FOR THE RECORD

Prediction of a common β-propeller catalytic domain for fructosyltransferases of different origin and substrate specificity

TIRSO PONS, LÁZARO HERNÁNDEZ, FRANK R. BATISTA, AND GLAY CHINEA

Centro de Ingeniería Genética y Biotecnología (CIGB), Havana 10600, Cuba

Abstract: The three-dimensional (3D) structure of fructan biosynthetic enzymes is still unknown. Here, we have explored folding similarities between reported microbial and plant enzymes that catalyze transfructosylation reactions. A sequence-structure compatibility search using TOPITS, SDP, 3D-PSSM, and SAM-T98 programs identified a β-propeller fold with scores above the confidence threshold that indicate a structurally conserved catalytic domain in fructosyltransferases (FTFs) of diverse origin and substrate specificity. The predicted fold appeared related to that of neuraminidase and sialidase, of glycoside hydrolase families 33 and 34, respectively. The most reliable structural model was obtained using the crystal structure of neuraminidase (Protein Data Bank file: 5nn9) as template, and it is consistent with the location of previously identified functional residues of bacterial levansucrases (Batista et al., 1999; Song & Jacques, 1999). The sequence–sequence analysis presented here reinforces the recent inclusion of fungal and plant FTFs into glycoside hydrolase family 32, and suggests a modified sequence pattern {H-x(2)-[PTV]-x(4)-[LIVMA]-[NSCAYG]-[DE]-P-[NDSC]-[GA]} for this family.

Keywords: fold recognition; glycoside hydrolases; levansucrase; sequence analysis

Fructans are commercially used in both food and nonfood applications. In nature, fructan biosynthesis occurs from sucrose by several microbial species and by about 15% of higher plants (Hendry & Wallace, 1993). Bacterial levansucrases (EC 2.4.1.10) are multifunctional enzymes capable of synthesizing high-molecular-mass levans directly from sucrose; however, plant fructans are synthesized by the concerted action of at least two fructosyltransferases (FTFs) exhibiting a distinct fructosyl-donor and fructosyl-acceptor specificities. Sucrose:sucrose 1-fructosyltransferase (1-SST) generally initiates fructan synthesis in plants by catalyzing the transfer of the fructosyl residue from one sucrose to another sucrose molecule, resulting in the formation of the trisaccharide 1-kestose. Then, structurally different fructans are formed by the action of fructan:fructan 1-fructosyltransferase (1-FFT), fructan:fructan 6G-fructosyltransferase (6G-FFT) or sucrose:fructan 6-fructosyltransferase (6-SFT) (for review see Vijn & Smeekens, 1999).

β-Fructofuranosidases are considered to function by a double displacement mechanism with an overall retention of the anomeric configuration of the fructosyl residue. These enzymes are grouped in the glycoside hydrolase family 32 (invertases, levanases, inulinases, sucrose-6-phosphate hydrolases, and fungal and plant FTFs), and glycoside hydrolase family 68 (bacterial FTFs and invertases from *Zymomonas mobilis* and *Bacillus sp.*) (http://afmb.cnrs-mrs.fr/~pedro/CAZY/ghf.html).

The three-dimensional (3D) structures and key residues at active sites of enzymes are generally better conserved than amino acid sequences. Consequently, structural studies combined with sequence comparisons have allowed many glycoside hydrolase families to be grouped according to a common fold and a common catalytic apparatus (Henrissat & Davies, 1997). This classification provides a predictive tool for the catalytic machinery of the glycoside hydrolase enzymes. In this work, we have investigated folding similarities between bacterial, fungal, and plant FTFs using a sequence-structure compatibility search approach.

