

Raffinose Family of Oligosaccharides: The plant specific compounds

Ishita Barary, *PG Department of Botany, Bethune College, Kolkata-6*

Sritama Mukherjee, *Assistant Professor, PG Department of Botany, Bethune College, Kolkata-6*

Abstract

Raffinose family of oligosaccharides are plant specific compounds. Biosynthesis of RFOs begin with the activity of Galactinol synthase, that galactosylates myo-inositol to produce galactinol. Raffinose and the subsequent higher molecular weight RFOs (Stachyose, Verbascose, and Ajugose) are synthesized from sucrose by the subsequent addition of activated galactose moieties donated by Galactinol. Raffinose Synthase (RafS) synthesizes the first RFO of the pathway i.e. trisaccharide raffinose from galactinol and sucrose. In the present work an attempt has been made to search the RafS sequences among the evolutionary diverse taxa to trace their origin and evolution. Also, the conserved domains are searched and a 3-D putative protein model of RafS is predicted.

Key words : Raffinose synthase, Galactinol synthase, RFO.

Introduction

Raffinose Family of Oligosaccharides (RFOs) are α -1,6-galactosyl extensions of sucrose (Suc). This group of oligosaccharides are found in plants and are known to serve as desiccation protectant in seeds, as transport sugar in phloem sap and as storage sugars. Galactinol synthase (GolS, EC 2.4.1.123) is the key enzyme that transfers a galactosyl group to RFOs which is donated by Galactinol (Gol; 1-O- α -D-galactopyranosyl-L-myo-inositol). Within the RFO biosynthetic pathway, the two other major enzymes are a) Raffinose Synthase (RafS, EC 2.4.1.82) that transfers galactosyl moiety from Gol to Suc and synthesizes Raffinose (Raf), b) Stachyose Synthase (StaS, EC 2.4.1.67) that further uses Gol as a galactosyl donor to Raf and produces the tetrasaccharide Stachyose (Sta). Generally the biosynthesis of RFOs proceeds by galactosyl transfer from a suitable galactosyl donor to an acceptor, which may be sucrose, raffinose or its higher homologues. In each step a molecule of myo-inositol is released as the leaving group (**Fig.1**). There are higher molecular weight RFOs (Stachyose, Verbascose, and Ajugose) present in some plants but Raf and Sta are the major sugars in this group and this study will concentrate on the structure, function and regulation pattern of RafS in plants.

RFOs have a wide range of predicted functions. They are synthesized and stored in seeds (Peterbayer and Richter, 2001) protecting the embryo from maturation associated desiccation (Downie et al., 2003), participate in several cellular functions including: transport and storage of carbon, signal transduction, membrane trafficking and mRNA export, act as signalling molecule following pathogen attack and wounding and accumulate in vegetative tissues in response to a range of abiotic stresses (Hinch et al., 2003; Taji et al., 2002; Peters et al., 2007; Zuther et al., 2012). RFOs are currently emerging as crucial molecules during stress response in plants, because of their membrane stabilizing, antioxidant and perhaps predicted signalling functions (Nishizawa et al., 2006; Valluru and Van den Ende, 2011). It has also been reported that Raf exists in the chloroplast (Foyer & Shigeoka, 2011) and may play a role in stabilizing photosystem II (Knaupp et al., 2011).

