Evolutionary relationships among prokaryotic fructosyltransferases

Yamira Quintero-de la Oliva*, Alexis Musacchio-Lasa, Ana Gabriela Martínez-Peña, Carmen Menéndez-Rodríguez, Lázaro Hernández-García

[1] Enzyme Technology Group, Agricultural Research Division, CIGB Havana, Cuba.
[2] System Biology Department, Biomedical Research Division, CIGB Havana, Cuba.
* yamira.quintero@cigb.edu.cu

ABSTRACT

In nature, fructans are synthesized from sucrose by few plant families and a wide range of microorganisms, including fungi, Archae and bacteria. Prokaryotic fructosyltransferases (levansucrase and inulosucrase) are grouped in the Glycosyl Hydrolase (GH) family -68 and the vast majority belong to five bacterial phyla (mainly Protobacteria, Actinobacteria and Firmicutes) and five archaeal families. These enzymes fold into a common 5-bladed beta-propeller with the active site located in a deep axial pocket. The catalytic triad (Asp/Asp/Glu) is strictly conserved at the pocket bottom in all GH-68 members. In this study, bioinformatics tools were used to predict distinctive structural features among GH-68 enzymes in regions that are not directly involved in substrate binding and catalysis. Comparative phylogenetic analysis and the distinctive non-catalytic characteristics were used as criteria for the evolutionary classification of proteins. Our in-silico predictions revealed several presumed cases of horizontal GH-68 gene acquisition from a taxonomically distant species.
Key words: fructosyltransferases, GH68, evolution.

RESUMEN

En la naturaleza los fructanos son sintetizados a partir de la sacarosa por algunas familias de plantas y un amplio rango de microorganismos, que incluyen hongos, arqueas y bacterias. Las fructosiltransferasas procariotas (levanasacarasa e inulosacarasa) son miembros de la familia glicosil hidrolasas (GH) -68 y la mayoría pertenecen a cinco filos bacterianos (principalmente Protobacteria, Actinobacteria y Firmicutes) y cinco familias de arqueas. Estas enzimas presentan una arquitectura del tipo propela β de 5 pétalos con la presencia de tres residuos catalíticos ácidos (Asp / Asp / Glu) en el sitio activo estrictamente conservados. En este estudio, se utilizaron herramientas bioinformáticas para predecir características estructurales distintivas entre las enzimas GH-68 en regiones que no están directamente involucradas en la unión del sustrato y la catálisis. El análisis filogenético comparativo y las características distintivas no catalíticas, se utilizaron como criterios para la clasificación evolutiva de las proteínas. Las predicciones in sílico revelaron varias fructosiltransferasas dentro de la familia GH-68 de adquisición de gen horizontal de una especie taxonómicamente distante.
Palabras clave: fructosiltransferasas, GH68, evolución.

INTRODUCTION

In nature, fructans are synthesized from sucrose by few plant families and a wide range of microorganisms, including bacteria, archaea and fungi. Plant and fungal fructosyltransferases (FTFs) are grouped in the Glycosyl Hydrolase (GH) family 32 together with eukaryotic and prokaryotic invertases, sucrose-6P hydrolases and fructanases. GH family 68 comprises prokaryotic FTFs (levansucrase and inulosucrase) and beta-fructofuranosidases sequenced so far from about 400 species of

three main bacterial phyla (Protobacteria, Actinobacteria and Firmicutes) and five archaea families. The vast majority of the GH-68 entries in database correspond to genome sequencing projects. Only bacterial members have been characterized functionally (46) or structurally (6) (http://www.cazy.org). All characterized bacterial FTFs catalyze both sucrose hydrolysis and transfructosylation reactions. GH-68 enzymes fold into a common 5-bladed beta-propeller containing the strictly conserved catalytic triad (DDE) at the bottom of a deep axial pocket. Natural point mutations in other few amino acids at the active site are supposed to be responsible for differences in the transferase/hydrolase rate, the linkage type, and the polymerization degree of the fructosylated product [1-6].