Results and discussion: We have compared the available amino acid sequences of β-fructofuranosidase enzymes and their phylogenetic relationship is shown in Figure 1. Our sequence comparison results (Fig. 2A,B) support the PFAM alignment accessible at http://afmb.cnrs-mrs.fr/~pedro/CAZY/ghf_32.html. Considering recent entries, including fungal and plant FTFs, into the glycoside hydrolase family 32, the PROSITE pattern (Bairoch et al., 1997) needs to be modified as follows:
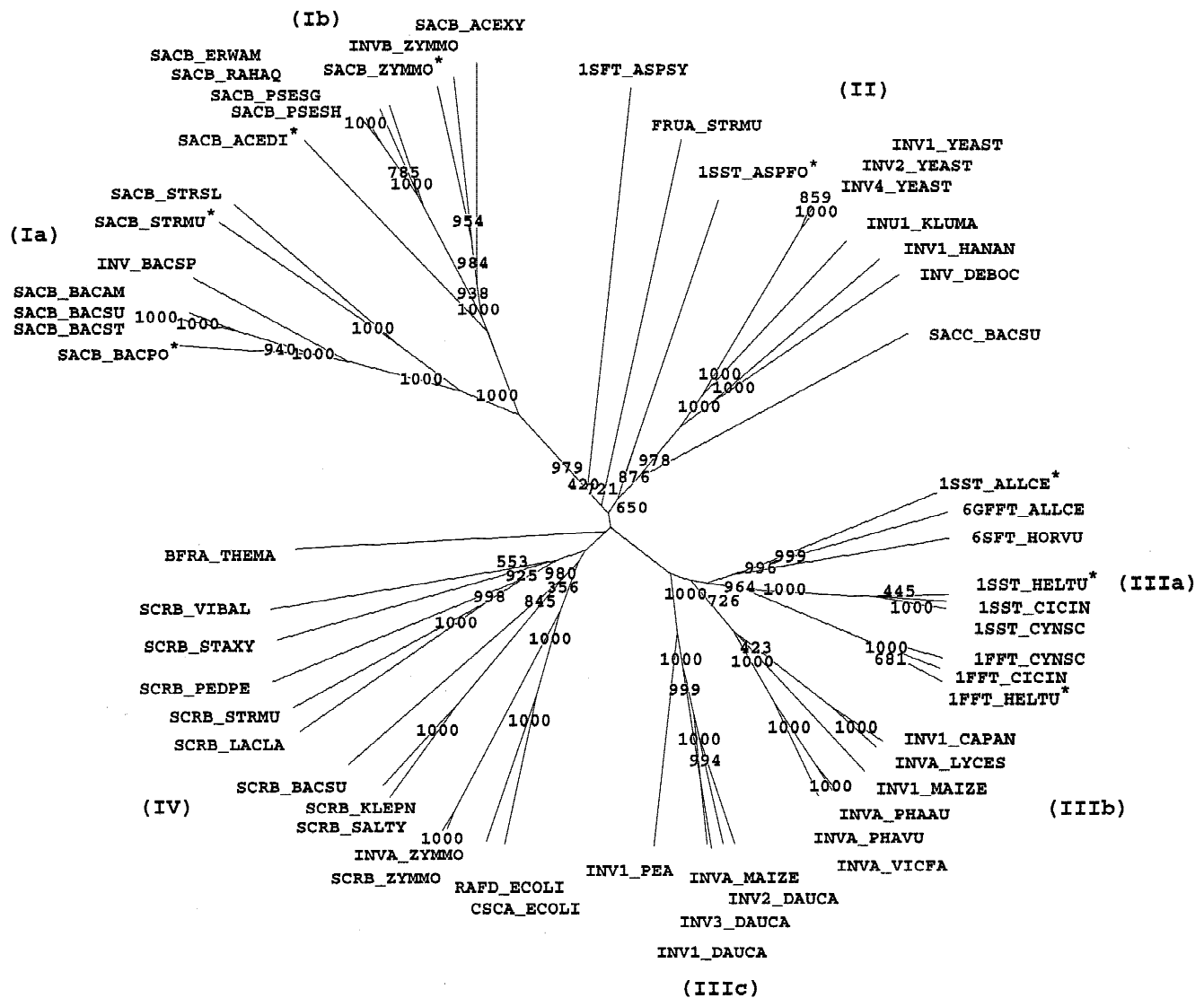

Fig. 1. Phylogenetic tree of β-fructofuranosidases. Sequence identifiers follow the SWISS-PROT conventions (name_specie). Asterisks indicate the eight representative FTF sequences used in the sequence-structure compatibility analysis. The tree displays bootstrap values from 1,000 simulations performed by CLUSTALW. Figure was generated by TREETOOL program (University of Illinois). Group I gathered levansucrases from Gram-positive (Ia) and Gram-negative (Ib) bacteria. Group II includes yeast invertases, fructanases and the so far available fungal FTF sequences (1SST_ASPFO and 1SFT_ASPSY). Plant FTFs (group IIIa) form a cluster with plant invertases (IIIb and IIIc). Group IV is represented by bacterial sucrose or raffinose hydrolases.