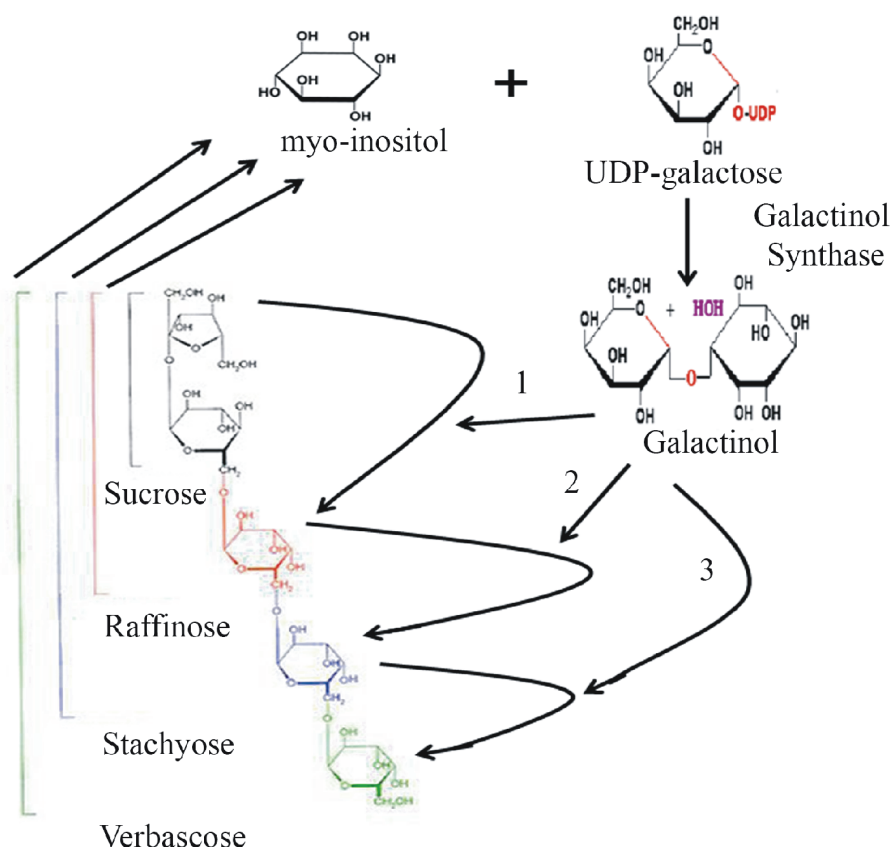


Fig.1. Synthesis of RFOs from galactinol where 1: Raffinose Synthase, 2: Stachyose Synthase, 3: Verbascose Synthase.

Materials and Methods

Building Sequence Repository

The gene/protein sequences for RafS were obtained from NCBI, Phytozome, Consurf, and SwissProt.

Analysis of Gene/Protein Sequences

Alignments of sequences of gene/protein were generated with the help of NCBI (<http://www.ncbi.nlm.nih.gov.in>) and Phytozome(<http://www.phytozome.net>).

Identification of Common Catalytic Domain

The conserved motifs and residues were derived from Consurf (<http://consurf.tau.ac.il/>; Landau et al., 2005). It is a web-based tool that automatically calculates evolutionary conservation scores and maps them on protein structures via a user-friendly interface. Protein structure were built with Swissprot, an annotated protein sequence database.

Analysis of evolutionary tree

Evolutionary relationships were deduced in ClustalW (www.ebi.ac.uk/Tools/msa/clustalw2/), a popular program used for multiple sequence alignment and for preparing phylogenetic trees. Hundred rounds of bootstrapping verified reliability of the trees.

Results

Identification of Conserved Domain among RafS

To identify a common catalytic domain, a BLAST search with functionally annotated RafS sequence from *Oryza sativa* (Accession No. NM_001063477) registered 21 monocot and 32 dicot RafS sequences (**Table. 1**).

