- и семилопастных  $\beta$ -пропеллеров отнесены соответственно к суперсемействам «Sialidases» и «Oligoxyloglucan reducing end-specific cellobiohydrolase» [82].

**Гликозидазы с каталитическим доменом в виде  $\beta$ -сэндвича ( $\beta$ -jelly roll fold).** Этот тип пространственной структуры представлен в гликозил-гидролазах двух кланов — GH-V и GH-C — оба они объединяют ферменты, действующие на экваториально ориентированные гликозильные связи с сохранением их аномерной конфигурации у продуктов гидролиза (табл. 2). В классификации Pfam семейства этих двух кланов отнесены к клану CL0004 [81], в классификации SCOP — к суперсемейству «Concanavalin A-like lectins/glucanases» [82], а в классификации C-семейств они отнесены к двум семействам: C4 и C5 [83]. Родство кланов GH-V и GH-C было продемонстрировано путем сравнения пространственных структур каталитических доменов семейств GH11 и GH16 [133]. Каталитический домен в виде  $\beta$ -сэндвича также обнаружен в семействе GH54 [134]. Для белков семейства GH16 была предложена подробная иерархическая классификация [133, 135], по несколько

подсемейств удается выделить в семействах GH7 и GH12 [22]. Альтернативная классификация выделяет в семействе GH7 только два подсемейства [73].

**Гликозидазы с каталитическим доменом в виде  $(\alpha/\alpha)_6$ -бочонка.** Этот тип пространственной структуры представлен в гликозил-гидролазах трех кланов — GH-G, GH-L и GH-M (табл. 2). Все соответствующие ферменты обращают аномерную конфигурацию субстрата у продуктов гидролиза. Такую же пространственную структуру имеют и гликозидазы семейств GH9, GH47, GH76, GH78, GH88, GH92, GH94 (ранее GT36), GH95, GH100, GH105 и GH116, а также полисахаридлиазы из семейства PL10. В классификации Pfam им соответствует клан CL0059 [81], а в классификации C-семейств — семейство C6 [83]. В классификации SCOP гликозидазы этой группы отнесены к двум суперсемействам — «Six-hairpin glycosidases» и «Seven-hairpin glycosidases» — в составе фолда «alpha/alpha toroid» [82]. Следует отметить, что каталитический домен семейства GH47 имеет модифицированную структуру  $(\alpha/\alpha)_7$ -бочонка. Филогенетический анализ позволяет выделить четыре группы белков в составе этого семейства [136]. В составе семейства GH8 было выделено три подсемейства [137], в GH15 — четыре подсемейства [73], а в семействе GH9 — три структурных субкласса растительных белков [138].

Итеративный скрининг базы данных аминокислотных последовательностей с помощью белков семейства GH78 позволил проследить их эволюционные связи с представителями семейств GH94, GH63, GH15, GH37, GH95 и GH65 гликозил-гидролаз. Статистически менее достоверные связи удалось также обнаружить с белками семейств GH92, GH9, GH8 и GH48 [99].

**Гликозидазы с каталитическим доменом лизоцимового типа.** Лизоцимы (К.Ф. 3.2.1.17) — обширная группа гликозил-гидролаз с особым типом пространственной структуры, характерной также для хитиназ (К.Ф. 3.2.1.14), хитозаназ (К.Ф. 3.2.1.132) и лактальбуминов. В классификации Pfam им соответствует клан CL0037 [81], в классификации SCOP — суперсемейство «Lysozyme-like» [82], а в классификации C-семейств — семейство C1 [83].

Каталитические домены рассматриваемой группы относятся к нескольким семействам гликозил-гидролаз: GH19, GH22, GH23, GH24, GH46, GH73, GH80, GH103, GH104, GH108 и GH124. В этой группе представлены ферменты как обращающие (GH19, GH24, GH46 и GH124), так и сохраняющие (GH22) экваториальную ориентацию гидролизуемой гликозильной связи у продуктов реакции [11]. Особый

случай представляет семейство GH23, содержащее «сохраняющие» пептидогликан-трансглюкозилазы и «обращающие» лизоцимы [64].

Лизоцимы исторически были первой группой гликозил-гидролаз, на примере которой была показана возможность эволюционного родства ферментов с разным механизмом катализируемой реакции [71, 139–141].

Следует отметить, что семейства GH24, GH46 и GH80 образуют клан GH-I (табл. 2). Более того, семейства GH46 и GH80 настолько близки, что аминокислотные последовательности относящихся к ним белков образуют континуум. Это позволило предложить рассматривать GH46 и GH80 как единое семейство [142]. Высокое сходство пространственных структур показано для белков семейств GH19 и GH23, что поставило вопрос об объединении этих двух семейств на более высоком иерархическом уровне [140]. В семействах GH22 и GH24 было выделено по два подсемейства [141].