PDOC00532: H-x(2)-P-x(4)-[LIVM]-N-D-P-N-G

(original pattern)

H-x(2)-[PTV]-x(4)-[LIVMA]-[NSCAYG]-[DE]-P-[NDSC]-[GA]

(new pattern).

A search in the SWISS-PROT database (84,622 sequences) by MOTIF program using the new pattern, revealed 35 matches, all corresponding to enzymes of family 32. Two of these sequences (FRUA_STRMU and INVA_MAIZE) were not detected using the original pattern (http://www.expasy.ch/cgi-bin/nicesite.pl?PS00609). Other modifications for the proposed pattern could be necessary, as the number of β-fructofuranosidase sequences increase.

We also searched for regions of local similarity among FTF enzymes. The MACAW program found two highly conserved blocks (Fig. 2C) with a probability of obtaining the observed level of similarity by chance (P-value) of 1.6×10^{-05} (search space $N = 2.106 \times 10^{58}$) and 0.0×10^{00} (search space $N = 1.607 \times 10^{60}$) for block1 and block 2, respectively. These low P-values show that the relationship is authentic.

The first conserved block is included in the conserved region called "sucrose box" present in levansucrases and enzymes of the glycoside hydrolase family 32, whereas the second block or "RDP" motif (numbered as IV in Fig. 2B) is highly conserved in all β-fructofuranosidases. The RDP motif of levansucrases from *Acetobacter diazotrophicus* (Batista et al., 1999) and *Streptococcus salivarius* (Song & Jacques, 1999) was found to be involved in
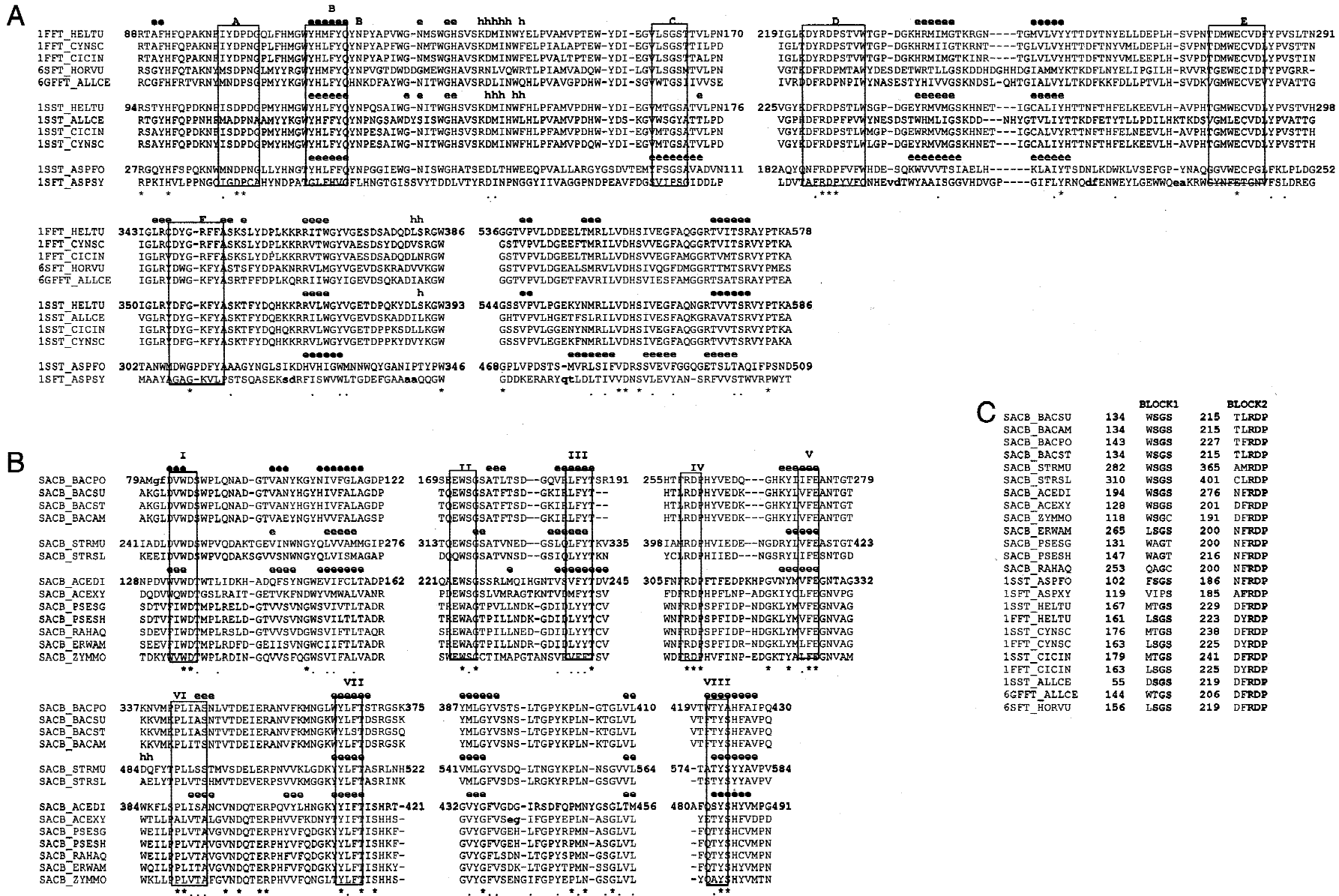
A