Table.1. List of organisms containing RafS with the key features of the gene/protein.							
Sl. No.	Organisms	Accession No.		Gene & Protein size		CDs	Expression
		Nucleotide	Protein	Gene(bp)	Protein (aa)		
MONOCOTS							
1.	<i>Oryza sativa Japonica</i>	NM_001063477	NP_001056942.2	2921	451	complete	hypothetical
2.	<i>Oryza brachyantha</i>	XM_006655778	XP_006655841.1	2696	786	complete	hypothetical
3.	<i>Sorghum bicolor</i>	XM_002437883	XP_002437928.1	2585	801	complete	hypothetical
4.	<i>Setaria italica</i>	XM_004964668	XP_004964725.1	2973	854	complete	predicted
5.	<i>Setaria italica</i>	XM_004964666	XP_004964723.1	2414	767	complete	predicted
6.	<i>Brachypodium distachyon</i>	XM_003564191	XP_003564239.2	2720	842	complete	predicted
7.	<i>Musa acuminata</i>	XM_009409099	XP_009407374.1	2881	851	complete	predicted
8.	<i>Phoenix dactylifera</i>	XM_008795738	XP_008793960.1	2712	775	complete	predicted
9.	<i>Phoenix dactylifera</i>	XM_008780978	XP_008779200.1	2425	794	complete	predicted
10.	<i>Phoenix dactylifera</i>	XM_008808824	XP_008807046.1	2912	871	complete	predicted
11.	<i>Setaria italica</i>	XM_004955741	XP_004955798.1	2657	781	complete	predicted
12.	<i>Brachypodium distachyon</i>	XM_003557346	XP_003557394.1	2926	762	complete	predicted
13.	<i>Oryza brachyantha</i>	XM_006651876	XP_006651939.1	2292	765	complete	predicted
14.	<i>Setaria italica</i>	XM_004981277	XP_004981334.1	2704	768	complete	predicted
15.	<i>Zea mays</i>	XM_008667421	XP_008665643.1	2982	764	complete	predicted
16.	<i>Musa acuminata</i>	XM_009400152	XP_009398427.1	3243	766	complete	predicted
17.	<i>Amborella trichopoda</i>	XM_006856458	XP_006856520.2	2742	761	complete	predicted
18.	<i>Musa acuminata</i>	XM_009410179	XP_009408454.1	2688	769	complete	predicted
19.	<i>Zea mays</i>	XM_008664846	XP_008663068.1	2544	678	complete	predicted
20.	<i>Phoenix dactylifera</i>	XM_008779856	XP_008778078.1	3175	793	complete	predicted
21.	<i>Elaeis guineensis</i>	XM_010908420	XP_010906722.1	2915	860	complete	predicted
DICOTS							
1.	<i>Eucalyptus grandis</i>	XM_010051754	XP_010050056.1	2699	774	complete	predicted
2.	<i>Eucalyptus grandis</i>	XM_010051753	XP_010050055.1	2715	784	complete	predicted
3.	<i>Erythranthe</i>	XM_012984876	XP_012840330.1	2879	849	complete	predicted
4.	<i>Fragaria vesca</i>	XM_004287234	XP_004287282.1	2454	732	complete	predicted
5.	<i>Pyrus x bretschneideri</i>	XM_009365707	XP_009363982.1	2726	749	complete	predicted
6.	<i>Pyrus x bretschneideri</i>	XM_009365706	XP_009363981.1	2845	832	complete	predicted
7.	<i>Nelumbo nucifera</i>	XM_010243275	XP_010241577.1	2978	786	complete	predicted
8.	<i>Nelumbo nucifera</i>	XM_010243274	XP_010241576.1	2967	786	complete	predicted

The *Oryza sativa* protein FASTA sequence is further analysed through consurf showing conserve and variable residues (**Fig.2**). The conserve stretches are FAWCTWDAFY, PPKFVIIDDG, IKENSKFQDG, DDPAAGIKTV, KYVYVWHAT and GYWGGVRPGV.

