



Gos NII Genetika
Moscow, Russia



Hierarchical Classification of Glycoside Hydrolases

Daniil G. Naumoff

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

<http://bioinform.genetika.ru/members/Naumoff/index.htm>

daniil_naumoff@yahoo.com

Иерархическая классификация гликозил-гидролаз

Даниил Геннадиевич Наумов

Лаборатория биоинформатики
Государственный Научно-Исследовательский Институт Генетики и Селекции
Промышленных Микроорганизмов, Москва

Классификация Ферментов (IUBMB)

К.Ф. 1 – Оксидоредуктазы

К.Ф. 2 – Трансферазы

К.Ф. 2.4 – Гликозил-трансферазы (в т.ч. трансгликозидазы)

К.Ф. 2.4.1 – **Гексозил-трансферазы**

К.Ф. 2.4.1.10 – **Левансахараза (β -фруктозил-трансфераза)**

К.Ф. 2.4.1.67 – **Стахиозосинтаза (α -галактозил-трансфераза)**

К.Ф. 2.4.2 – Пентозил-трансферазы

К.Ф. 3 – Гидролазы

К.Ф. 3.2 – Гликозил-гидролазы

К.Ф. 3.2.1 – **О- и S-гликозил-гидролазы**

К.Ф. 3.2.1.22 – **Мелибиаза (α -галактозидаза)**

К.Ф. 3.2.1.26 – **Сахараза (инвертаза, β -фруктофуранозидаза)**

К.Ф. 3.2.2 – N-гликозил-гидролазы

К.Ф. 4 – Лиазы

К.Ф. 5 – Изомеразы

К.Ф. 6 – Лигазы

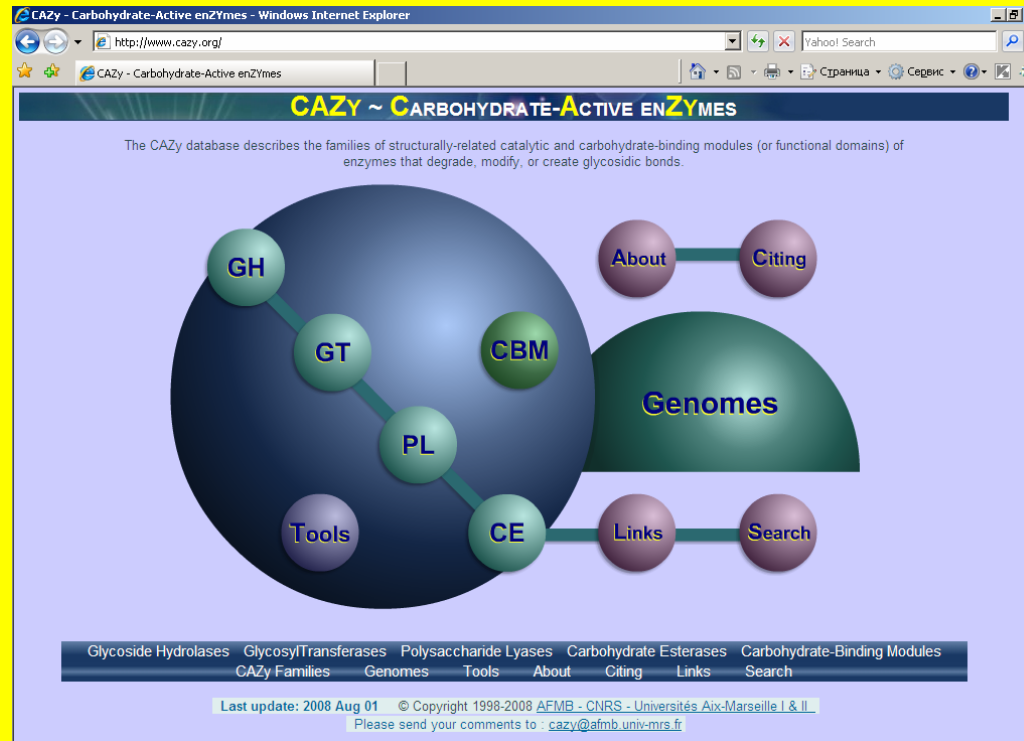
Гликозил-гидролазы = К.Ф. 3.2.1.X; X=1...165

Progress in the Sequence Based Classification of Glycoside Hydrolases and Their Homologues

Number of proteins	Number of families	Reference
300	36	Henrissat. <i>Biochem. J.</i> 1991, 280:309-316
480	45	Henrissat & Bairoch. <i>Biochem. J.</i> 1993, 293:781-788
950	57 (17 => 5 clans)	Henrissat & Bairoch. <i>Biochem. J.</i> 1996, 316:695-696
...	62 (19 => 5 clans)	Henrissat & Davies. <i>Curr. Opin. Struct. Biol.</i> 1997, 7:637-644
2.200	70 (...)	Coutinho & Henrissat. <i>CAZy server.</i> 1998
>20.000	101 (46 => 14 clans)	Coutinho & Henrissat. <i>CAZy server.</i> 2005
~30.000	106 (46 => 14 clans)	Coutinho & Henrissat. <i>CAZy server.</i> 2006
>30.000	110 (46 => 14 clans)	Coutinho & Henrissat. <i>CAZy server.</i> 2007
>40.000	114 (49 => 14 clans)	Coutinho & Henrissat. <i>CAZy server.</i> 2008

CAZy = the Carbohydrate-Active enZYmes server (<http://www.cazy.org/>)

Sequence Based Classification of the Carbohydrate-Active Enzymes at the CAZy server (www.cazy.org/)



- Glycoside Hydrolases (including transglycosidases) => 114 GH families (14 clans)
- Glycosyltransferases => 91 GT families (3 folds)
- Polysaccharide Lyases => 20 PL families
- Carbohydrate Esterases => 16 CE families
- Carbohydrate-Binding Modules => 52 CBM families

Объединение семейств гликозидаз в кланы

Клан	Семейства (GH)	Оптическая конфигурация	Трёхмерная структура
GH-A	1, 2, 5, 10, 17, 26, 30, 35, 39, 42, 50, 51, 53, 59, 72, 79, 86, 113	сохраняется (экв.)	$(\beta/\alpha)_8$ -barrel
GH-B	7, 16	сохраняется (экв.)	β -jelly roll
GH-C	11, 12	сохраняется (экв.)	β -jelly roll
GH-D	27, 31, 36	сохраняется (акс.)	$(\beta/\alpha)_8$ -barrel
GH-E	33, 34, 83	сохраняется (экв.)	6-fold β -propeller
GH-F	43, 62	меняется (экв.)	5-fold β -propeller
GH-G	37, 63	меняется (акс.)	$(\alpha/\alpha)_6$
GH-H	13, 70, 77	сохраняется (акс.)	$(\beta/\alpha)_8$ -barrel
GH-I	24, 46, 80	меняется (экв.)	$\alpha+\beta$
GH-J	32, 68	сохраняется (β -фуранозид)	5-fold β -propeller
GH-K	18, 20, 85	сохраняется (экв.)	$(\beta/\alpha)_8$ -barrel
GH-L	15, 65	меняется (акс.)	$(\alpha/\alpha)_6$
GH-M	8, 48	меняется (экв.)	$(\alpha/\alpha)_6$
GH-N	28, 49	меняется (акс.)	$(\beta)_3$ -solenoid

Пространственная структура белков семейства GH27 гликозидаз (α -галактозидаза *Trichoderma reesei*)



GH27N

N-концевой
каталитический домен в
виде $(\beta/\alpha)_8$ -бочонка
(TIM barrel-type structure)

GH27C

C-концевой домен,
состоящий из восьми
антипараллельных
 β -слоёв (β -sandwich
fold, Greek key motif)

Объединение семейств гликозидаз в кланы

Клан	Семейства (GH)	Оптическая конфигурация	Трёхмерная структура
GH-A	1, 2, 5, 10, 17, 26, 30, 35, 39, 42, 50, 51, 53, 59, 72, 79, 86, 113	сохраняется (экв.)	$(\beta/\alpha)_8$ -barrel
GH-B	7, 16	сохраняется (экв.)	β -jelly roll
GH-C	11, 12	сохраняется (экв.)	β -jelly roll
GH-D	27, 31, 36	сохраняется (акс.)	$(\beta/\alpha)_8$ -barrel
GH-E	33, 34, 83	сохраняется (экв.)	6-fold β -propeller
GH-F	43, 62	меняется (экв.)	5-fold β -propeller
GH-G	37, 63	меняется (акс.)	$(\alpha/\alpha)_6$
GH-H	13, 70, 77	сохраняется (акс.)	$(\beta/\alpha)_8$ -barrel
GH-I	24, 46, 80	меняется (экв.)	$\alpha+\beta$
GH-J	32, 68	сохраняется (β -фуранозид)	5-fold β -propeller
GH-K	18, 20, 85	сохраняется (экв.)	$(\beta/\alpha)_8$ -barrel
GH-L	15, 65	меняется (акс.)	$(\alpha/\alpha)_6$
GH-M	8, 48	меняется (экв.)	$(\alpha/\alpha)_6$
GH-N	28, 49	меняется (акс.)	$(\beta)_3$ -solenoid

K.Φ. 3.2.1.22

Common name: α -galactosidase

Reaction: Hydrolysis of terminal, non-reducing α -D-galactose residues in α -D-galactosides, including galactose oligosaccharides, galactomannans, and galactolipids

Molecular mechanism of hydrolyzing reaction: double displacement with overall retention of the anomeric configuration of the axial glycosidic bond

Other names: melibiase; α -D-galactosidase; α -galactosidase A; α -D-galactoside galactohydrolase; α -galactoside galactohydrolase