Функционально слабоизученные белки семейства COG5526 структурно похожи на лизоцимы и вероятно участвуют в разрушении клеточной стенки бактерий [143].

**Гликозидазы с каталитическим доменом в виде правозакрученного ( $\beta$ )<sub>3</sub>-соленоида.** Считается, что все белковые домены с этим типом трехмерной структуры обладают общностью эволюционного происхождения [95, 144]. В классификации Pfam им соответствует клан CL0268 [81], в классификации SCOP – суперсемейство «Pectin lyase-like» [82], а в классификации C-семейств – семейство C2 [83]. Структуру ( $\beta$ )<sub>3</sub>-соленоида имеют гликозил-гидролазы семейств GH28, GH49, GH55, GH82, GH87, GH90 и GH91 (ранее PL19). При этом семейства GH28 и GH49 образуют клан GH-N (табл. 2). В классификации SCOP соответствующее суперсемейство также содержит представителей семейств PL1, PL3, PL6 и PL9 полисахаридлиаз и семейства CE8 карбогидрат-эстераз [82]. Структуру правозакрученного ( $\beta$ )<sub>3</sub>-соленоида, вероятно, имеют также гликозидазы семейства GH120, поскольку они обладают сходством с представителями семейства DUF1565 из клана CL0268 (неопубликованные данные).

Итеративный скрининг базы данных аминокислотных последовательностей с помощью белков семейства GH28 позволил проследить их эволюционные связи с гликозидазами семейств GH49, GH55, GH82 и GH87, а также с полисахаридлиазами семейства PL1 [99, 145], а с помощью белка семейства GH82 – с гликозидазами семейств GH28, GH55 и GH87 [145].

Филогенетический анализ позволил выделить несколько групп белков в пределах семейства GH28 гликозил-гидролаз [146].

**Другие семейства гликозил-гидролаз.** Выше были рассмотрены шесть основных «фолдов» и «групп фолдов» каталитических доменов гликозил-гидролаз. Однако некоторые гликозидазы обладают другими типами каталитических доменов. Так в классификации Pfam [81] есть два содержащих семейства гликозидаз клана – CL0063 и CL0199 – в которых большинство семейств не содержат ферментов с гликозил-гидролазными активностями.

Клан CL0063 включает в себя семейства GH4 (PF02056) и GH109 (PF01408), для которых характерен необычный для гликозил-гидролаз механизм с использованием NAD<sup>+</sup> в качестве кофактора [11]. Семейство GH4 уникально тем, что часть относящихся к нему гликозидаз использует субстраты с аксиально, а другая – с экваториально ориентированной гликозидной связью [11]. Филогенетический анализ позволил выделить в семействе GH4 четыре монофилетические группы белков, характеризующихся различающимися ферментативными активностями [147].

Клан CL0199 содержит семейства GH45 (PF02015) и GH102 (PF06725). В семействе GH45 было выделено пять подсемейств [148].

В заключение следует отметить, что почти все семейства каталитических доменов гликозил-гидролаз удалось объединить в небольшое число фолдов. По всей видимости, в пределах каждого из них все домены обладают общностью эволюционного происхождения. Однако несколько семейств остались неклассифицированными. Большинство из них – недавно обнаруженные семейства, для которых нет экспериментальных данных о пространственной структуре каталитических доменов. Дальнейшие исследования этих семейств позволят определить их положение в иерархической классификации гликозил-гидролаз. Например, недавно опубликовано предварительное сообщение о кристаллизации белка, содержащего домен семейства GH115 гликозил-гидролаз [149]. Это позволяет полагать, что в ближайшее время появятся данные о его пространственной структуре, необходимые для классификации.

После прохождения этой статьей стадии рецензирования в базе данных CAZy появились два новых семейства гликозил-гидролаз: GH124 и GH125. При этом семейство GH124 оказалось близкородственным семейству GH23, а семейство GH125 было включено в клан GH-L [11].

Работа выполнена при финансовой поддержке РФФИ (грант 11-04-00571-а).