In this study, we search for distinctive features between GH-68 enzymes in regions that are not directly involved in substrate binding and catalysis, as a supporting criterion to establish the evolutionary relationship among GH-68 enzymes and to identify horizontal gene transfer events between taxonomically distant species.

MATERIAL AND METHODS

Data analysis

Sequences were retrieved by BlastP comparisons of levansucrases from representative families against the database of non-redundant protein sequences (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) and Simple Phylogeny (https://www.ebi.ac.uk/Tools/phylogeny/simple_phylogeny/) were used for sequence alignment and tree generation, respectively.

RESULTS AND DISCUSSION

Protein phylogenetic analysis was accompanied with a comparative search of characteristic traits in regions that are not directly involved in substrate binding and catalysis but may better reflect the enzyme evolutionary history. Available bioinformatics tools allowed to predict: -the presence/absence of a signal peptide and a signature motif in the protein N-terminal region, -the size and isoelectric point of the putative mature enzyme, and -the formation of a fold-stabilizing disulfide bridge or its topologically/functionally complementary calcium-binding site. The conservation degree of distinctive non-catalytic traits within the members of adjacent phylogenetic clades was used as a supporting criterion to establish the evolutionary connections among GH-68 proteins and to identify candidate HGT (Horizontal Gene Transfer) events between taxonomically distant species with similar habitats. At least 21 clades are clearly defined in Phylogenetic tree the of 392 available GH-68 sequences. Table I it summary of distinctive non-catalytic traits in GH68 enzymes grouped by their phylogenetic relationships.

The all 58 Archaea species clustered together in exclusive neighboring clades (1, 2 and 3) reflecting no evidence of horizontal gene acquisition. The N-terminal signature motif W(T/S)(R/I)AD(A/I/M/V)(L/M/R/I) of unknown function is shared by most Proteobacteria (clades 7, 8, 9, 10, 11, 12 and 14) and Actinobacteria (clades 6, 12, 13) suggesting a common old ancestor for two groups that evolved separately. In this presumption, it is remarkable the combined presence of signal peptide and fold-stabilizing disulfide bond in all 91 proteins within clades 6, 12, 13 and 14, and their absolute absence in the 62 proteins (Alpha- and Gammaproteobacteria) comprising the clades 7, 8, 9, 10, 11.

Table 1. Summary of distinctive non-catalytic traits in GH68 enzymes grouped by their phylogenetic relationships