B

C

		BLOCK1		BLOCK2
SACB_BACSU	134	WSGS	215	TLRDP
SACB_BACAM	134	WSGS	215	TLRDP
SACB_BACPO	143	WSGS	227	TFRDP
SACB_BACST	134	WSGS	215	TLRDP
SACB_STRMU	282	WSGS	365	AMRDP
SACB_STRSL	310	WSGS	401	CLRDP
SACB_ACEDI	194	WSGS	276	NFRDP
SACB_ACEXY	128	WSGS	201	DFRDP
SACB_ZYMMO	118	WSGC	191	DFRDP
SACB_ERWAM	265	LSGS	200	NFRDP
SACB_PSESG	131	WAGT	200	NFRDP
SACB_PSESH	147	WAGT	216	NFRDP
SACB_RAHAQ	253	QAGC	200	NFRDP
1SST_ASPFO	102	FSGS	186	NFRDP
1SFT_ASPXY	119	VIPS	185	AFRDP
1SST_HELTU	167	MTGS	229	DFRDP
1FFT_HELTU	161	LSGS	223	DYRDP
1SST_CYNSC	176	MTGS	238	DFRDP
1FFT_CYNSC	163	LSGS	225	DYRDP
1SST_CICIN	179	MTGS	241	DFRDP
1FFT_CICIN	163	LSGS	225	DYRDP
1SST_ALLCE	55	DSGS	219	DFRDP
6GFFT_ALLCE	144	WTGS	206	DFRDP
6SFT_HORVU	156	LSGS	219	DFRDP