## СПИСОК ЛИТЕРАТУРЫ

1. Barrett, A.J. (1994) *Methods Enzymol.*, **244**, 1–15.
2. Кочетков Н.К., Бочков А.Ф., Дмитриев Б.А., Усов А.И., Чижов О.С., Шибяев В.Н. (1967) *Химия углеводов*, Изд-во Химия, Москва.
3. Красиков В.В., Карелов Д.В., Фирсов Л.М. (2001) *Биохимия*, **66**, 332–348.
4. Sinnott, M.L. (1990) *Chem. Rev.*, **90**, 1171–1202.
5. Хорлин А.Я. (1974) В кн. *Структура и функции активных центров ферментов*, Изд-во Наука, Москва, с. 39–69.
6. Хорлин А.Я. (1988) В кн. *Химическая энциклопедия*, том 1, Изд-во Советская энциклопедия, Москва, с. 575–576.
7. Рабинович М.Л., Мельник М.С., Болобова А.В. (2002) *Прикладная биохимия и микробиология*, **38**, 355–373.
8. Видершайн Г.Я. (1980) *Биологические основы гликозидов*, Изд-во Медицина, Москва.
9. Рабинович М.Л., Мельник М.С., Болобова А.В. (2002) *Биохимия*, **67**, 1026–1050.
10. Рабинович М.Л., Мельник М.С. (2000) *Успехи биологической химии*, **40**, 205–266.
11. *Carbohydrate-Active Enzymes server* (2011), (<http://www.cazy.org/>).
12. Santarel, V.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) *Nucleic Acids Res.*, **37**, D233–D238.
13. Hazlewood, G.P., and Gilbert, H.J. (1998) *Biochem. Soc. Trans.*, **26**, 185–190.
14. Naumoff, D.G. (2010) В кн. *Доклады III международной конференции «Математическая биология и биоинформатика», Пушкино, 10–15 октября 2010*, Изд-во МАКС Пресс, Москва, с. 139–140.
15. Parrent, J.L., James, T.Y., Vasaitis, R., and Taylor, A.F.S. (2009) *BMC Evol. Biol.*, **9**, Art.148.
16. Наумов Д.Г. (2004) *Молекуляр. биология*, **38**, 463–476.
17. Naumoff, D.G. (2005) *BMC Genomics*, **6**, Art.112.
18. Turakainen, H., Aho, S., and Korhola, M. (1993) *Appl. Environ. Microbiol.*, **59**, 2622–2630.
19. Tomme, P., Warren, R.A.J., and Gilkes, N.R. (1995) *Adv. Microb. Physiol.*, **37**, 1–81.
20. Garcia-Vallve, S., Romeu, A., and Palau, J. (2000) *Mol. Biol. Evol.*, **17**, 352–361.
21. Tymowska-Lalanne, Z., and Kreis, M. (1998) *Adv. Botanical Res.*, **28**, 71–117.
22. Vlasenko, E., Schulein, M., Cherry, J., and Xu, F. (2010) *Bioresour. Technol.*, **101**, 2405–2411.
23. Coutinho, P.M., Stam, M., Blanc, E., and Henrissat, B. (2003) *Trends Plant Sci.*, **8**, 563–565.
24. Andrews, S.R., Charnock, S.J., Lakey, J.H., Davies, G.J., Claeysens, M., Nerinckx, W., Underwood, M., Sinnott, M.L., Warren, R.A., and Gilbert, H.J. (2000) *J. Biol. Chem.*, **275**, 23027–23033.
25. Mian, I.S. (1998) *Blood Cells Mol. Dis.*, **24**, 83–100.
26. Наумов Д.Г. (2007) *Молекуляр. биология*, **41**, 1056–1068.
27. Karlsson, M., and Stenlid, J. (2009) *J. Mol. Microbiol. Biotechnol.*, **16**, 208–223.
28. Park, B.H., Karpinets, T.V., Syed, M.H., Leuze, M.R., and Uberbacher, E.C. (2010) *Glycobiology*, **20**, 1574–1584.
29. Coutinho, P.M., and Henrissat, B. (1999) in *Genetics, biochemistry and ecology of cellulose degradation* (Ohmiya, K., Hayashi, K., Sakka, K., Kobayashi, Y., Karita, S., and Kimura, T., ed.), Uni. Publishers Co., Tokyo, pp. 15–23.
30. Buchner, E. (1897) *Ber. Dt. Chem. Ges.*, **30**, 117–124, (цитируется по англ. переводу: <http://bip.cnrs-mrs.fr/bip10/buchner0.htm>).
31. Jacob, F., and Monod, J. (1961) *J. Mol. Biol.*, **3**, 318–356.
32. Blake, C.C., Koenig, D.F., Mair, G.A., North, A.C., Phillips, D.C., and Sarma, V.R. (1965) *Nature*, **206**, 757–761.