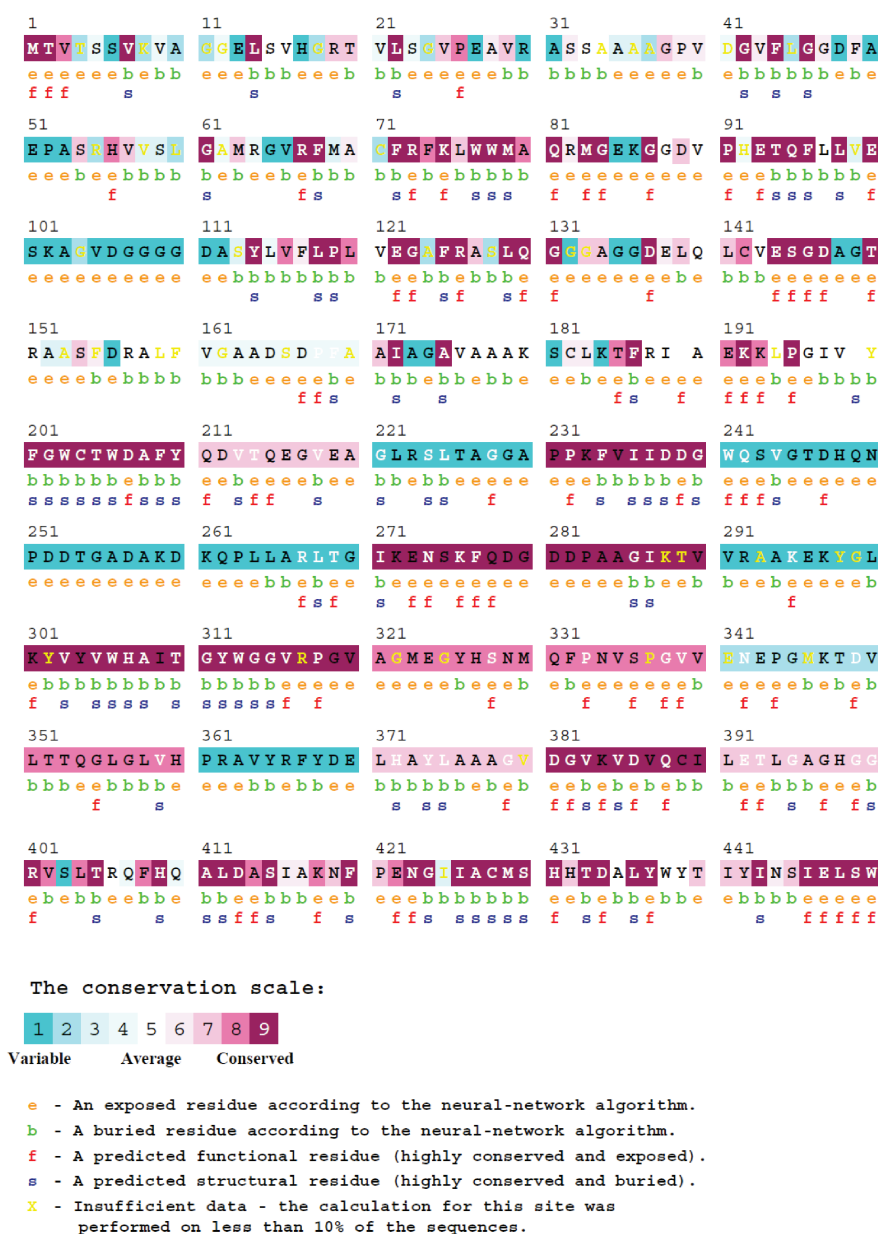


Fig.2. Consurf colour coded MSA result of *Oryza sativa* showing conserved and variable stretches in 3-D protein model of Raffinose Synthase(RafS)

The protein modelling is done through SwissProt (<https://swissmodel.expasy.org/interactive>). Crystal structure of a putative alpha-N-acetylgalactosaminidase (BH1870) from *Bacillus halodurans* is considered as the template. The oligostate of crystal is homo-tetramer. The *Zea mays* RafS sequence is the target sequence (Fig.3).