Fig. 2. Comparisons of the amino acid sequence of microbial and plant FTFs. Conserved regions in the sequence alignments of (**A**) fungal and plant FTFs and (**B**) bacterial FTFs are boxed. At the bottom, asterisks indicate identity and dots indicate conservative changes (STA, LVIM, KR, DE, QN, FYW). Numbering refers to the precursor proteins and lowercase letters in bold represent insertions. **C:** Conserved regions (block1 and block2) in all FTF enzymes, obtained from multiple alignments using the MACAW and BLOCKS programs. The numbers give the starting position of blocks in the mature proteins. The sequences are named as in Table 1, and the conserved residues are in bold.
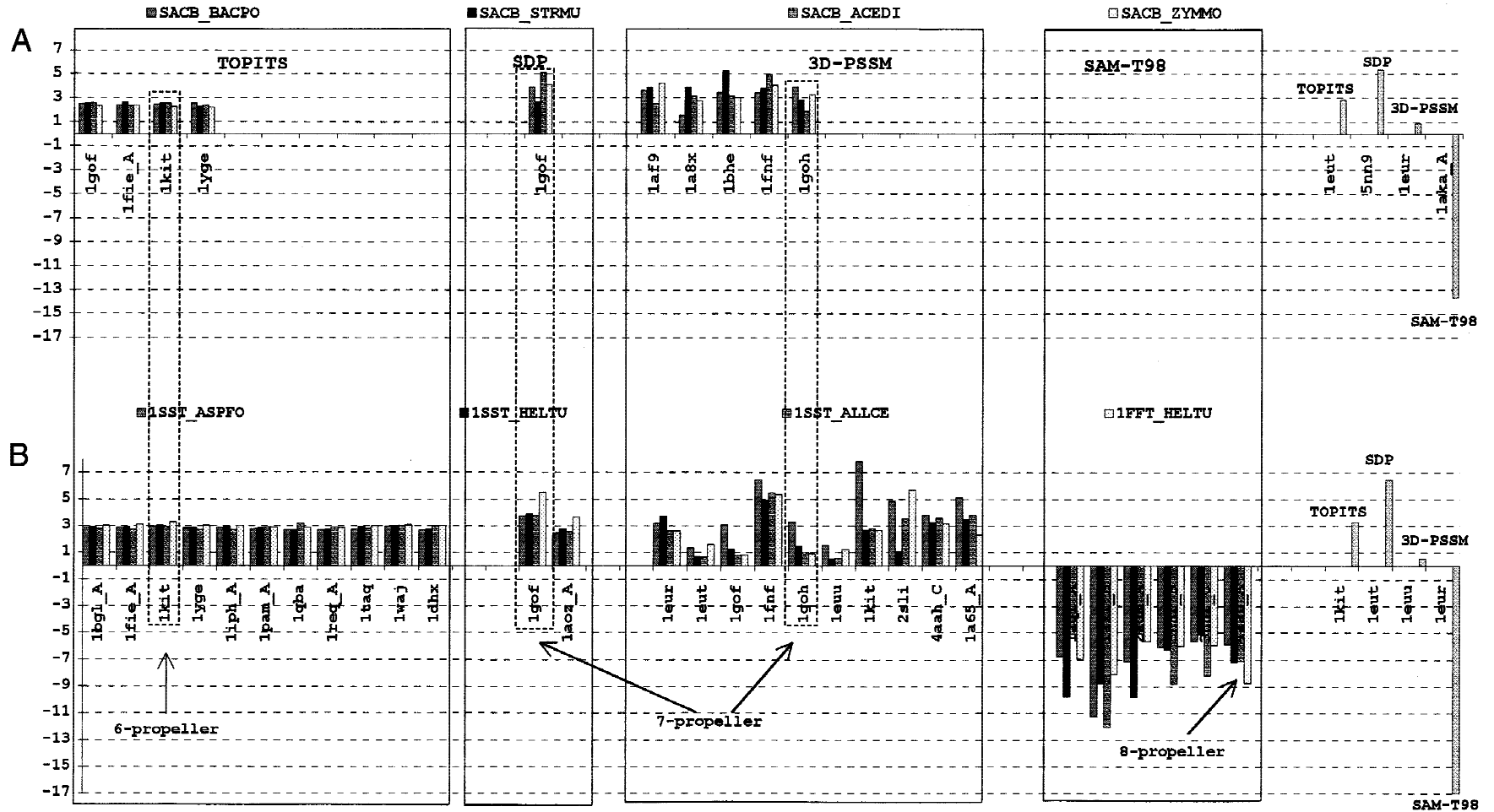
Fig. 3. Graphic representation of sequence-structure compatibility scores obtained by TOPITS, SDP, 3D-PSSM, and SAM-T98 programs. Panels show the sequence-structure compatibility scores for the selected sequence (most divergent) of (**A**) bacterial FTFs and (**B**) fungal and plant FTFs. Boxes represent a comparison of scores corresponding to the remote homologous structures identified by each of the four theoretical methods for bacterial, fungal and plant fructosyltransferases. The four columns to the right (without boxes) represent the better scores obtained by each sequence-structure compatibility search method. The β-propeller fold is represented by sialidases (2sil, 1kit, 1eut, 1eur, 1euu), neuraminidase (5nn9), galactose oxidase (1gof, 1goh), and methanol dehydrogenase (4aah). Only the remote homologous structures (PDB file) identified simultaneously with each of the selected sequences are represented. Dashed lines represent the β-propeller fold identified by TOPITS, SDP and 3D-PSSM programs using bacterial, fungal and plant FTF sequences. SAM-T98 identified an 8-propeller fold using the fungal and plant FTF sequences. The second best score (-13.23) obtained by SAM-T98 using the sequence of *A. diazotrophicus* levansucrase corresponds to the galactose oxidase structure.