33. Seidl, V., Huemer, B., Seiboth, B., and Kubicek, C.P. (2005) *FEBS J.*, **272**, 5923–5939.
34. Wolfenden, R., Lu, X., and Young, G. (1998) *J. Am. Chem. Soc.*, **120**, 6814–6815.
35. *Enzyme Nomenclature. Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the Nomenclature and Classification of Enzymes* (1992) Acad. Press, San Diego.
36. *ExplorEnz – The Enzyme Database* (2011), (<http://www.enzyme-database.org/index.php>).
37. Laine, R.A. (1994) *Glycobiology*, **4**, 759–767.
38. Watt, G.M., Lowden, P.A.S., and Flitsch, S.L. (1997) *Curr. Opin. Struct. Biol.*, **7**, 652–660.
39. Sinnott, M.L. (1998) *Biochem. Soc. Trans.*, **26**, 160–164.
40. Juers, D.H., Huber, R.E., and Matthews, B.W. (1999) *Protein Sci.*, **8**, 122–136.
41. Davies, G., and Henrissat, B. (1995) *Structure*, **3**, 853–859.
42. Davies, G.J. (1998) *Biochem. Soc. Trans.*, **26**, 167–173.
43. McCarter, J.D., and Withers, S.G. (1994) *Curr. Opin. Struct. Biol.*, **4**, 885–892.
44. Wang, Q., Graham, R.W., Trimbur, D., Warren, R.A.J., and Withers, S.G. (1994) *J. Am. Chem. Soc.*, **116**, 11594–11595.
45. Kasumi, T., Tsumuraya, Y., Brewer, C.F., Kersters-Hilderson, H., Claeysens, M., and Hehre, E.J. (1987) *Biochemistry*, **26**, 3010–3016.
46. Henrissat, B., and Davies, G. (1997) *Curr. Opin. Struct. Biol.*, **7**, 637–644.
47. Coutinho, P.M., and Henrissat, B. (1999) in *Recent advances in carbohydrate bioengineering* (Gilbert, H.J., Davies, G., Henrissat, B., and Svensson, B., ed.), The Royal Society of Chemistry, Cambridge, UK, pp. 3–12.
48. Nagano, N., Porter, C.T., and Thornton, J.M. (2001) *Protein Eng.*, **14**, 845–855.
49. Vasella, A., Davies, G.J., and Bohm, M. (2002) *Curr. Opin. Chem. Biol.*, **6**, 619–629.
50. White, A., and Rose, D.R. (1997) *Curr. Opin. Struct. Biol.*, **7**, 645–651.
51. Zechel, D.L., and Withers, S.G. (2000) *Acc. Chem. Res.*, **33**, 11–18.
52. Rye, C.S., and Withers, S.G. (2000) *Curr. Opin. Chem. Biol.*, **4**, 573–580.
53. Beguin, P. (1990) *Ann. Rev. Microbiol.*, **44**, 219–248.
54. Gilkes, N.R., Henrissat, B., Kilburn, D.G., Miller, Jr. R.C., and Warren, R.A.J. (1991) *Microbiol. Rev.*, **55**, 303–315.
55. Henrissat, B., Claeysens, M., Tomme, P., Lemesle, L., and Moron, J.-P. (1989) *Gene*, **81**, 83–95.
56. Henrissat, B. (1991) *Biochem. J.*, **280**, 309–316.
57. Henrissat, B., and Bairoch, A. (1993) *Biochem. J.*, **293**, 781–788.
58. Henrissat, B., and Bairoch, A. (1996) *Biochem. J.*, **316**, 695–696.
59. Himmel, M.E., Karplus, P.A., Sakon, J., Adney, W.S., Baker, J.O., and Thomas, S.R. (1997) *Appl. Biochem. Biotech.*, **63–65**, 315–325.
60. Henrissat, B. (1998) *Biochem. Soc. Trans.*, **26**, 153–156.
61. Henrissat, B., and Romeu, A. (1995) *Biochem. J.*, **311**, 350–351.
62. Marques, A.R., Coutinho, P.M., Videira, P., Fialho, A.M., and Sa-Correia, I. (2003) *Biochem. J.*, **370**, 793–804.
63. Gebler, J., Gilkes, N.R., Claeysens, M., Wilson, D.B., Beguin, P., Wakarchuk, W.W., Kilburn, D.G., Miller, Jr.R.C., Warren, R.A.J., and Withers, S.G. (1992) *J. Biol. Chem.*, **267**, 12559–12561.
64. Gloster, T.M., Turkenburg, J.P., Potts, J.R., Henrissat, B., and Davies, G.J. (2008) *Chem. Biol.*, **15**, 1058–1067.