Phylum or Class	Family	Clades (# proteins)	Encoded protein (# aa)	Signal peptide (# aa)	Mature protein (Daltons)	pI	Disulfide bond	
							Yes	No
Euryarchaeota	Haloarculaceae	1(12)	442	0/12	49230.08	4.61	3	9
	Halobacteriaceae	2(2)	429	0/2	47834.00	4.78	2	0
		3(4)	435	1/4 (38aa)	48532.04	4.69	0	4
	Haloferacaceae	1(5)	435	0/5	49313.60	4.66	5	0
		2(10)	432	0/10	48180.20	4.56	9	1
		3(1)	422	0/1	46795.00	4.41	0	1
	Halorubraceae	1(10)	465	0/10	50882.40	4.52	3	7
		2(4)	446	0/3	49217.75	4.66	4	0
	Natrialbaceae	1(3)	452	0/3	50493.00	4.53	0	3
		3(7)	448	7/7 (37aa)	45918.43	4.49	0	7
Alphaproteobacteria	Acetobacteraceae	7(14)	438	0/14	48258.36	4.99	0	14
		14(9)	566	7/9 (31aa)	59082.67	5.91	9	0
	Caulobacteraceae	19(2)	384	0/2	41758.00	5.83	0	2
	Erythrobacteraceae	19(14)	370	0/14	40313.50	5.28	0	14
	Rhodobacteraceae	17(3)	468	3/3 (26aa)	48738.33	4.38	0	3
	Sphingomonadaceae	7(2)	418	0/2	46430.50	4.84	0	2
		19(29)	376	0/29	40890.48	5.62	0	29
Betaproteobacteria	Burkholderiaceae	14(25)	526	25/25 (31aa)	53356.16	6.23	25	0
Gammaproteobacteria	Aeromonadaceae	12(1)	524	1/1 (25aa)	54793.00	5.16	1	0
	Alteromonadaceae	12(1)	528	1/1 (27aa)	55508.00	4.78	1	0
	Enterobacteriaceae	9(1)	426	0/1	47091.00	4.69	0	1
	Erwiniaceae	8(5)	420	0/5	46543.00	4.91	0	5
		10(3)	415	0/3	46439.00	4.87	0	3
	Halomonadaceae	8(1)	416	0/1	46200.00	4.54	0	1
	Moraxellaceae	10(2)	434	0/2	47964.00	5.19	0	2
	Morganellaceae	8(1)	423	0/1	47698.00	5.08	0	1
	Oceanospirillaceae	12(1)	524	1/1 (25aa)	55185.00	4.75	1	0
	Pectobacteriaceae	9(6)	428	0/6	47588.00	5.02	0	6
	Pseudoalteromonadaceae	12(2)	525	2/2 (24aa)	55652.50	4.73	2	0
	Pseudomonadaceae	9(6)	424	0/6	47002.00	4.86	0	6
		10(6)	426	0/6	47174.67	4.95	0	6
		11(10)	416	0/10	46103.60	5.13	0	10
		12(2)	525	2/2 (27aa)	54595.50	4.95	2	0
	Vibrionaceae	8(2)	419	0/2	46592.00	4.73	0	2
		12(1)	523	1/1 (27aa)	54991.00	4.67	1	0
	Yersiniaceae	10(3)	416	0/3	46146.33	4.81	0	3
Deltaproteobacteria	Desulfovibrionaceae	15(3)	627	3/3 (51aa)	61950.67	3.85	2	1
Bacteroidetes	Cyclobacteriaceae	16(1)	456	1/1 (33aa)	47199.00	4.74	0	1
Actinobacteria	Actinomycetaceae	12(6)	622	6/6 (33aa)	63725.83	4.87	6	0
	Cellulomonadaceae	6(2)	467	2/2 (28aa)	48778.50	4.75	2	0
		12(1)	558	1/1 (29aa)	57952.00	4.74	1	0
	Dermabacteraceae	13(2)	524	2/2 (39aa)	53276.00	4.66	0	2
	Microbacteriaceae	12(10)	541	10/10 (34aa)	55689.00	4.78	10	0
	Micrococcaceae	12(3)	530	3/3 (35aa)	54817.33	5.16	3	0
		13(12)	532	12/12 (37aa)	54050.50	5.74	12	0
	Mycobacteriaceae	4(1)	445	1/1 (29aa)	48787.00	4.22	0	1
		13(1)	514	1/1 (42aa)	52426.00	5.63	0	1
		21(1)	475	1/1 (31aa)	50041.00	6.18	0	1
	Pseudonocardiaceae	18(2)	375	2/2	40964.00	5.82	0	2
	Streptomycetaceae	6(12)	441	12/12 (27aa)	46070.42	8.56	12	0
		18(1)	373	0/1	40786.00	5.90	0	1
	Streptosporangiaceae	18(11)	371	0/11	40396.00	6.13	0	11
Firmicutes	Bacillaceae	4(14)	506	11/14 (30aa)	53802.36	4.83	6	8
		21(30)	481	30/30 (29aa)	50858.90	5.79	0	30
	Clostridiaceae	4(3)	428	0/3	48348.00	7.84	0	3
		21(2)	498	2/2 (31aa)	52266.00	5.92	0	2
	Lactobacillaceae	20(19)	816	19/19 (35aa)	85839.21	6.53	0	19
	Leuconostocaceae	20(1)	900	1/1 (46aa)	91999.00	5.12	0	1
		21(3)	1251	3/3 (39aa)	124256.75	5.29	0	3
	Paenibacillaceae	4(3)	424	2/3 (29aa)	44084.00	5.62	2	1
		21(20)	493	20/20 (28aa)	51822.35	5.41	0	20
	Ruminococcaceae	21(1)	485	1/1 (29aa)	50503.00	5.00	0	1
	Sporolactobacillaceae	21(2)	485	2/2 (29aa)	51271.50	5.70	0	2
	Streptococcaceae	20(13)	851	13/13 (42aa)	87768.15	4.75	0	13
Deinococcus-Thermus	Deinococcaceae	5(2)	514	2/2 (31aa)	52133.00	5.14	2	0