binding and/or split of sucrose. The BLOCKS program (http://www.public.iastate.edu/~pedro/blocks_query.html) corroborated that only block1 and block2 are conserved in more than 95% of the FTF enzymes. Sequence similarity between bacterial levansucrases, and fungal and plant FTFs is well below the "twilight zone" (as low as 7%). Because PSI-BLAST—a sensitive sequence similarity search program—failed to detect a reliable relationship (E-values better than 0.001) between FTFs and those proteins with determined 3D structure, we used as approach a sequence-structure compatibility search that combined the TOPITS, SDP, 3D-PSSM, and SAM-T98 programs.

For the study, we selected the most divergent sequences within the FTF enzymes (asterisks in Fig. 1) to cover a wider range of sequence diversity into the FTF family, and because the accuracy of the native fold selection in sequence-structure compatibility analyses can be drastically improved by using a few homologs with low sequence similarity (Reva et al., 1999). The most conserved region (comprising more than 76% of the total residues, except for *Streptococcus mutans* FTF which is 54.6%) of the eight selected sequences, predicted as β-domain by PHD program, was analyzed by using our sequence-structure compatibility approach.

The sequence-structure compatibility search identified a β-propeller fold with scores above the confidence threshold to indicate a structural homology for the catalytic domain of FTFs (details are given in the caption to Fig. 3) and predicted that FTFs are related to the known 3D structures of neuraminidase and sialidase (glycoside hydrolases families 33 and 34). The compatibility scores produced by TOPITS, SDP, and 3D-PSSM programs using the bacterial FTFs are lower than those of fungal and plant enzymes,

and in some cases are below the confidence threshold. Considering that the top ranks produced by TOPITS, SDP, and 3D-PSSM programs included several β-propeller folds for bacterial, fungal, and plant FTFs, we concluded that our prediction is reliable.

The catalytic residues Asp23 and Glu204 of the *Saccharomyces cerevisiae* invertase (Reddy & Maley, 1990, 1996) are highly conserved in fungal and plant FTFs (motifs A and E in Fig. 2B). Based on the high sequence similarity and the sequence-structure compatibility results obtained here (Fig. 3B), we propose to extend the β-propeller structural model for glycoside hydrolase family 32 (Pons et al., 1998; http://www.cnb.uam.es/~cnbprot/Glico/fam32.html) to fungal and plant FTFs. This proposal is in accordance with the assumption that plant FTFs evolved from invertases (Vijn & Smeekens, 1999) and reinforce the recent inclusion of fungal and plant FTFs into glycoside hydrolase family 32.

It is well known in fold recognition that identifying the correct fold in a set of structures is a much easier task than providing the correct alignment between the probe sequence and the target protein structure. Although it is not possible to obtain atomic details from sequence-structure compatibility models, the information derived from the conserved active sites of sialidase and neuraminidase was good enough to allow location of known functional residues of distinct bacterial levansucrases. Additionally, the sequence-structure alignment in Figure 4B showed a high correspondence between the secondary structure elements, which are characteristics of the pseudo sixfold symmetry. The equivalent residues Asp309 and Asp397 in the RDP motif of levansucrases from *A. diazotrophicus and S. salivarius*, respectively, were found to be implicated in

Table 1. *β-Fructosyltransferase enzymes used in this study*[a]