65. Thompson, J., Ruvinov, S.B., Freedberg, D.I., and Hall, B.G. (1999) *J. Bacteriol.*, **181**, 7339–7345.
66. Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J.-P., and Davies, G. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 7090–7094.
67. Rawlings, N.D., and Barrett, A.J. (1993) *Biochem. J.*, **290**, 205–218.
68. Birsan, C., Johnson, P., Joshi, M., MacLeod, A., McIntosh, L., Monem, V., Nitz, M., Rose, D.R., Tull, D., Wakarchuck, W.W., Wang, Q., Warren, R.A.J., White, A., and Withers, S.G. (1998) *Biochem. Soc. Trans.*, **26**, 156–160.
69. Okuyama, M., Kitamura, M., Hondoh, H., Kang, M.S., Mori, H., Kimura, A., Tanaka, I., and Yao, M. (2009) *J. Mol. Biol.*, **392**, 1232–1241.
70. Kuroki, R., Weaver, L.H., and Matthews, B.W. (1993) *Science*, **262**, 2030–2033.
71. Kuroki, R., Weaver, L.H., and Matthews, B.W. (1995) *Nat. Struct. Biol.*, **2**, 1007–1011.
72. Stam, M., Bernard, T., Rancurel, C., Coutinho, P.M., and Henrissat, B. (2005) in *Journées Ouvertes Biologique Informatique Mathématiques (JOBIM'2005)*, July 6–8, 2005, Lyon, (<http://pbil.univ-lyon1.fr/events/jobim2005/proceedings/P133Stam.pdf>).
73. Stam, M. (2006) *Evolution et prediction des activites des glycoside hydrolases*. Ph.D. thesis, Universite Aix-Marseille 1, Marseille, France, ([www.sfb.fr/Theses/2006\\_Stam\\_Mark.pdf](http://www.sfb.fr/Theses/2006_Stam_Mark.pdf)).
74. Svensson, B. (1988) *FEBS Lett.*, **230**, 72–76.
75. Наумов Д. Г., Карперас М. (2009) *Молекуляр. биология*, **43**, 709–721.
76. Rigden, D.J. (2002) *FEBS Lett.*, **523**, 17–22.
77. Janecek, S., Svensson, B., and MacGregor, E.A. (2007) *FEBS Lett.*, **581**, 1261–1268.
78. Naumoff, D.G. (2000) in *Abstracts of Fourth International Fructan Symposium «Fructan 2000», August 16–20, 2000*, Arolla, Switzerland, P.1.4, (<http://www.kokkinias.com/fructan/admin/abstracts/pdf/naumoff.pdf>).
79. Naumoff, D.G. (2004) in *Proceedings of the Fourth International Conference on Bioinformatics of Genome Regulation and Structure, July 25–30, 2004*, vol. 1, Novosibirsk, Russia, pp. 315–318.
80. Naumoff, D.G. (2006) in *Proceedings of the Fifth International Conference on Bioinformatics of Genome Regulation and Structure, July 16–22, 2006*, vol. 1, Novosibirsk, Russia, pp. 294–298.
81. Finn, R.D., Mistry, J., Tate, J., Coghill, P., Heger, A., Pollington, J.E., Gavin, O.L., Gunasekaran, P., Ceric, G., Forslund, K., Holm, L., Sonnhammer, E.L., Eddy, S.R., and Bateman, A. (2010) *Nucl. Acids Res.*, **38**, D211–D222.
82. Andreeva, A., Howorth, D., Chandonia, J.M., Brenner, S.E., Hubbard, T.J., Chothia, C., and Murzin, A.G. (2008) *Nucleic Acids Res.*, **36**, D419–D425.
83. Roka, P., Czanik, P., Ponyi, T., and Fulop, L. (2008) *Bulletin of the Szent István University*, **1**, 5–10.
84. Banner, D.W., Bloomer, A.C., Petsko, G.A., Phillips, D.C., Pogson, C.I., Wilson, I.A., Corran, P.H., Furth, A.J., Milman, J.D., Offord, R.E., Priddle, J.D., and Waley, S.G. (1975) *Nature*, **255**, 609–614.
85. Hocker, B., Jurgens, C., Wilmanns, M., and Sterner, R. (2001) *Curr. Opin. Biotechnol.*, **12**, 376–381.
86. Henn-Sax, M., Hocker, B., Wilmanns, M., and Sterner, R. (2001) *Biol. Chem.*, **382**, 1315–1320.
87. Orengo, C.A., Jones, D.T., and Thornton, J.M. (1994) *Nature*, **372**, 631–634.
88. Rigden, D.J., Jedrzejewski, M.J., and de Mello, L.V. (2003) *FEBS Lett.*, **544**, 103–111.
89. Nagano, N., Orengo, C.A., and Thornton, J.M. (2002) *J. Mol. Biol.*, **321**, 741–765.