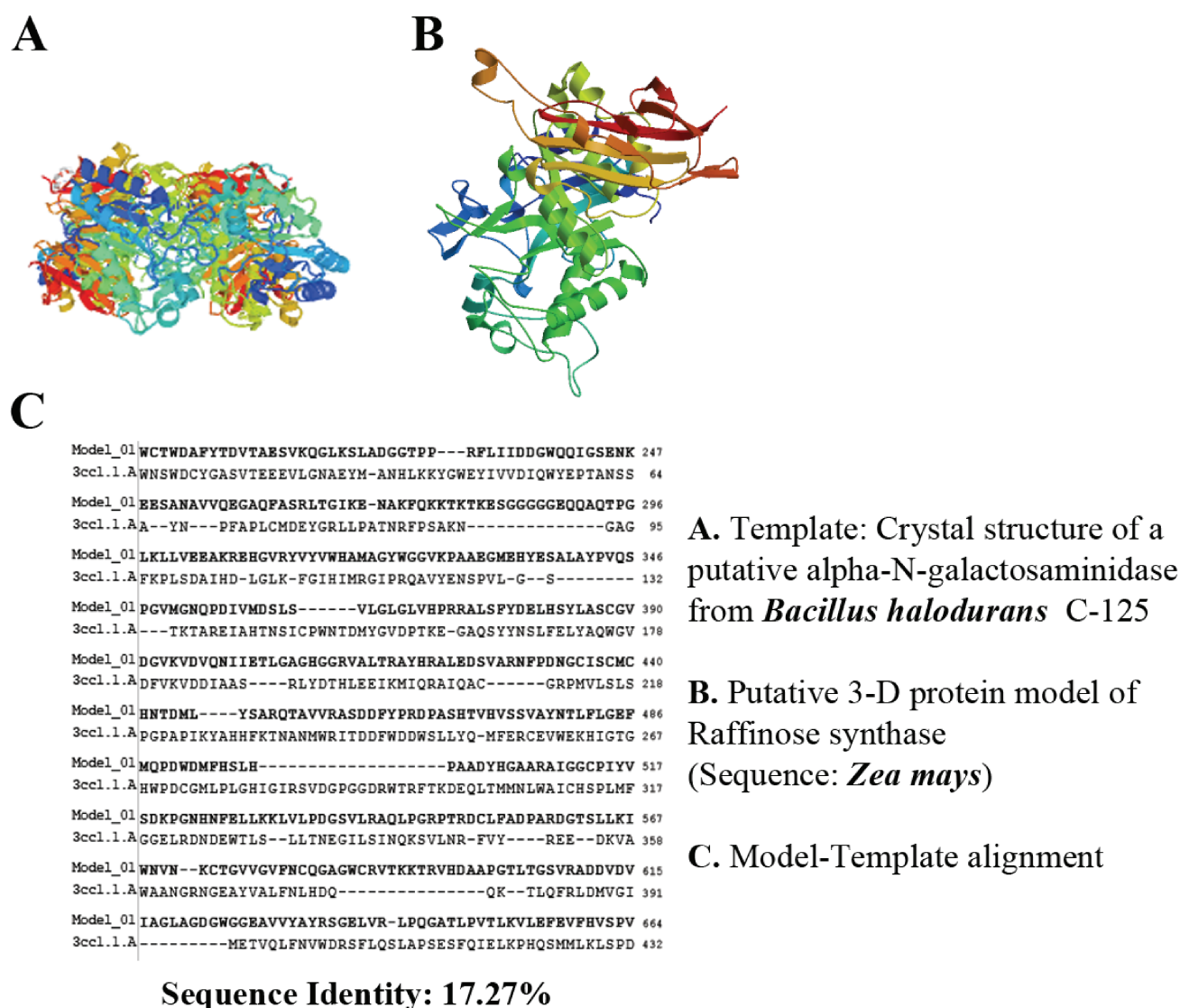


Fig.3. 3-D protein model of Raffinose Synthase (B) alongwith the structure of Template (A) and Model-Template alignment (C).

Phylogenetic Analysis of RafS sequences

The cladogram of all RafS or RafS-like sequences extracted from genomic sequences of plants shows a distinct separation of monocot and dicot sequences (Fig.4). The evolutionary history was inferred using the Neighbor-Joining method. The bootstrap consensus tree inferred from 21 replicates is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 100% bootstrap replicates are collapsed. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree most of the plants exhibit more than one isoforms of RafS. Diversification is high among the monocots and the highly monophyletic clade of monocots suggests a single ancestral sequence.

Table.2 List of organisms used in fig.4.

Sl. No.	Name of the organism	Organism code
1.	<i>Amborella trichopoda</i>	Atr
2.	<i>Brachypodium distachyon</i>	Bdi
3.	<i>Brassica napus</i>	Bna
4.	<i>Brassica oleovar</i>	Bol
5.	<i>Brassica rapa</i>	Bra
6.	<i>Citrus sinensis</i>	Csi
7.	<i>Eucalyptus grandis</i>	Egr
8.	<i>Elaeis guineensis</i>	Egu
9.	<i>Erythranthe guttatus</i>	Egu
10.	<i>Fragaria vesca</i>	Fve
11.	<i>Glycine max</i>	Gma
12.	<i>Musa acuminata</i>	Mac
13.	<i>Malus domestica</i>	Mdo
14.	<i>Nelumbo nucifera</i>	Nnu
15.	<i>Oryza brachyantha</i>	Obr
16.	<i>Oryza sativa</i>	Osa
17.	<i>Pyrus x bretschneideri</i>	Pbr
18.	<i>Phoenix dactylifera</i>	Pda
19.	<i>Sorghum bicolor</i>	Sbi
20.	<i>Setaria italica</i>	Sit
21.	<i>Vitis vinifera</i>	Vvi
22.	<i>Zea mays</i>	Zma

Discussion

From the search of RafS sequences from different domains of life, it becomes apparent that RafS is a plant-specific protein subfamily. These Rafs sequences are reported from Monocot and Dicot plants such as *Oryza sativa Japonica*, *Oryza brachyantha*, *Brachypodium distachyon*, *Musa acuminata*, *Zea mays*, *Setaria italica*, *Amborella trichopoda*, *Phoenix dactylifera*, *Eucalyptus grandis*, *Erythranthe guttatus* etc. The crystal structure of RafS is unavailable. The RafS sequences are showing small stretch of characteristic of conserved and variable residue for identification of common catalytic domain. The 3-D protein model of RafS from *Zea mays* establishes that small changes in amino acid residues may result in change of biological fitness of the protein. The phylogenetic tree of RafS sequences represents distinct separation for dicot and monocot sequences and a strongly monophyletic clade for monocots suggesting a single ancestral sequence. From the present study it can be concluded that more genomic data from lower plants supported by crystal structures of RafS are required to search for the ancestry of the gene, in relevance of its preferential evolution in higher plants.