Systematic name: α -D-galactoside galactohydrolase

Comments: Also hydrolyses α -D-fucosides

Семейства гликозидаз, содержащие α -галактозидазы

- GH4** – α -galactosidase from *Escherichia coli* and bifunctional enzymes with α -galactosidase and α -glucosidase (EC 3.2.1.20) activities from *Thermotoga maritime* and *Thermotoga neapolitana*
+ 6-phospho- α -glucosidases (EC 3.2.1.122) from Eubacteria
+ 6-phospho- β -glucosidases (EC 3.2.1.86) from Eubacteria
- GH27** – mainly α -galactosidases from Eukaryota (animals, plants, fungi, etc.) and some Eubacteria
+ α -N-acetylgalactosaminidases (EC 3.2.1.49) from Eukaryota (animals and fungi)
+ isomalto-dextranase (EC 3.2.1.94) from *Arthrobacter globiformis*
- GH36** – mainly α -galactosidases from Eubacteria and Eukaryota (fungi and plants)
+ α -galactosyltransferases (EC 2.4.1.67 and EC 2.4.1.82) from plants
+ α -N-acetylgalactosaminidase from *Clostridium perfringens*
+ ORFs from *Sulfolobus solfataricus* and *Sulfolobus tokodaii*
- GH57** – α -galactosidases from *Pyrococcus furiosus* and *Thermococcus alcaliphilus*
+ α -amylases (EC 3.2.1.1), amylopullulanase (EC 3.2.1.41), and
4- α -glucanotransferases (EC 2.4.1.x) from Eubacteria and Archaea
- GH97** – α -galactosidase from *Bacteroides thetaiotaomicron*
+ α -glucosidases (EC 3.2.1.20) from *B. thetaiotaomicron* and *Tannerella forsythia*
- GH110** – α -galactosidases from *Bacteroides fragilis*, *B. thetaiotaomicron* and *Streptomyces avermitilis*

clan
GH-D

Family GH27

Subfamily 27a:

- plant α -galactosidases: *Arabidopsis*, *Carica*, *Coffea*, *Cyamopsis*, *Glycine*, *Helianthus*, *Hordeum*, *Lycopersicon*, *Oryza*, *Petunia*, *Phaseolus*, and *Senna*
- α -galactosyltransferase from *Ajuga reptans*
- animal α -galactosidases and α -N-acetylgalactosaminidases: *Anopheles*, *Apis*, *Ateles*, *Brugia*, *Caenorhabditis*, *Ciona*, *Danio*, *Drosophila*, *Gallus*, *Homo*, *Mus*, *Rattus*, *Takifugu*, and *Tetraodon*
- fungal and yeast α -galactosidases: *Aspergillus*, *Ganoderma*, *Gibberella*, *Magnaporthe*, *Mortierella*, *Penicillium*, *Phanerochaete*, *Saccharomyces*, *Schizosaccharomyces*, *Thermomyces*, *Torulaspora*, *Trichoderma*, *Ustilago*, and *Zygosaccharomyces*
- α -N-acetylgalactosaminidase from *Acremonium* sp.
- ORFs from *Dictyostelium discoideum* and *Toxoplasma gondii*
- bacterial α -galactosidases: *Cellvibrio mixtus*, *Clostridium josui*, *Pseudomonas fluorescens*, *Saccharopolyspora erythraea*, and *Streptomyces coelicolor*
- bacterial ORFs: *Bacteroides*, *Fibrobacter*, *Microbulbifer*, *Porphyromonas*, and *Streptomyces*

Subfamily 27b:

- plant ORFs: *Arabidopsis thaliana* and *Oryza sativa*
- bacterial ORFs: *Bacillus halodurans*, *Bifidobacterium longum*, *Kineococcus radiotolerans*, and *Ruminococcus albus*

Subfamily 27c:

- α -galactosidase from *Trichoderma reesei*
- ORF from *Aspergillus nidulans*

Unclassified proteins

- isomalto-dextranase from *Arthrobacter globiformis*
- bacterial ORFs: *Bacteroides thetaiotaomicron* and *Streptomyces avermitilis*

Family GH36

Subfamily GH36A:

- bacterial α -galactosidases: *Azotobacter*, *Bacillus*, *Geobacillus*, *Bifidobacterium*, *Carnobacterium*, *Clostridium*, *Enterococcus*, *Erwinia*, *Escherichia*, *Geobacillus*, *Klebsiella*, *Lactobacillus*, *Lactococcus*, *Leuconostoc*, *Novosphingobium*, *Oenococcus*, *Pediococcus*, *Ruminococcus*, *Streptococcus*, *Streptomyces*, *Thermoanaerobacter*, *Treponema*, *Vibrio*, and *Yersinia*
- fungal α -galactosidases: *Absidia*, *Aspergillus*, *Gibberella*, *Penicillium*, and *Trichoderma*

Subfamily GH36B:

- bacterial α -galactosidases: *Burkholderia*, *Leptospira*, *Streptomyces*, *Thermotoga*, *Thermus*, and *Vibrio*

Subfamily GH36C:

- plant alkaline α -galactosidases or seed imbibition proteins: *Arabidopsis*, *Brassica*, *Cicer*, *Cucumis*, *Hordeum*, *Lycopersicon*, *Malus*, *Oryza*, and *Persea*
- plant α -galactosyltransferases: *Alonsoa*, *Arabidopsis*, *Cucumis*, *Medicago*, *Oryza*, *Pisum*, *Stachys*, and *Vigna*
- fungal ORFs: *Aspergillus*, *Coccidioides*, *Cryptococcus*, *Gibberella*, *Magnaporthe*, *Neurospora*, and *Ustilago*
- ORF from *Toxoplasma gondii*
- α -galactosidase from *Bifidobacterium breve*
- bacterial ORFs: *Bifidobacterium longum*, *Bacteroides fragilis*, and *Bacteroides thetaiotaomicron*
- ORFs from *Sulfolobus solfataricus* and *Sulfolobus tokodaii*

Subfamily GH36D:

- α -N-acetylgalactosaminidase from *Clostridium perfringens*
- bacterial ORFs: *Aeromonas*, *Clostridium*, *Escherichia*, *Photobacterium*, and *Streptococcus*

Четыре консервативных участка в белках α -галактозидазного суперсемейства

MEL1_YEAST	65	GYKYIILDDCWS	141	NRVDYLKYDNCY	204	WRMSGDV	258	WNLDLNL
AGAL_PHAVU	106	GYQYINIDDCWG	184	WGIDYLKYDNCE	242	WRTTGDI	274	WNDPML
NAGA_HUMAN	71	GYTYLNIDDCWI	148	WKVDMLKLDGCF	212	WRNYDDI	248	WNDPML
AGL1_BIFLO	48	GWDTLVIDIDWY	178	WGLDFLKVDMMQ	240	WRISDDL	272	WADADMV
AGL3_HYPJE	234	GYDLCSLDSGWQ	339	WGVDMMLKLDFLT	407	MRTDQDL	451	YPDMDAL
IMD_ARTGO	100	GYDIACD-GWI	219	LGVPYLRIDFLS	290	VRINADA	338	ILDGDFM
AGAL_LACPL	367	GIEMFVLDDGWF	478	VPIDYIKWDMNR	551	QSWPSDN	584	GTSPDEL
AGL2_HYPJE	375	GIKLFVLDDGWF	496	ASISYVKWDMNR	561	HIWTSDD	594	SAVPNGQ
AGL7_ASPFU	413	GAGYFVIDAGWY	524	YGVGYFKFDYNI	595	LQSSSDQ	626	WAYPQPA
AGAL_THEMA	213	PFEVFQIDDAYE	319	MGYRYFKIDFLF	382	MRIGPDT	423	LNDPCL
AGAL_VIBPA	227	DLEWVLLDDGYQ	341	WGVELFKLDANY	404	MRVSDDV	436	QIDPCA
AGL3_STRCO	329	GLKWAVLDDGWQ	438	WGYEGLKIDGQH	512	QYPSSDP	540	SYSGDHV
AGAL_SULSO	256	RLNWFVIIDGWQ	360	RDFDLVKVDNQW	421	MRNSIDY	454	YPDYDMF
AGAL_BIFBR	339	PVSWVLIDGWS	458	AGVDFVKVDSQS	523	TRTSDDF	556	HCDWDMF
GALT_VIGAN	251	APRFVVIDGWQ	476	TGVTGVKIDVIH	544	GRVGDDF	585	QPDWDMF
NAGA_CLOPE	253	TLDAFVVDGWA	355	YDISYWKIDGML	429	IQTSQDV	...	
ORF1_CLOPE	515	PIDSYVVDGWH	634	FDIDYWKLDGFA	715	IQNSQDT	...	
ORF1_ECOLI	297	ALDAFLLDDGWD	393	EHITSFKLDGMG	458	WRQGDDI	...	
ORF2_CLOPE	112	PKGIIMIDGWS	220	YGVDFGFKFDAGD	286	HSWEYNG	344	ALMPMMQ
XYLQ_LACPE	297	PLDVHFHDCFQ	406	MGVDSFKTDFGE	474	IQYTGAA	534	LLSSHRS
XYLS_SULSO	250	PLDVIVLDWRYW	345	LGIDAYWLDASE	423	ISWSGDV	483	TFCPILR
LYAG_HUMAN	397	PLDVQWNLDYD	510	VPFDMGWI DMNE	611	GHWTGDV	666	AFYPFMR