90. Farber, G.K., and Petsko, G.A. (1990) *Trends Biochem. Sci.*, **15**, 228–234.
91. Hocker, B., Schmidt, S., and Sterner, R. (2002) *FEBS Lett.*, **510**, 133–135.
92. Lang, D., Thoma, R., Henn-Sax, M., Sterner, R., and Wilmanns, M. (2000) *Science*, **289**, 1546–1550.
93. Richter, M., Bosnali, M., Carstensen, L., Seitz, T., Durchschlag, H., Blanquart, S., Merkl, R., and Sterner, R. (2010) *J. Mol. Biol.*, **398**, 763–773.
94. Jenkins, J., Leggio, L.L., Harris, G., and Pickersgill, R. (1995) *FEBS Lett.*, **362**, 281–285.
95. Pickersgill, R., Harris, G., Leggio, L.L., Mayans, O., and Jenkins, J. (1998) *Biochem. Soc. Trans.*, **26**, 190–198.
96. St John, F.J., Gonzalez, J.M., and Pozharski, E. (2010) *FEBS Lett.*, **584**, 4435–4441.
97. Rigden, D.J. (2005) *FEBS Lett.*, **579**, 5466–5472.
98. Lo Leggio, L., and Larsen, S. (2002) *FEBS Lett.*, **523**, 103–108.
99. Stam, M.R., Blanc, E., Coutinho, P.M., and Henrissat, B. (2005) *Carbohydrate Research*, **340**, 2728–2734.
100. Наумов Д.Г. (2008) В кн. *Материалы Пятого съезда общества биотехнологов России им. Ю.А. Овчинникова, 2–4 декабря 2008*, Москва, с. 116–117, (<http://bioros.tmweb.ru/Vcongress/Naumov.pdf>).
101. Наумов Д. Г., Степушенко О.О. (2011) *Молекуляр. биология*, **45**(3), (в печати).
102. Naumoff, D.G. (2010) *J. Bioinform. Comput. Biol.*, **8**, 437–451.
103. Dagnall, B.H., Paulsen, I.T., and Saier, Jr.M.H. (1995) *Biochem. J.*, **311**, 349–350.
104. Naumoff, D.G. (2001) *Glycoconjugate J.*, **18**, 109.
105. Naumoff, D.G. (2003) in *Program and Abstracts of the 5th Carbohydrate Bioengineering Meeting. 6–9 April, 2003*, Groningen, The Netherlands, pp. 81.
106. Naumoff, D.G. (2002) in *International Summer School «From Genome to Life: Structural, Functional and Evolutionary Approaches», July 15–27, 2002*, Cargèse, Corsica, France, pp. 40, (<http://www-archbac.u-psud.fr/Meetings/cargese2002/abstracts/NAUMOFF.html>).
107. Alcutt, M.J., Hsieh, H.-Y., Chapman, L.F., and Smith, D.S. (2002) *FEMS Microbiol. Lett.*, **214**, 77–80.
108. Gizatullina, D.I., and Naumoff, D.G. (2009) in *Proceedings of the International Moscow Conference on Computational Molecular Biology, July 20–23, 2009*, Moscow, Russia, pp. 249–250, ([http://mccmb.belozersky.msu.ru/2009/MCCMB09\\_Proceedings.pdf](http://mccmb.belozersky.msu.ru/2009/MCCMB09_Proceedings.pdf)).
109. Stam, M.R., Danchin, E.G., Rancurel, C., Coutinho, P.M., and Henrissat, B. (2006) *Protein Eng. Des. Sel.*, **19**, 555–562.
110. Naumoff, D.G., and Naumov, G.I. (2010) В кн. *Доклады III международной конференции «Математическая биология и биоинформатика», Пуццо, 10–15 октября 2010*, Изд-во МАКС Пресс, Москва, с. 135–136.
111. Gizatullina, D.I., and Naumoff, D.G. (2008) in *The Sixth International Conference on Bioinformatics of Genome Regulation and Structure, June 22–28, 2008*, Novosibirsk, Russia, pp. 82, ([http://www.bionet.nsc.ru/meeting/bgrs2008/BGRS2008\\_Proceedings.pdf](http://www.bionet.nsc.ru/meeting/bgrs2008/BGRS2008_Proceedings.pdf)).
112. Kitamura, M., Okuyama, M., Tanzawa, F., Mori, H., Kitago, Y., Watanabe, N., Kimura, A., Tanaka, I., and Yao, M. (2008) *J. Biol. Chem.*, **283**, 36328–36337.
113. Hidaka, M. (2010) *Trends Glycosci. Glycotechnol.*, **22**, 41–44.
114. Naumoff, D.G. (2005) *FEBS J.*, **272**, 94–95.
115. Naumoff, D.G. (2009) *Glycoconjugate J.*, **26**, 847.
116. Naumoff, D., and Stepuschenko, O. (2010) *FEBS J.*, **277**, 233–234.
117. Intra, J., Pavesi, G., and Horner, D.S. (2008) *BMC Evol. Biol.*, **8**, Art.214.