Enzyme	Identifier	GH family	EC	SPTREMBL accession no.	SWISS-PROT accession no.	Source
Fructan:fructan 6G-FTF	6GFFT_ALLCE	32	—	P92916	—	*Allium cepa*
Sucrose:fructan 6-FTF	6SFT_HORVU	32	—	Q96466	—	*Hordeum vulgare*
Sucrose:sucrose 1-FTF	1SST_ALLCE	32	—	O81082	—	*Allium cepa*
Sucrose:sucrose 1-FTF	1SST_HELTU	32	2.4.1.99	O81986	—	*Helianthus tuberosus*
Sucrose:sucrose 1-FTF	1SST_CICIN	32	2.4.1.99	O24459	—	*Cichorium intybus*
Sucrose:sucrose 1-FTF	1SST_CYNSC	32	—	O23786	—	*Cynara scolymus*
Sucrose:sucrose 1-FTF	1SST_ASPFO	32	3.2.1.26	O42801	—	*Aspergillus foetidus*
Fructosyltransferase	1SFT_ASPSY	32	—	[b]	—	*Aspergillus sydowii*
Fructan:fructan 1-FTF	1FFT_CYNSC	32	—	O65778	—	*Cynara scolymus*
Fructan:fructan 1-FTF	1FFT_CICIN	32	2.4.1.100	Q9ZR96	—	*Cichorium intybus*
Fructan:fructan 1-FTF	1FFT_HELTU	32	2.4.1.100	O81985	—	*Helianthus tuberosus*
Levansucrase	SACB_BACSU	68	2.4.1.10	—	P05655	*Bacillus subtilis*
Levansucrase	SACB_BACST	68	2.4.1.10	—	P94468	*Bacillus stearothermophilus*
Levansucrase	SACB_BACPO	68	—	Q9Z5E5	—	*Paenibacillus polymyxa*
Levansucrase	SACB_BACAM	68	2.4.1.10	—	P21130	*Bacillus amyloliquefaciens*
Fructosyltransferase	SACB_STRMU	68	2.4.1.10	—	P11701	*Streptococcus mutans*
Levansucrase	SACB_STRSL	68	2.4.1.10	—	Q55242	*Streptococcus salivarius*
Levansucrase	SACB_ACEDI	68	2.4.1.10	—	Q43998	*Acetobacter diazotrophicus*
Levansucrase	SACB_ACEXY	68	—	[c]	—	*Acetobacter xylinus*
Levansucrase	SACB_PSESG	68	2.4.1.10	—	O52408	*Pseudomonas syringae pv glycinea*
Levansucrase	SACB_PSESH	68	2.4.1.10	—	O68609	*Pseudomonas syringae pv phaseolicola*
Levansucrase	SACB_RAHAQ	68	2.4.1.10	—	O54435	*Rhanella aquatilis*
Levansucrase	SACB_ERWAM	68	2.4.1.10	—	Q46654	*Erwinia amylovora*
Levansucrase	SACB_ZYMMO	68	2.4.1.10	—	Q60114	*Zymomonas mobilis*