GH27

GH36A

GH36B

GH36C

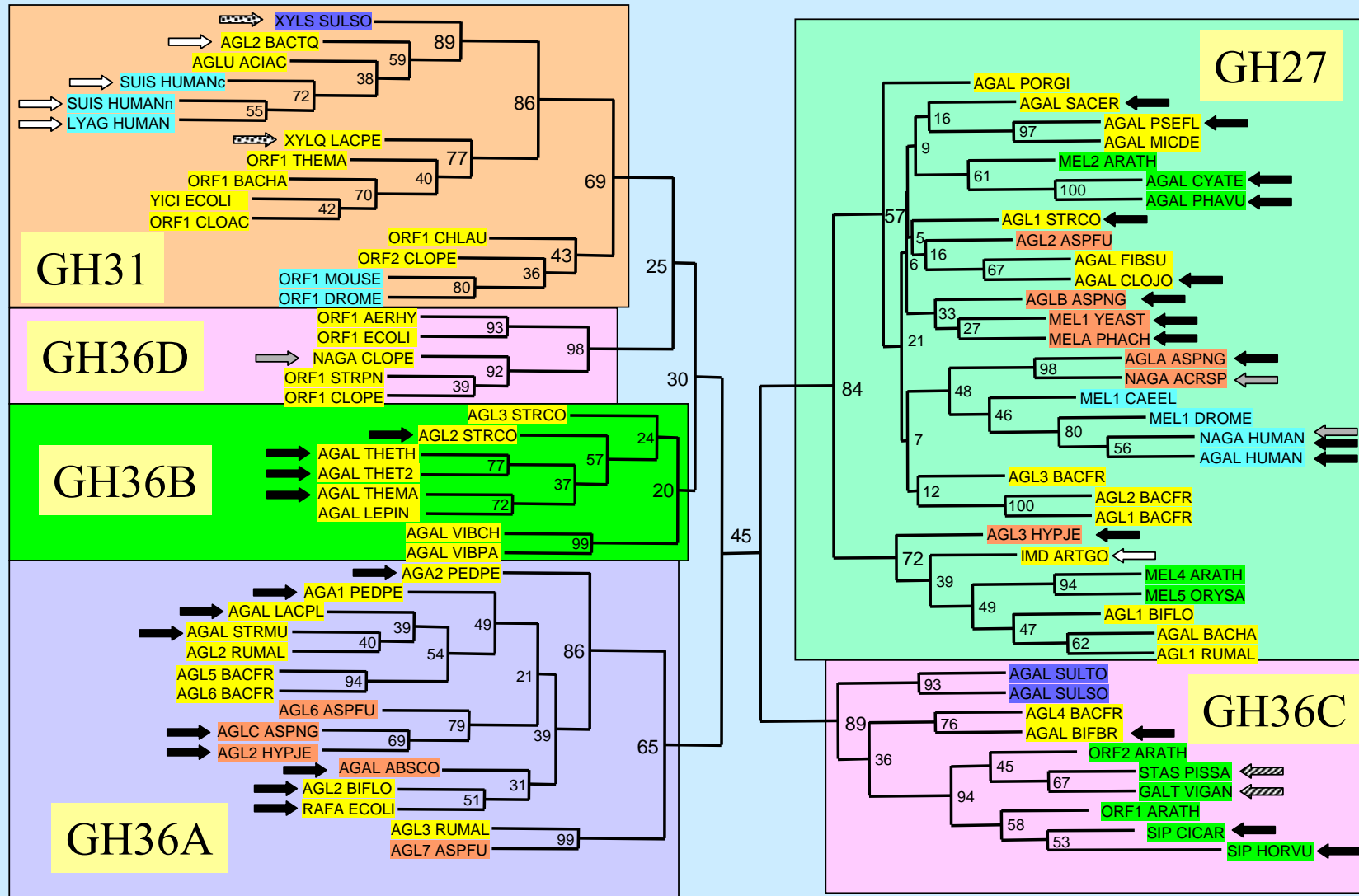
GH36D

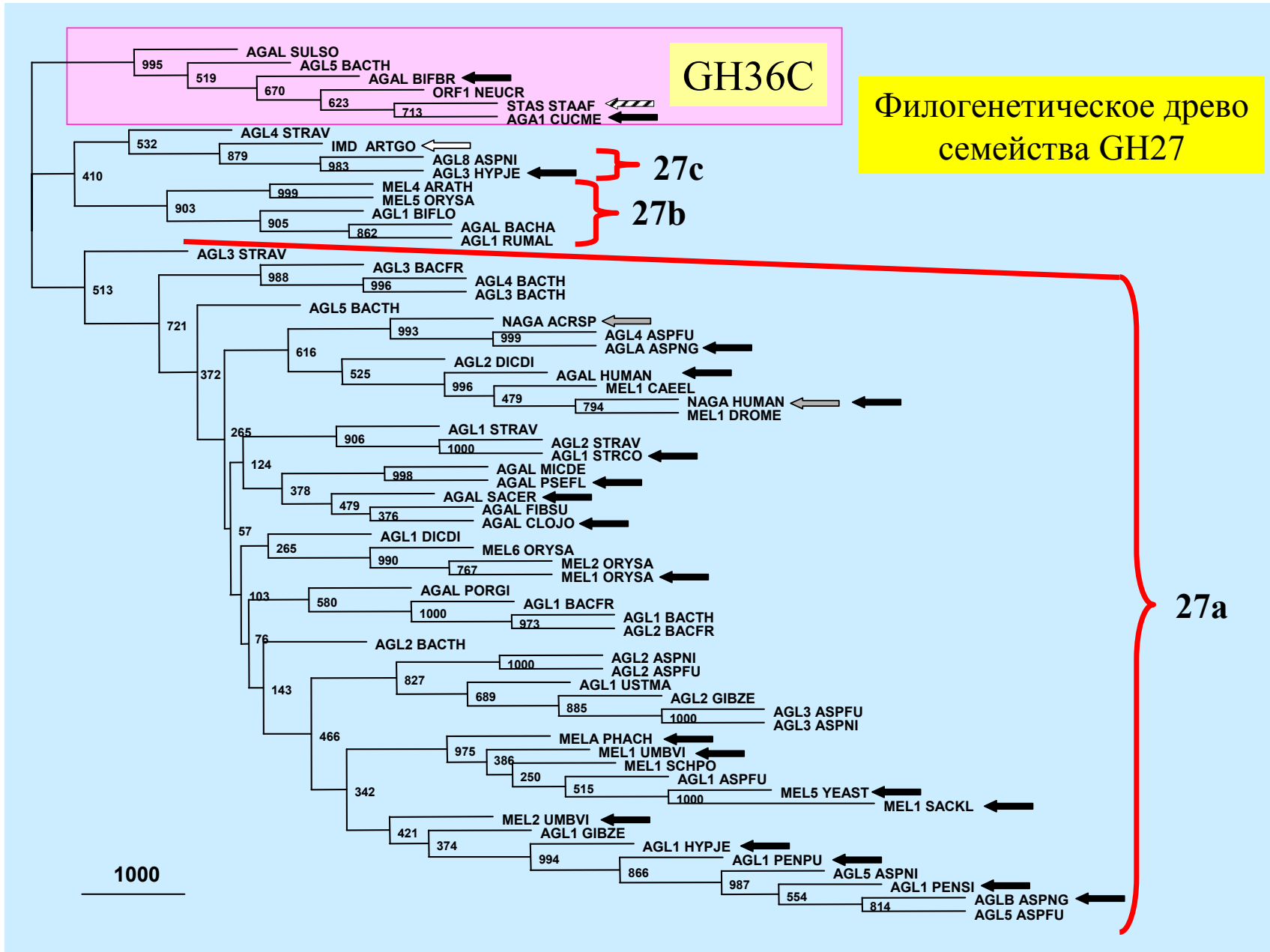
GH31

Нуклеофил ↑

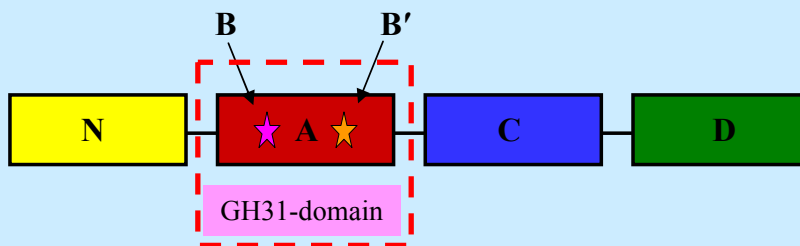
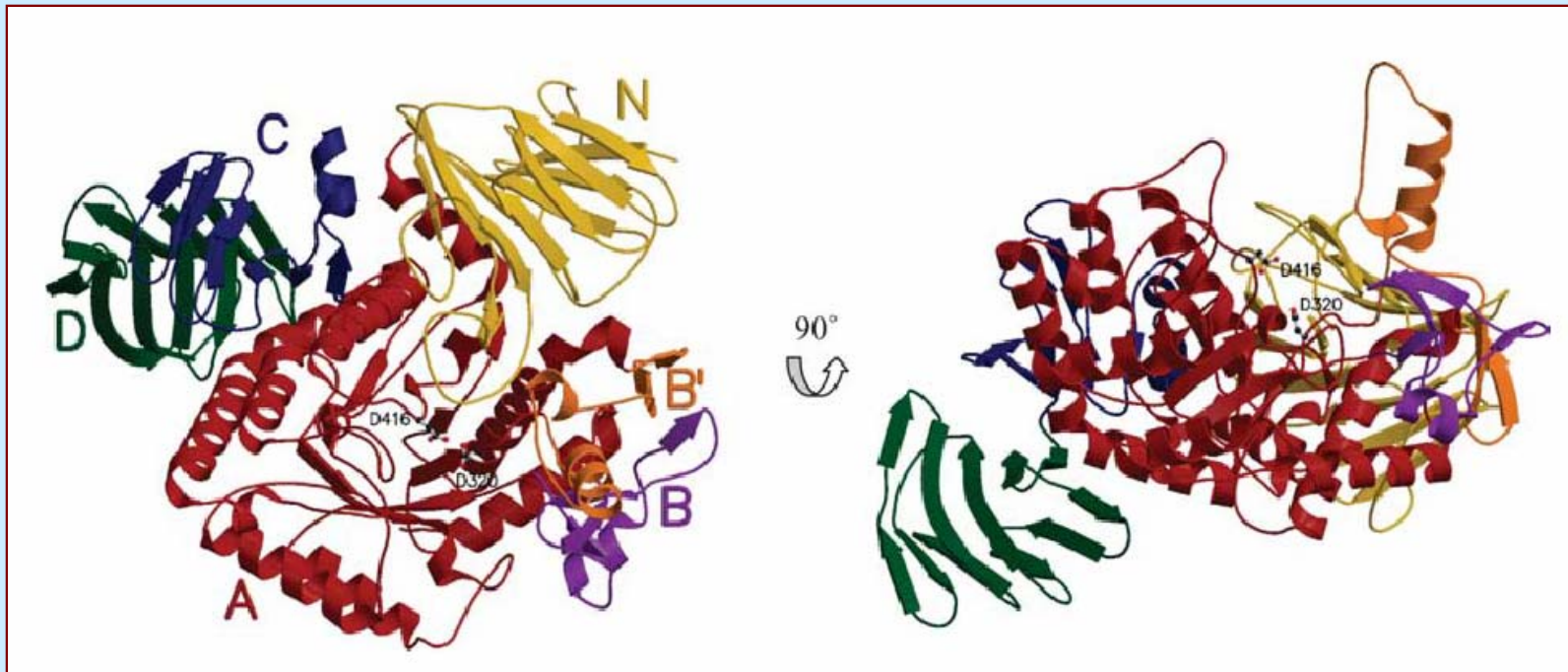
↑ Донор H⁺