118. Shallom, D., Golan, G., Shoham, G., and Shoham, Y. (2004) *J. Bacteriol.*, **186**, 6928–6937.
119. Cournoyer, B., and Faure, D. (2003) *J. Mol. Microbiol. Biotechnol.*, **5**, 190–198.
120. Hidaka, M., Nishimoto, M., Kitaoka, M., Wakagi, T., Shoun, H., and Fushinobu, S. (2009) *J. Biol. Chem.*, **284**, 7273–7283.
121. Sumida, T., Fujimoto, K., and Ito, M. (2011) *J. Biol. Chem.* (in press). [PMID: 21297160]
122. Imamura, H., Fushinobu, S., Jeon, B.-S., Wakagi, T., and Matsuzawa, H. (2001) *Biochem.*, **40**, 12400–12406.
123. Janecek, S. (1998) *Folia Microbiol.*, **43**, 123–128.
124. Gonzalez, D.S., and Jordan, I.K. (2000) *Mol. Biol. Evol.*, **17**, 292–300.
125. Zona, R., Chang-Pi-Hin, F., O'Donohue, M.J., and Janecek, S. (2004) *Eur. J. Biochem.*, **271**, 2863–2872.
126. Mertz, B., Kuczenski, R.S., Larsen, R.T., Hill, A.D., and Reilly, P.J. (2005) *Biopolymers*, **79**, 197–206.
127. Chaudhuri, I., Soding, J., and Lupas, A.N. (2008) *Proteins*, **71**, 795–803.
128. Chaudhuri, I., Coles, M., Martin, J., and Lupas, A.N. (2005) *FEBS J.*, **272**, 79.
129. Наумов Д.Г., Дорошенко В.Г. (1998) *Молекуляр. биология*, **32**, 902–907.
130. Naumoff, D.G. (2001) *Proteins*, **42**, 66–76.
131. Наумов Д.Г. (2000) В кн. *Биосфера и человечество. Материалы конференции, посвященной 100-летию со дня рождения Н.В. Тимофеева-Ресовского, 20–21 сентября 2000*, Обнинск, с. 116–122.
132. Pons, T., Naumoff, D.G., Martinez-Fleites, C., and Hernandez, L. (2004) *Proteins*, **54**, 424–432.
133. Strohmeier, M., Hrmova, M., Fischer, M., Harvey, A.J., Fincher, G.B., and Pleiss, J. (2004) *Protein Sci.*, **13**, 3200–3213.
134. Fushinobu, S., Hidaka, M., Miyanaga, A., and Imamura, H. (2007) *J. Appl. Glycosci.*, **54**, 95–102.
135. *The family GH16 glycoside hydrolase database* (2011), (<http://www.ghdb.uni-stuttgart.de>).
136. Jordan, I.K., Bishop, G.R., and Gonzalez, D.S. (2001) *Bioinformatics*, **17**, 965–976.
137. Adachi, W., Sakihama, Y., Shimizu, S., Sunami, T., Fukazawa, T., Suzuki, M., Yatsunami, R., Nakamura, S., and Takenaka, A. (2004) *J. Mol. Biol.*, **343**, 785–795.
138. Urbanowicz, B.R., Bennett, A.B., Del Campillo, E., Catala, C., Hayashi, T., Henrissat, B., Hofte, H., McQueen-Mason, S.J., Patterson, S.E., Shoseyov, O., Teeri, T.T., and Rose, J.K. (2007) *Plant Physiol.*, **144**, 1693–1696.
139. Holm, L., and Sander, C. (1994) *FEBS Lett.*, **340**, 129–132.
140. Monzingo, A.F., Marcotte, E.M., Hart, P.J., and Robertus, J.D. (1996) *Nat. Struct. Biol.*, **3**, 133–140.
141. Wohlkonig, A., Huet, J., Looze, Y., and Wintjens, R. (2010) *PLoS One*, **5**, e15388.
142. Tremblay, H., Blanchard, J., and Brzezinski, R. (2000) *Can. J. Microbiol.*, **46**, 952–955.
143. Pei, J., and Grishin, N.V. (2005) *Protein Sci.*, **14**, 2574–2581.
144. Jenkins, J., Mayans, O., and Pickersgill, R. (1998) *J. Struct. Biol.*, **122**, 236–246.
145. Rigden, D.J., and Franco, O.L. (2002) *FEBS Lett.*, **530**, 225–232.
146. Markovic, O., and Janecek, S. (2001) *Protein Eng.*, **14**, 615–631.
147. Березина О.В., Лунина Н.А., Зверлов В.В., Наумов Д.Г., Либель В., Великодворская Г.А. (2003) *Молекуляр. биология*, **37**, 801–809.
148. Igarashi, K., Ishida, T., Hori, C., and Samejima, M. (2008) *Appl. Environ. Microbiol.*, **74**, 5628–5634.
149. Fujimoto, Z., Ichinose, H., Biely, P., and Kaneko, S. (2011) *Acta Crystallogr.*, **67**, 68–71.
150. Moller, P.L., Jorgensen, F., Hansen, O.C., Madsen, S.M., and Stougaard, P. (2001) *Appl. Environ. Microbiol.*, **67**, 2276–2283.
151. Henrissat, B., Teeri, T.T., and Warren, R.A.J. (1998) *FEBS Lett.*, **425**, 352–354.
152. Dive, C., Stahlberg, J., Reinikainen, T., Ruchonen, L., Pettersson, G., Knowles, J.K.C., Teeri, T.T., and Jones, T.A. (1994) *Science*, **265**, 524–528.
153. Torronen, A., Kubicek, C.P., and Henrissat, B. (1993) *FEBS Lett.*, **321**, 135–139.
154. Margolles-Clark, E., Tenkanen, M., Luonteri, E., and Penttila, M. (1996) *Eur. J. Biochem.*, **240**, 104–111.
155. Liebl, W., Wagner, B., and Schellhase, J. (1998) *System. Appl. Microbiol.*, **21**, 1–11.
156. Crennell, S.J., Garman, E.F., Laver, W.G., Vimr, E.R., and Taylor, G.L. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 9852–9856.
157. Ferretti, J.J., Gilpin, M.L., and Russell, R.R.B. (1987) *J. Bacteriol.*, **169**, 4271–4278.
158. Rojas, A., Garcia-Vallve, S., Palau, J., and Romeu, A. (1999) *Biologia*, **54**, 255–277.
159. MacGregor, E.A., Jespersen, H.M., and Svensson, B. (1996) *FEBS Lett.*, **378**, 263–266.
160. MacGregor, E.A., Janecek, S., and Svensson, B. (2001) *Biochim. Biophys. Acta*, **1546**, 1–20.
161. Mooser, G., Hefta, S.A., Paxton, R.J., Shively, J.E., and Lee, T.D. (1991) *J. Biol. Chem.*, **266**, 8916–8922.
162. Janecek, S. (2005) *Biologia, Bratislava*, **60** (Suppl. 16), 177–184.
163. Pons, T., Olmea, O., China, G., Beldarrain, A., Marquez, G., Acosta, N., Rodriguez, L., and Valencia, A. (1998) *Proteins*, **33**, 383–395.
164. Sprenger, N., Bortlik, K., Brandt, A., Boller, T., and Wiemken, A. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 11652–11656.