Other Actinobacteria (clade 18) and Alphaproteobacteria (clade 19) produce shorter proteins with a rather ambiguous N-terminal motif and without either signal peptide or disulfide bond. The proteins of 90 Firmicutes species (clades 20 and 21), containing signal peptide and lacking the N-terminal motif, are particularly distinguishable by the Ca^{2+}-cofactor binding site that assumes the fold-stabilizing function as alternative to the general absence of Cys residues. Orthologues from other 20 Firmicutes species (clade 4), often with saline habitats, share the highest identities (above 40%) and signature non-catalytic traits with *Euryarchaeota* (clades 1, 2 and 3) and the Actinobacteria families Cellulomonadaceae and Streptomycetaceae (clade 6). The separate clades 5 (Deinococcus-Thermus), 15

(Deltaproteobacteria), 16 (Bacteroidetes), and 17 (Alphaproteobacteria, family Rhodobacteraceae), with only 9 members in total, showed very low identities (below 35%) in independent BlastP comparisons. A different catalytic triad (EED) makes the three reported sequences from *Desulfovibrio africanus* (clade 15) even more peculiar. Several candidate HGT events between bacteria were detected by comparing the topologies of the GH-68 tree against the corresponding 16S rRNA "reference" tree. We consider that the double advantage of utilizing exogenous sucrose as an energy source and producing a protective polysaccharide is a positive selective force driving retention of horizontally acquired GH-68 genes.

CONCLUSIONS

In this study, phylogenetic analysis and distinctive features between GH-68 enzymes in regions that are not directly involved in substrate binding and catalysis, permitted establish the enzyme evolutionary relationship.

Archaea species reflecting no evidence of horizontal gene acquisition. The N-terminal signature motif W(T/S)(R/I)AD(A/I/M/V)(L/M/R/I) of unknown function is shared by most Proteobacteria (clades 7, 8, 9, 10, 11, 12 and 14) and Actinobacteria (clades 6, 12, 13) suggesting a common old ancestor for two groups that evolved separately. Several candidate HGT events between bacteria were detected by comparing the topologies of the GH-68 tree. We consider that the double advantage of utilizing exogenous sucrose as an energy source and producing a protective polysaccharide is a positive selective force driving retention of horizontally acquired GH-68 genes.

REFERENCES

1. Meng G, Fütterer K (2003) Nat Struct Biol 10:935-941 (DOI:10.1038/nsb974).
2. Martinez-Fleites et al. (2005) Biochem J 390:19-27 (DOI:10.1042/BJ20050324).
3. Pijning et al. (2011) J Mol Biol 412:80-93 (DOI:10.1016/j.jmb.2011.07.031).
4. Strube et al. (2011) J Biol Chem 286:17593-17600 (DOI:10.1074/jbc.M110.203166).
5. Tonozuka et al. (2012) Enz Microb Technol 51:359-365 (DOI:10.1016/j.enzmictec.2012.08.004).
6. Wuerges et al. (2015) J Struct Biol 191:290-298 (DOI:10.1016/j.jsb.2015.07.010).