The neighbor-joining phylogenetic tree of the α -galactosidase superfamily



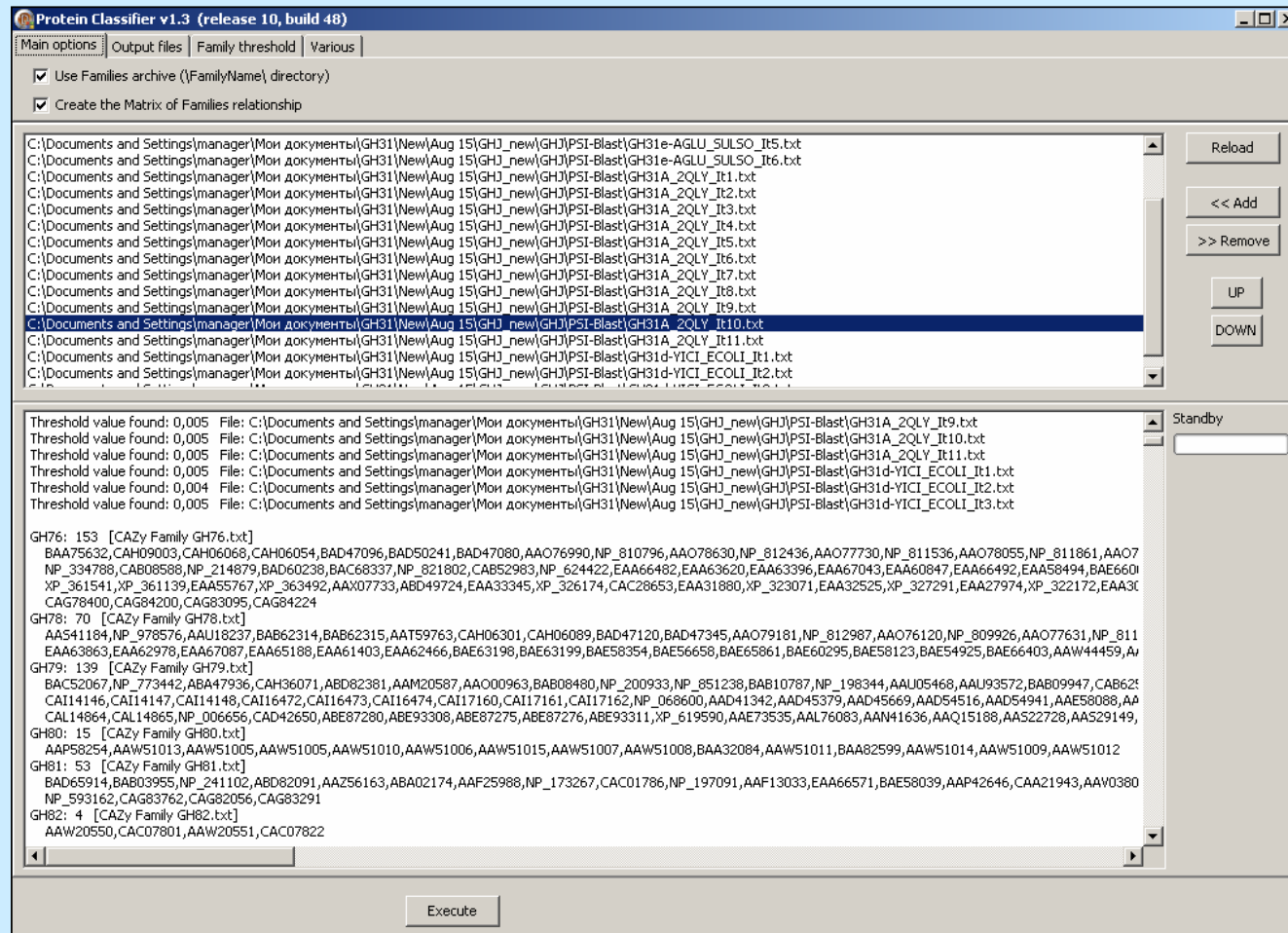


Domain structure of α -glucosidase MalA from *Sulfolobus solfataricus* (PDB, 2G3M)



MalA consists of four main domains (N, A, C, and D) and two sub-domains (B and B'). Domain A has the $(\beta/\alpha)_8$ -barrel type structure and belongs to the GH31 family.

Screenshot of PSI Protein Classifier



PSI Protein Classifier program was used for automatic analysis of PSI-BLAST outputs.

BLAST results (Descriptions)

Sequences with E-value BETTER than threshold

Sequences producing significant alignments:

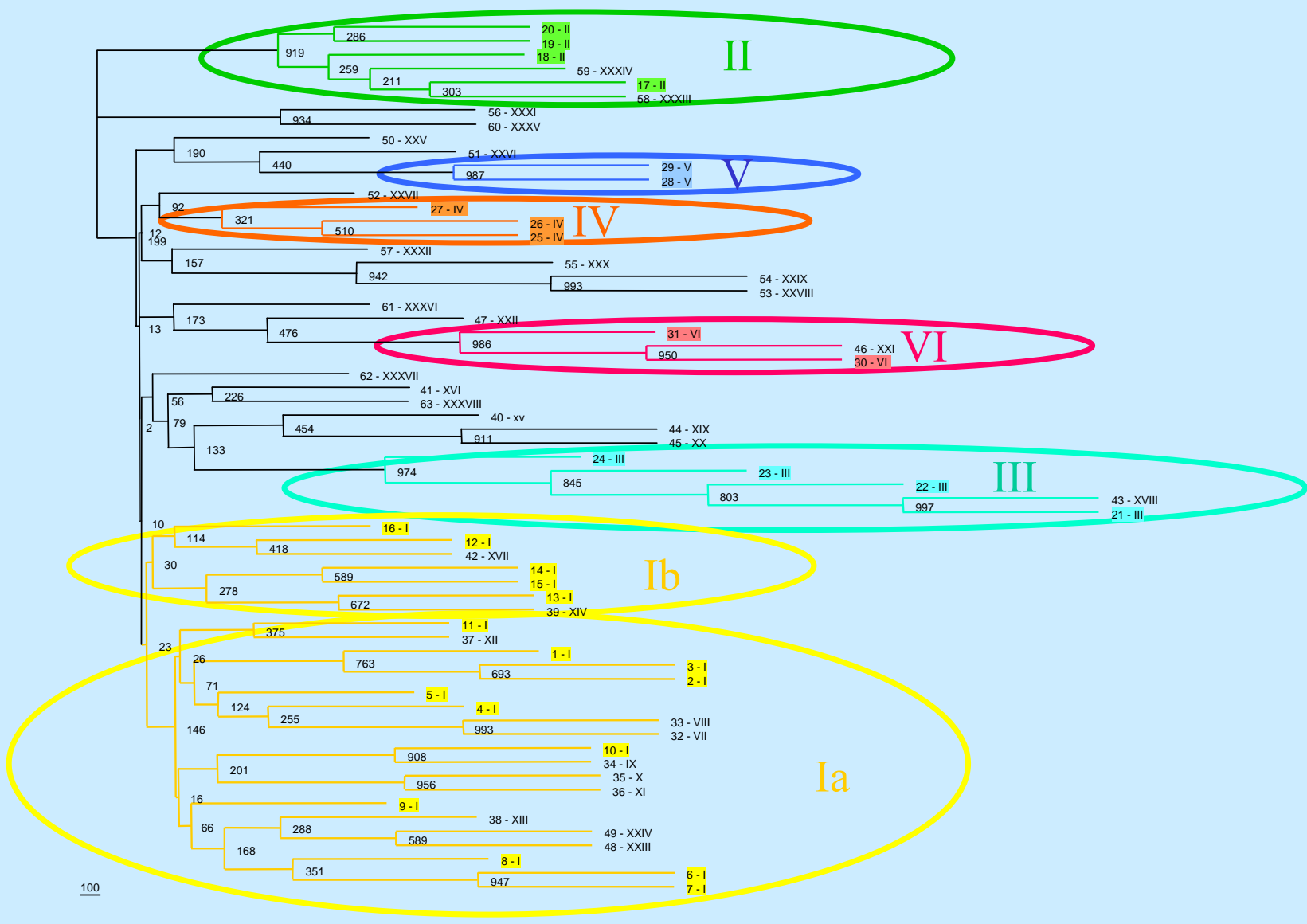
			Score (Bits)	E Value
NEW	<input checked="" type="checkbox"/>	gi 39965881 ref XP_365144.1 	1752	0.0
NEW	<input checked="" type="checkbox"/>	gi 88176755 gb EAQ84223.1 	489	2e-136
NEW	<input checked="" type="checkbox"/>	gi 88181573 gb EAQ89041.1 	473	2e-131
NEW	<input checked="" type="checkbox"/>	gi 46110355 ref XP_382235.1 	461	2e-125
NEW	<input checked="" type="checkbox"/>	gi 88181171 gb EAQ88639.1 	320	3e-85
NEW	<input checked="" type="checkbox"/>	gi 88181510 gb EAQ88978.1 		
NEW	<input checked="" type="checkbox"/>	gi 88181988 gb EAQ89456.1 		
NEW	<input checked="" type="checkbox"/>	gi 85090316 ref XP_958358.1 		
NEW	<input checked="" type="checkbox"/>	gi 89296464 gb EAR94452.1 	275	7e-72
NEW	<input checked="" type="checkbox"/>	gi 3777481 dbj BAA33931.1 	274	2e-71
NEW	<input checked="" type="checkbox"/>	gi 46139043 ref XP_391212.1 	273	2e-71
NEW	<input checked="" type="checkbox"/>	gi 22331822 ref NP_191190.2 	273	3e-71
NEW	<input checked="" type="checkbox"/>	gi 7572929 emb CAB87430.1 	272	5e-71
NEW	<input checked="" type="checkbox"/>	gi 17902582 emb CAC85738.1 	270	2e-70
NEW	<input checked="" type="checkbox"/>	gi 39951521 ref XP_363477.1 	270	3e-70
NEW	<input checked="" type="checkbox"/>	gi 39966419 ref XP_365195.1 	267	2e-69
NEW	<input checked="" type="checkbox"/>	gi 51701639 sp Q99172 MEL_ZYGCI	266	3e-69
NEW	<input checked="" type="checkbox"/>	gi 88176179 gb EAQ83647.1 	266	4e-69
NEW	<input checked="" type="checkbox"/>	gi 3821271 emb CAA08915.1 	264	1e-68
NEW	<input checked="" type="checkbox"/>	gi 29827169 ref NP_821803.1 	263	4e-68
NEW	<input checked="" type="checkbox"/>	gi 15788341 gb AAL07760.1 	263	4e-68
NEW	<input checked="" type="checkbox"/>	gi 89303594 gb EAS01582.1 	262	5e-68
NEW	<input checked="" type="checkbox"/>	gi 51701634 sp Q03647 MEL_SACPS	261	1e-67
NEW	<input checked="" type="checkbox"/>	gi 51701642 sp Q9P4V4 MEL_ZYGMR	260	2e-67
NEW	<input checked="" type="checkbox"/>	gi 30725668 gb AAP37856.1 	260	2e-67
NEW	<input checked="" type="checkbox"/>	gi 51963650 ref XP_506569.1 	260	2e-67
NEW	<input checked="" type="checkbox"/>	gi 70990042 ref XP_749870.1 	260	3e-67
NEW	<input checked="" type="checkbox"/>	gi 18292 emb CAA32772.1 	259	3e-67
NEW	<input checked="" type="checkbox"/>	gi 52699549 gb AAU86897.1 	259	4e-67
NEW	<input checked="" type="checkbox"/>	gi 53747927 emb CAF34023.1 	259	5e-67
NEW	<input checked="" type="checkbox"/>	gi 15241601 ref NP_196455.1 	259	5e-67
NEW	<input checked="" type="checkbox"/>	gi 92872552 gb ABE81161.1 	259	6e-67
NEW	<input checked="" type="checkbox"/>	gi 1246856 emb CAA64760.1 	256	5e-66