## HIERARCHICAL CLASSIFICATION OF GLYCOSIDE HYDROLASES

**D. G. Naumoff**

*State Institute for Genetics and Selection of Industrial  
Microorganisms, Pervy Dorozhny proezd 1, Moscow 117545,  
Russia; E-mail: daniil\_naumoff@yahoo.com*

Received January 21, 2011

Structural and functional features of glycoside hydrolases are considered in this review. Glycoside hydrolases are a widespread group of enzymes that are represented in almost all living organisms. Their catalytic domains are grouped into 120 amino acid sequence-based families in the international classification of carbohydrate-active enzymes (CAZy database). Some of these families compose 14 clans at a higher hierarchical level. Enzymes of the same clan have common evolutionary origin of their genes and share the most important functional characteristics: composition of the active center, anomeric configuration of the hydrolyzed glycosidic bonds, and molecular mechanism of the catalyzed reaction (either inverting or retaining). Currently, extensive data on the relationship between glycoside hydrolase families belonging to different clans and/or not included into any clans are available in the literature, as well as information on phylogenetic protein relationship within particular families. Based on the data, a multilevel hierarchical classification of glycoside hydrolases and their homologs is proposed. Almost all families of glycoside hydrolase catalytic domains can be assigned to six basic folds, which are large groups of proteins having essentially the same three-dimensional structure and common evolutionary origin.

*Key words:* glycoside hydrolase, CAZy, protein hierarchical classification, protein evolution, protein family, clan, superfamily, subfamily, search of homologs, TIM-barrel, SCOP, Pfam, PSI-BL