E-value < 0.01 or 0.001

Разбиение семейства GH31 на 38 подсемейств на основе попарного сравнения последовательностей с помощью программы PSI Protein Classifier

Подсемейства	Группы белков	Активности	Число найденных белков	
			31 января – 1 февраля 2008	17-18 сентября 2008
I	1-16	α -глюкозидаза [К.Ф. 3.2.1.20], α -ксилозидаза [К.Ф. 3.2.1.-], α -1,3-глюкозидаза [К.Ф. 3.2.1.84], сахараза-изомальтаза [К.Ф. 3.2.1.48] и [К.Ф. 3.2.1.10]	596	701
II	17-20		52	76
III	21-24		41	55
IV	25-27	α -ксилозидаза [К.Ф. 3.2.1.-]	92	118
V	28, 29	α -глюкозидаза [К.Ф. 3.2.1.20]	43	60
VI	30, 31		22	32
VII	32	α -1,4-глюканлиаза [К.Ф. 4.2.2.13]	5	5
VIII	33	α -1,4-глюканлиаза [К.Ф. 4.2.2.13]	7	8
IX	34		3	3
X	35		1	1
XI	36		1	1
XII	37		75	84
XIII	38		2	2
XIV	39	6- α -глюкозилтрансфераза [К.Ф. 2.4.1.-]	3	3
XV	40	6- α -глюкозилтрансфераза [К.Ф. 2.4.1.-]	14	10
XVI	41		1	1
XVII	42		2	2
XVIII	43		1	1
XIX	44		1	1
XX	45		1	1
XXI	46		6	6
XXII	47		4	4
XXIII	48		1	1
XXIV	49		1	1
XXV	50		1	1
XXVI	51		2	2
XXVII	52		3	3
XXVIII	53		1	1
XXIX	54		1	1
XXX	55		3	6
XXXI	56		1	1
XXXII	57	3- α -изомальтозилтрансфераза [К.Ф. 2.4.1.-]	21	15
XXXIII	58		12	29
XXXIV	59		5	3
XXXV	60		0	1
XXXVI	61		0	1
XXXVII	62		0	1
XXXVIII	63		0	1
прочие			74	80
всего			1099	1323

Филогенетическое древо семейства GN31



Семейства α -галактозидазного суперсемейства и семейство GH97

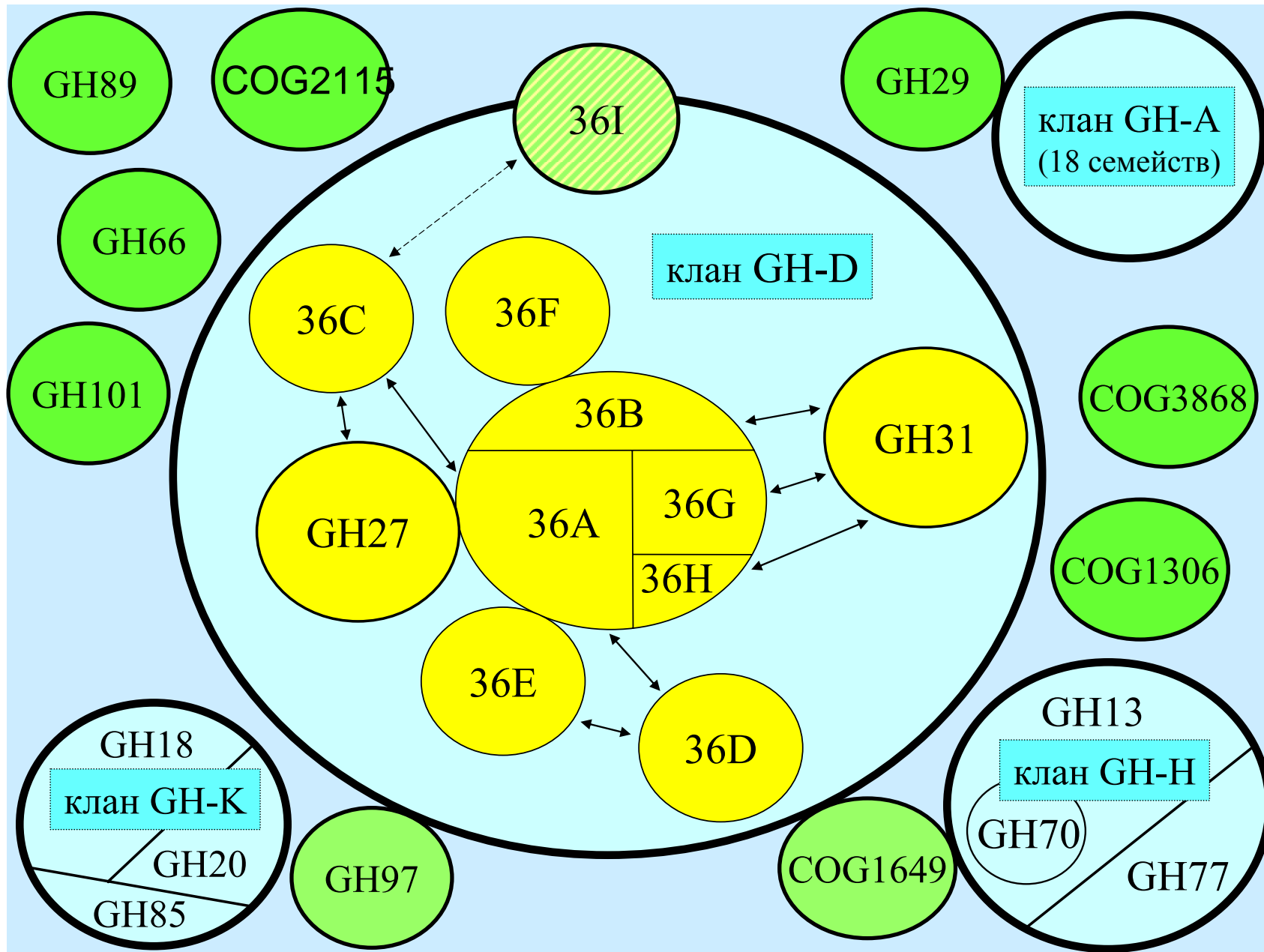
Family	GH27	GH31	GH36A	GH36B	GH36C	GH36D	GH97
Clan	GH-D	GH-D	GH-D	GH-D	GH-D	GH-D	None
COG/KOG	KOG2366	COG1501 KOG1065	COG3345	COG3345	None	None	None
Known enzymatic activities	EC 2.4.1.x <u>EC 3.2.1.22</u> <u>EC 3.2.1.49</u> EC 3.2.1.94	EC 2.4.1.x EC 3.2.1.10 <u>EC 3.2.1.20</u> EC 3.2.1.48 EC 3.2.1.84 EC 4.2.2.13	<u>EC 3.2.1.22</u>	<u>EC 3.2.1.22</u>	EC 2.4.1.67 EC 2.4.1.82 <u>EC 3.2.1.22</u>	<u>EC 3.2.1.49</u>	<u>EC 3.2.1.20</u> <u>EC 3.2.1.22</u>
Molecular mechanism	Retaining	Retaining	Retaining	Not known	Not known	Not known	Retaining, Inverting
Origin	Eukaryota: Alveolata Fungi Metazoa Mycetozoa Viridiplantae Eubacteria: Acidobacteria Actinobacteria Bacteroidetes Fibrobacteres Firmicutes Proteobacteria Verrucomicrobia	Eukaryota: Alveolata Entamoebidae Euglenozoa Fungi Metazoa Mycetozoa Rhodophyta Viridiplantae Eubacteria: Acidobacteria Actinobacteria Bacteroidetes Cyanobacteria Firmicutes Proteobacteria Spirochaetes Thermotogales Thermus Verrucomicrobia Archaea: Crenarchaeota Euryarchaeota	Eukaryota: Fungi Eubacteria: Acidobacteria Actinobacteria Bacteroidetes Firmicutes Proteobacteria Spirochaetes Verrucomicrobia	Eubacteria: Acidobacteria Actinobacteria Deinococcus Proteobacteria Spirochaetes Thermotogales Thermus	Eukaryota: Alveolata Fungi Viridiplantae Eubacteria: Actinobacteria Bacteroidetes Verrucomicrobia Archaea: Crenarchaeota	Eubacteria: Firmicutes Proteobacteria Verrucomicrobia	Eukaryota: Metazoa (?) Eubacteria: Acidobacteria Actinobacteria Bacteroidetes Planctomycetes Proteobacteria Verrucomicrobia Archaea: Euryarchaeota

Применение программы PSI Protein Classifier для поиска дальних гомологов семейства GH31

Семейство	Минимальное число итераций с данным белком			Число найденных представителей
	MgaM (подсемейство I)	MalA (подсемейство I)	YicI (подсемейство IV)	
GH5	–	– (10)*	17	1
GH13	2	4	2	3061
GH27	3	3	2	415
GH31	1	1	1	1323
GH36A	2	2	2	295
GH36B	2	2	1	58
GH36C	4	4	2	130
GH36D	2	5	2	31
GH36E	6	10	24	2
GH36F	9	6	4	10
GH36G	2	2	2	4
GH36H	3	3	2	7
GH66	10	–	–	1
GH101	–	20	6	20
COG1306	–	6	–	1
COG1649	4	6	5	24
COG3868	–	– (6)*	–	0
XP_001744488	–	–	10	1
XP_001813879	–	–	15	1

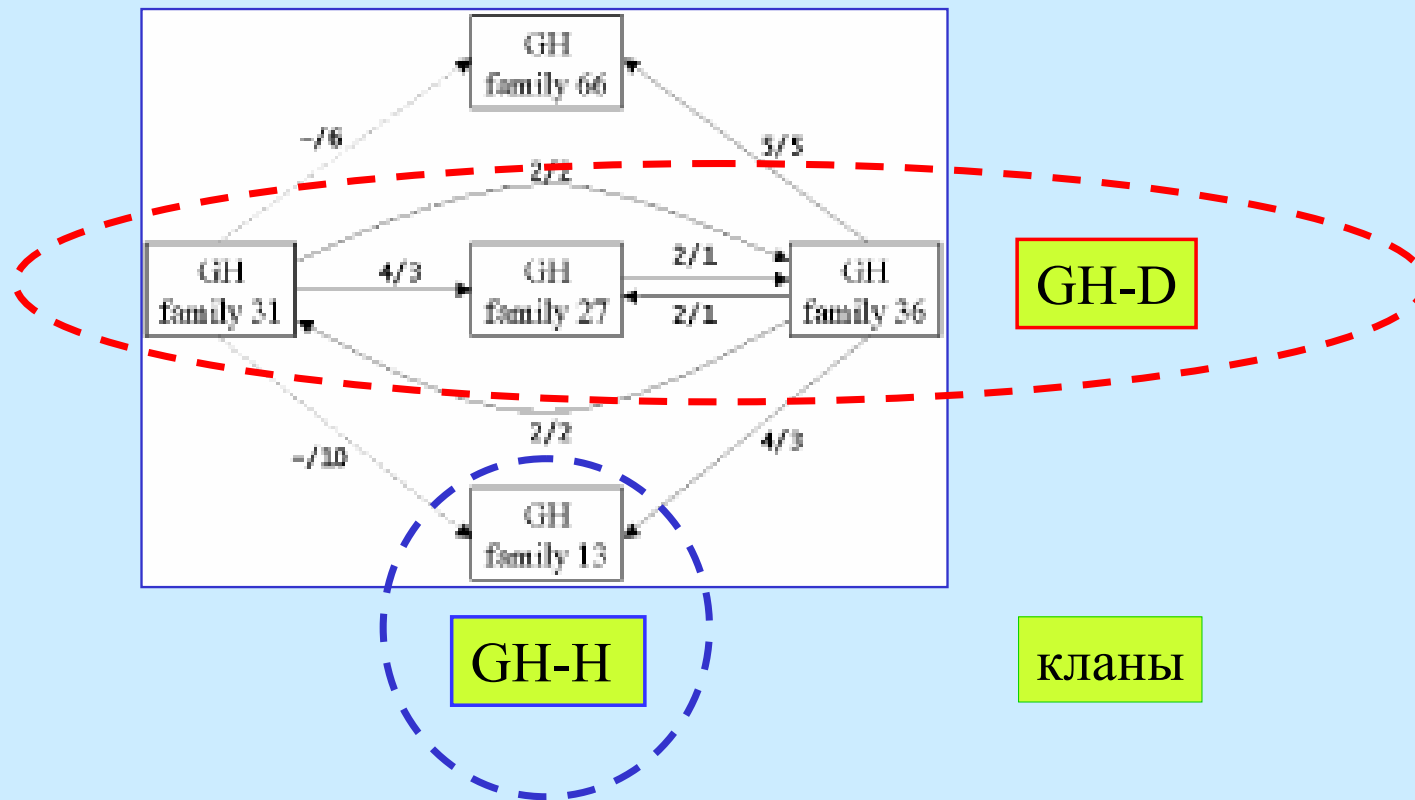
В колонках MgaM, MalA и YicI указано минимальное число итераций программы PSI-BLAST, необходимое для обнаружения с помощью последовательностей GH31-доменов этих белков представителя данного семейства (первая колонка). Число найденных представителей – общее число белков данного семейства, найденное в течение 10 итераций с помощью MgaM и 25 итераций с помощью MalA и YicI.

* При скрининге 19 сентября 2008 года программой PSI-BLAST базы данных аминокислотных последовательностей с помощью последовательности каталитического домена белка MalA в течение первых 25 итераций представителей семейств GH5 и COG3868 обнаружено не было. Однако, при аналогичных скринингах 20 августа 2007 года по итогам шестой итерации был обнаружен белок NP_214368 (*E*-value = 0.004), принадлежащий семейству COG3868, а 3 июня 2008 года по итогам десятой итерации был обнаружен белок AAC36862 (*E*-value = 0.0008), принадлежащий семейству GH5.



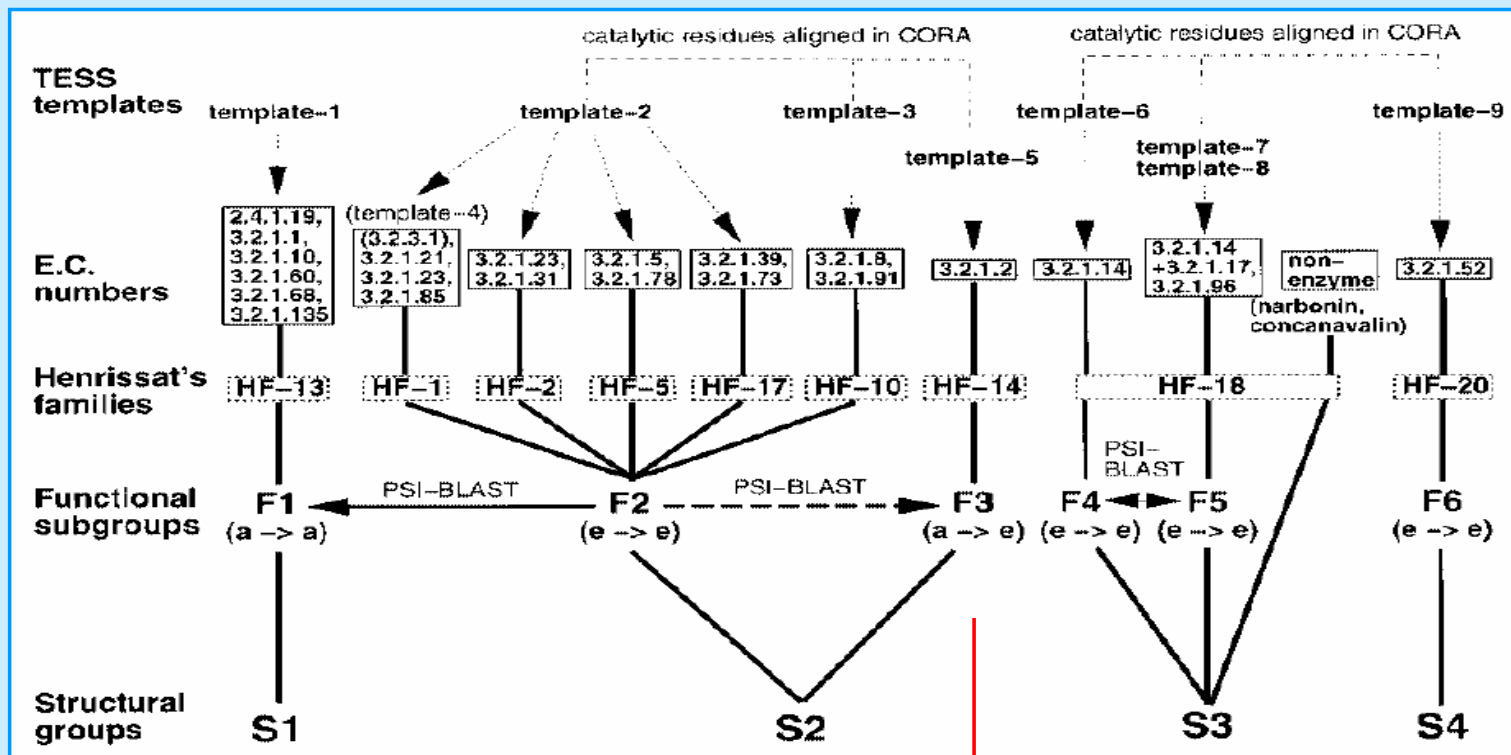
Родственные связи α -галактозидаз с другими гликозидазами (Rigden, 2002)

Rigden DJ. Iterative database searches demonstrate that glycoside hydrolase families 27, 31, 36, and 66 share a common evolutionary origin with family 13. *FEBS Lett.* 2002, 523(1-3):17-22.



Родственные связи $(\beta/\alpha)_8$ гликозидаз (Nagano et al., 2001)

Nagano N, Porter CT, Thornton JM. The $(\beta/\alpha)_8$ glycosidases: sequence and structure analyses suggest distant evolutionary relationships. Protein Eng. 2001, 14(11):845-855.



Кланы:

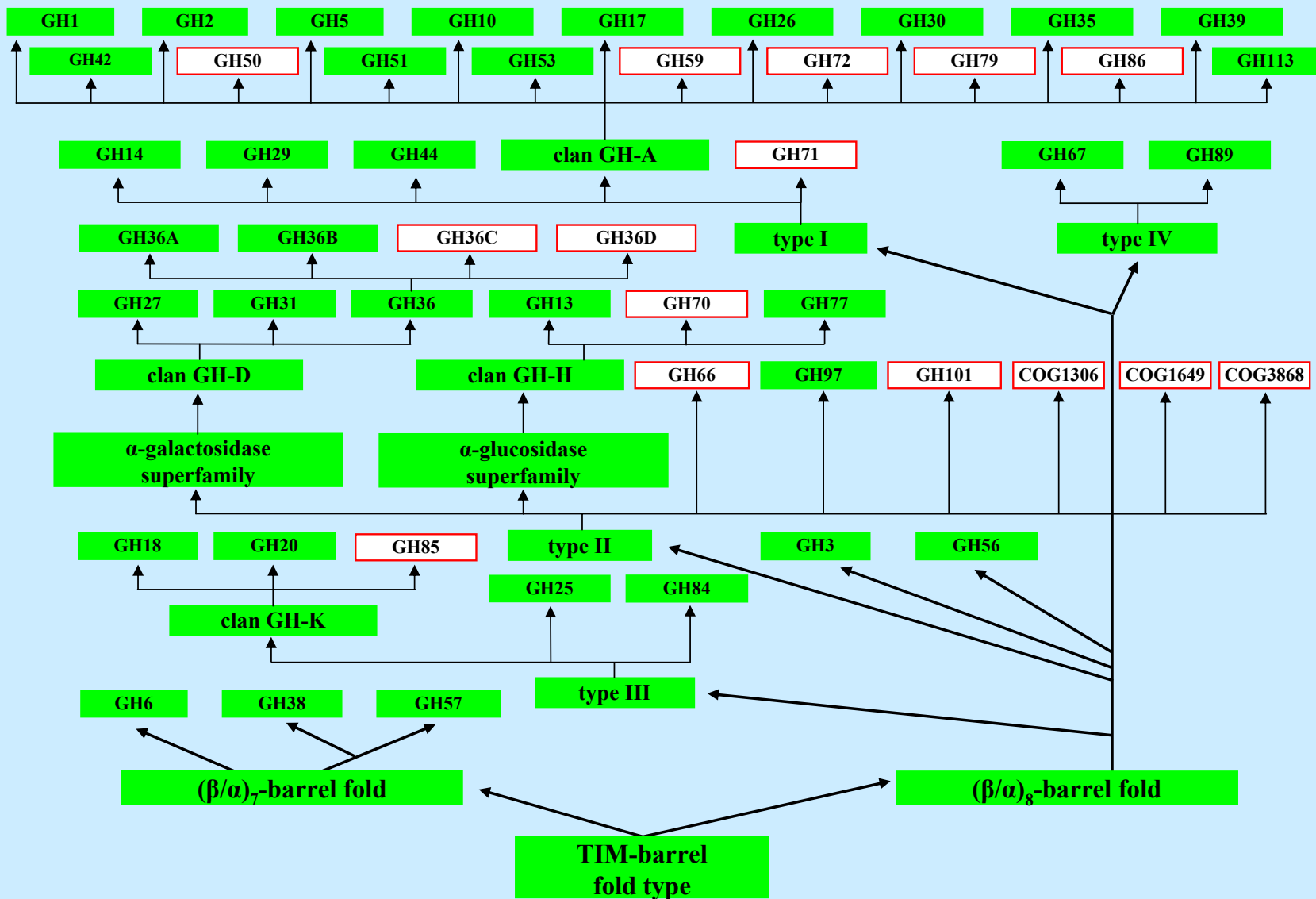
GH-H

GH-A

?

GH-K

Иерархическая классификация каталитических доменов гликозил-гидролаз



Объединение семейств гликозидаз в кланы

Клан	Семейства (GH)	Оптическая конфигурация	Трёхмерная структура
GH-A	1, 2, 5, 10, 17, 26, 30, 35, 39, 42, 50, 51, 53, 59, 72, 79, 86, 113	сохраняется (экв.)	$(\beta/\alpha)_8$ -barrel
GH-B	7, 16	сохраняется (экв.)	β -jelly roll
GH-C	11, 12	сохраняется (экв.)	β -jelly roll
GH-D	27, 31, 36	сохраняется (акс.)	$(\beta/\alpha)_8$ -barrel
GH-E	33, 34, 83	сохраняется (экв.)	6-fold β -propeller
GH-F	43, 62	меняется (экв.)	5-fold β -propeller
GH-G	37, 63	меняется (акс.)	$(\alpha/\alpha)_6$
GH-H	13, 70, 77	сохраняется (акс.)	$(\beta/\alpha)_8$ -barrel
GH-I	24, 46, 80	меняется (экв.)	$\alpha+\beta$
GH-J	32, 68	сохраняется (β -фуранозид)	5-fold β -propeller
GH-K	18, 20, 85	сохраняется (экв.)	$(\beta/\alpha)_8$ -barrel
GH-L	15, 65	меняется (акс.)	$(\alpha/\alpha)_6$
GH-M	8, 48	меняется (экв.)	$(\alpha/\alpha)_6$
GH-N	28, 49	меняется (акс.)	$(\beta)_3$ -solenoid

Объединение семейств гликозидаз в кланы

Клан	Семейства (GH)	Оптическая конфигурация	Трёхмерная структура
GH-A	1, 2, 5, 10, 17, 26, 30, 35, 39, 42, 50, 51, 53, 59, 72, 79, 86, 113	сохраняется (экв.)	$(\beta/\alpha)_8$ -barrel
GH-B	7, 16	сохраняется (экв.)	β -jelly roll
GH-C	11, 12	сохраняется (экв.)	β -jelly roll
GH-D	27, 31, 36	сохраняется (акс.)	$(\beta/\alpha)_8$ -barrel
GH-E	33, 34, 83	сохраняется (экв.)	6-fold β -propeller
GH-F	43, 62	меняется (экв.)	5-fold β -propeller
GH-G	37, 63	меняется (акс.)	$(\alpha/\alpha)_6$
GH-H	13, 70, 77	сохраняется (акс.)	$(\beta/\alpha)_8$ -barrel
GH-I	24, 46, 80	меняется (экв.)	$\alpha+\beta$
GH-J	32, 68	сохраняется (β -фуранозид)	5-fold β -propeller
GH-K	18, 20, 85	сохраняется (экв.)	$(\beta/\alpha)_8$ -barrel
GH-L	15, 65	меняется (акс.)	$(\alpha/\alpha)_6$
GH-M	8, 48	меняется (экв.)	$(\alpha/\alpha)_6$
GH-N	28, 49	меняется (акс.)	$(\beta)_3$ -solenoid

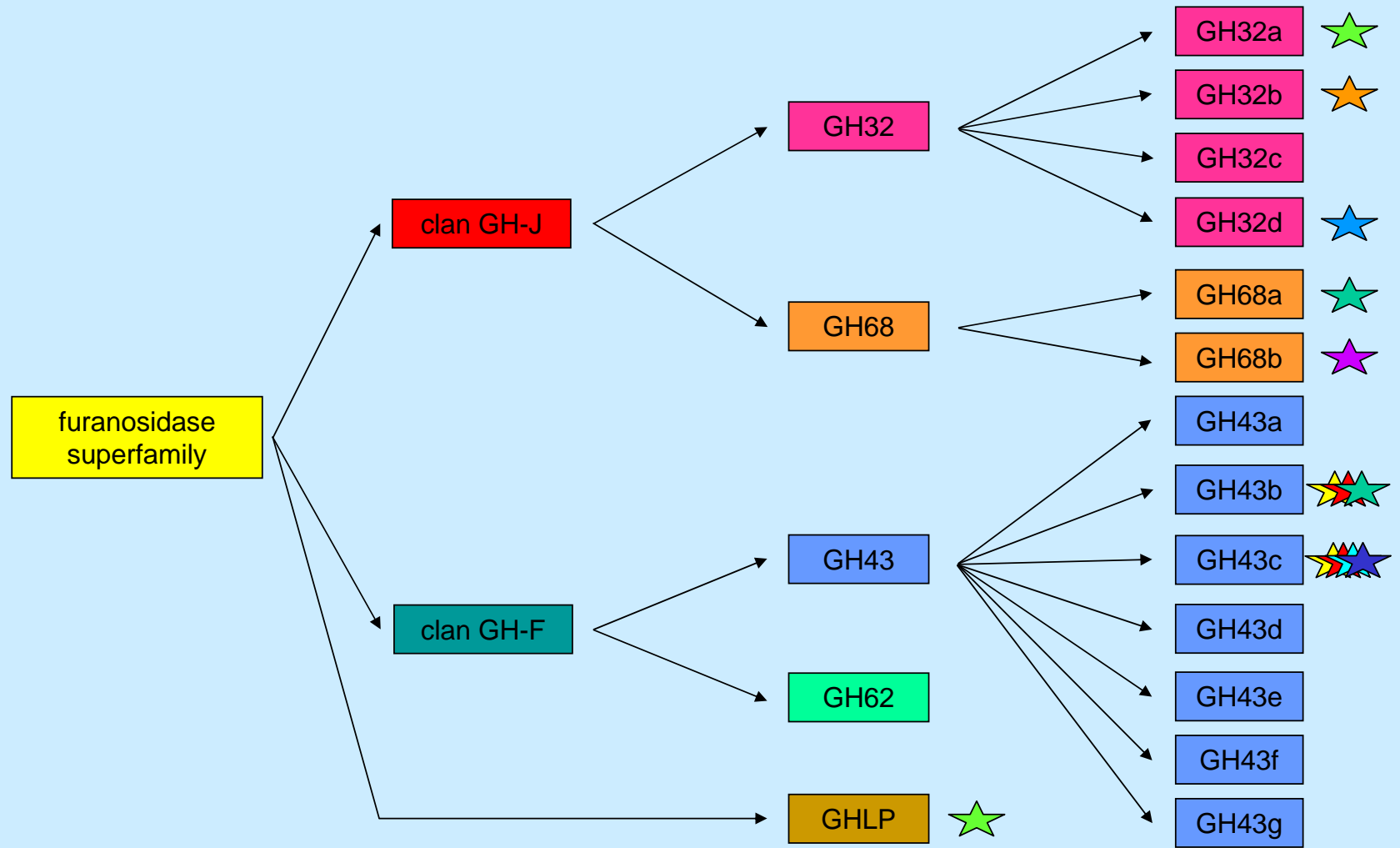
Объединение семейств гликозидаз в кланы

Клан	Семейства (GH)	Оптическая конфигурация	Трёхмерная структура
GH-A	1, 2, 5, 10, 17, 26, 30, 35, 39, 42, 50, 51, 53, 59, 72, 79, 86, 113	сохраняется (экв.)	$(\beta/\alpha)_8$ -barrel
GH-B	7, 16	сохраняется (экв.)	β -jelly roll
GH-C	11, 12	сохраняется (экв.)	β -jelly roll
GH-D	27, 31, 36	сохраняется (акс.)	$(\beta/\alpha)_8$ -barrel
GH-E	33, 34, 83	сохраняется (экв.)	6-fold β -propeller
GH-F	43, 62	меняется (экв.)	5-fold β -propeller
GH-G	37, 63	меняется (акс.)	$(\alpha/\alpha)_6$
GH-H	13, 70, 77	сохраняется (акс.)	$(\beta/\alpha)_8$ -barrel
GH-I	24, 46, 80	меняется (экв.)	$\alpha+\beta$
GH-J	32, 68	сохраняется (β -фуранозид)	5-fold β -propeller
GH-K	18, 20, 85	сохраняется (экв.)	$(\beta/\alpha)_8$ -barrel
GH-L	15, 65	меняется (акс.)	$(\alpha/\alpha)_6$
GH-M	8, 48	меняется (экв.)	$(\alpha/\alpha)_6$
GH-N	28, 49	меняется (акс.)	$(\beta)_3$ -solenoid

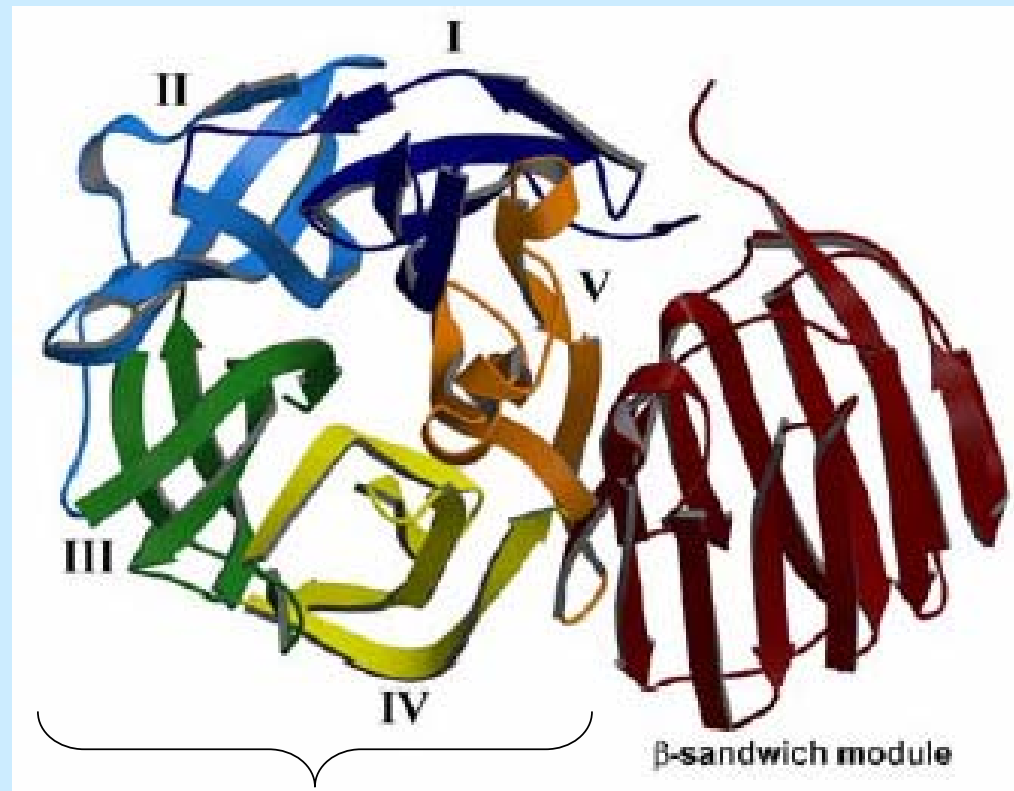
β-fructosidase (furanosidase) superfamily

Family	GH32	GH43	GH62	GH68	GHLP
Clan	GH-J	GH-F	GH-F	GH-J	Not known
COG / KOG	COG1621 / KOG0228	COG3507	None	None	COG2152
Known enzymatic activities	EC 2.4.1.99 EC 2.4.1.100 EC 2.4.1.x EC 3.2.1.7 <u>EC 3.2.1.26</u> EC 3.2.1.65 EC 3.2.1.80	EC 3.2.1.8 EC 3.2.1.37 <u>EC 3.2.1.55</u> EC 3.2.1.99	<u>EC 3.2.1.55</u>	EC 2.4.1.x EC 2.4.1.9 EC 2.4.1.10 <u>EC 3.2.1.26</u>	Not known
Molecular mechanism	Retaining	Inverting	Not known	Retaining	Not known
Origin	Eukaryota: Euglenozoa Fungi Parabasalidea Viridiplantae Eubacteria: Acidobacteria Actinobacteria Bacteroidetes Chloroflexi Firmicutes Fusobacteria Planctomycetes Proteobacteria Thermotogales Archaea: Euryarchaeota	Eukaryota: Fungi Viridiplantae Eubacteria: Acidobacteria Actinobacteria Bacteroidetes Firmicutes Planctomycetes Proteobacteria Thermotogales Verrucomicrobia	Eukaryota: Fungi Eubacteria: Actinobacteria Proteobacteria Verrucomicrobia	Eubacteria: Actinobacteria Firmicutes Proteobacteria Archaea: Euryarchaeota	Eukaryota: Fungi Metazoa Viridiplantae Eubacteria: Actinobacteria Aquificales Cyanobacteria Firmicutes Proteobacteria Thermotogales Archaea: Crenarchaeota Euryarchaeota

A hierarchical structure of the β -fructosidase (furanosidase) superfamily



**Пространственная структура белков семейства GH32
(инулиназа *Thermotoga maritima*)**



N-концевой
каталитический домен в
виде пяти-лопастного
 β -пропеллера

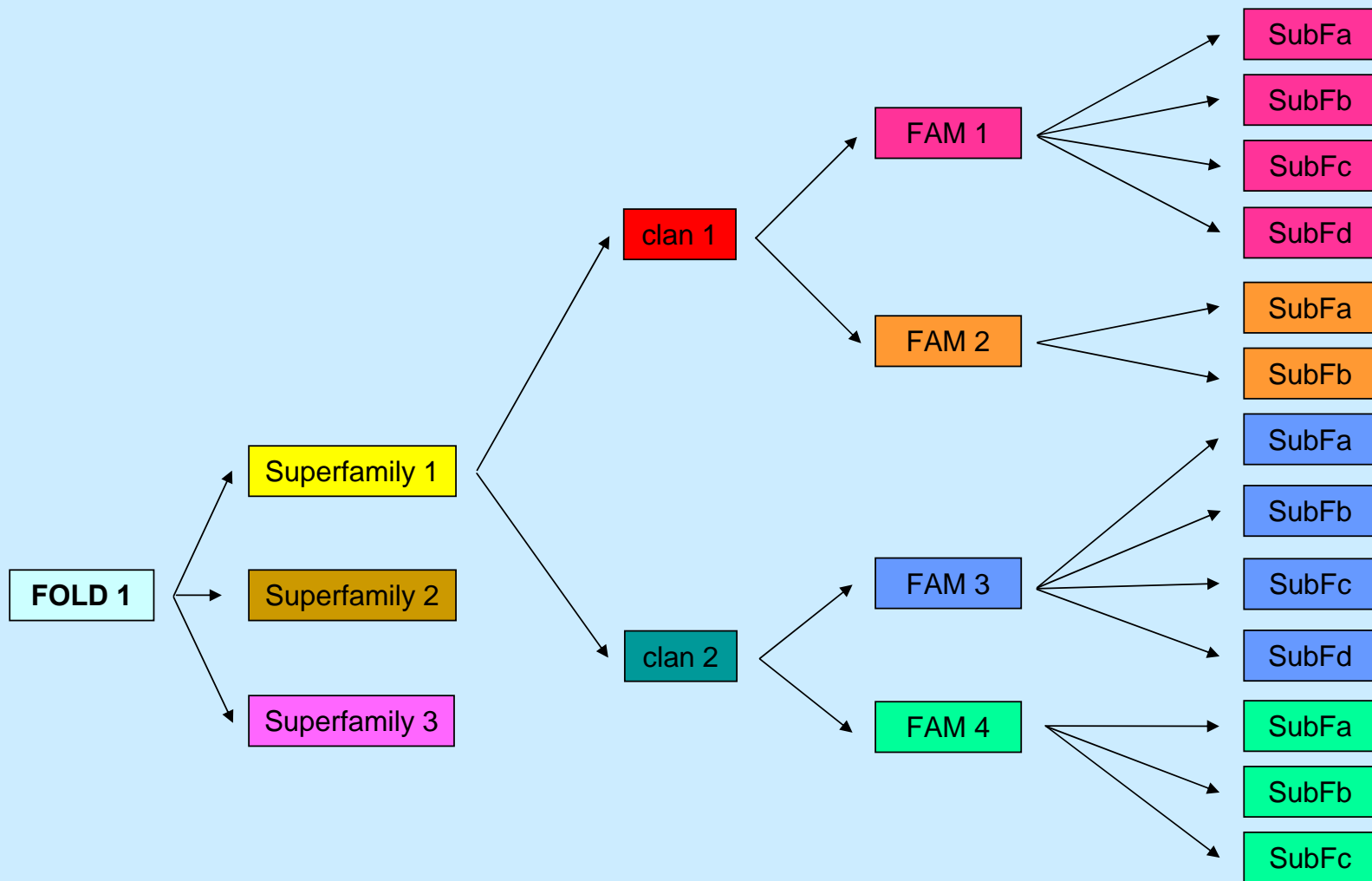
Объединение семейств гликозидаз в кланы

Клан	Семейства (GH)	Оптическая конфигурация	Трёхмерная структура
GH-A	1, 2, 5, 10, 17, 26, 30, 35, 39, 42, 50, 51, 53, 59, 72, 79, 86, 113	сохраняется (экв.)	$(\beta/\alpha)_8$ -barrel
GH-B	7, 16	сохраняется (экв.)	β -jelly roll
GH-C	11, 12	сохраняется (экв.)	β -jelly roll
GH-D	27, 31, 36	сохраняется (акс.)	$(\beta/\alpha)_8$ -barrel
GH-E	33, 34, 83	сохраняется (экв.)	6-fold β -propeller
GH-F	43, 62	меняется (экв.)	5-fold β -propeller
GH-G	37, 63	меняется (акс.)	$(\alpha/\alpha)_6$
GH-H	13, 70, 77	сохраняется (акс.)	$(\beta/\alpha)_8$ -barrel
GH-I	24, 46, 80	меняется (экв.)	$\alpha+\beta$
GH-J	32, 68	сохраняется (β -фуранозид)	5-fold β -propeller
GH-K	18, 20, 85	сохраняется (экв.)	$(\beta/\alpha)_8$ -barrel
GH-L	15, 65	меняется (акс.)	$(\alpha/\alpha)_6$
GH-M	8, 48	меняется (экв.)	$(\alpha/\alpha)_6$
GH-N	28, 49	меняется (акс.)	$(\beta)_3$ -solenoid

Strohmeier
et al.
(2004)

?

Иерархическая классификация гликозил-гидролаз



Thank you for the attention!

Please, visit my web-site to find a list of publications:
<http://bioinform.genetika.ru/members/Naumoff/index.htm>