[a]EC, enzyme classification number according to the International Union of Biochemistry and Molecular Biology (IUBMB) recommendations; GH, glycoside hydrolase; accession numbers in the EMBL database; [b]AJ289046 and [c]AB034152.
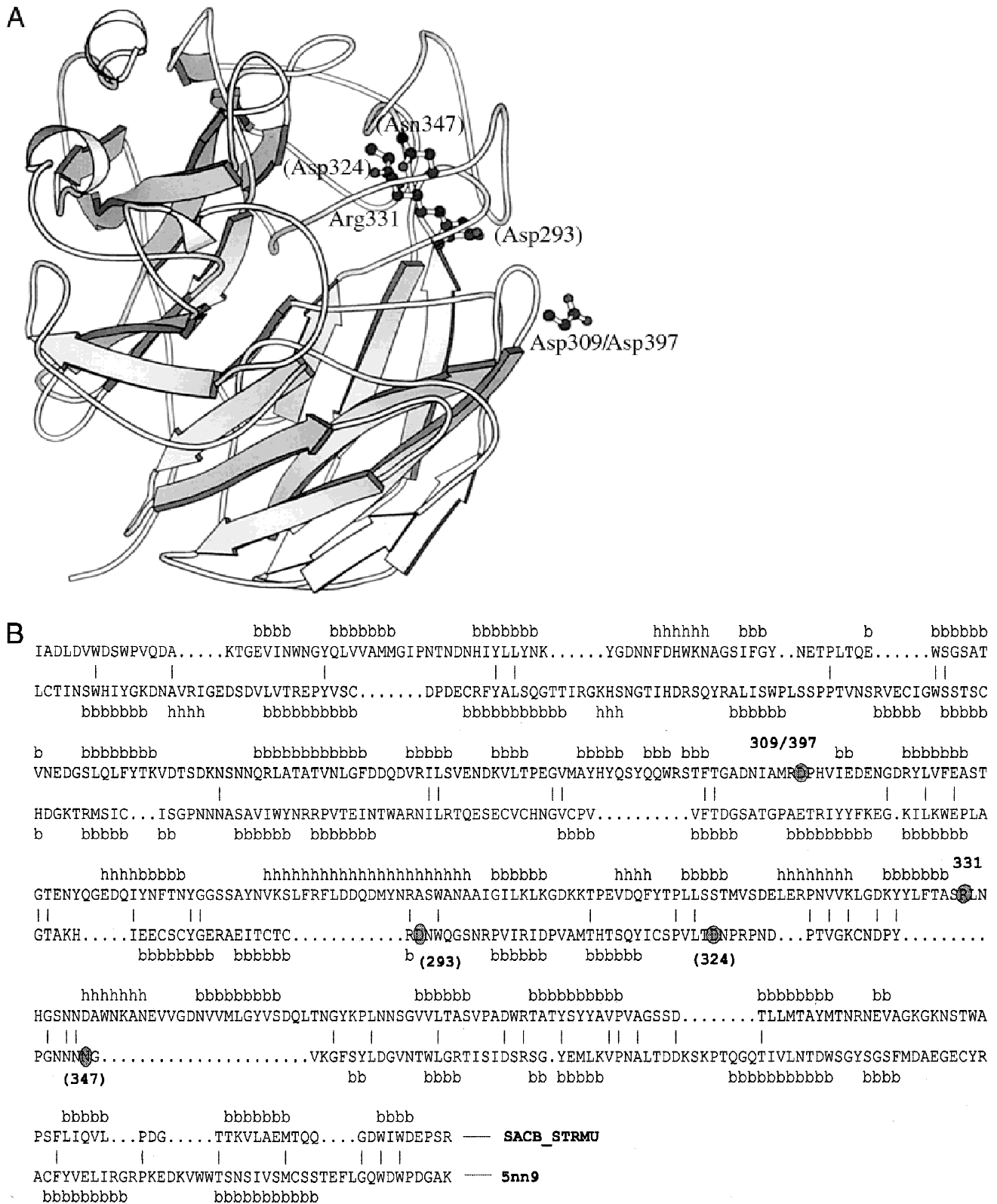
Fig. 4. Putative location of known functional residues of bacterial levansucrases in the β-propeller catalytic domain of the crystal structure of influenza virus neuraminidase. Top view of the active site of neuraminidase N9 crystal structure (PDB file: 5nn9). **A:** Ball-and-stick models represent known functional residues (Arg331, Asp309, and Asp397) of distinct levansucrases. The catalytic residues (Asp293, Asp324, and Asn347) from neuraminidase N9 are shown in parenthesis. The figure was generated by MOLSCRIPT program (Kraulis, 1991). **B:** The alignment produced by SDP between the FTF sequence from *S. mutans* and neuraminidase N9. Above the FTF sequence is the secondary structure predicted by PHD. Below the neuraminidase sequence is the known secondary structure of neuraminidase as determined by DSSP program (Kabsch & Sander, 1983). Helices are indicated by h, and β-strands by b.

binding or splitting of sucrose (Batista et al., 1999; Song & Jacques, 1999). The RDP motif is close in space to Arg331 (in the vicinity of the conserved region VII in Fig. 2A), involved in the polymerase activity of *Bacillus subtilis* levansucrase (Chambert & Petit-Glatron, 1991). All these residues are exposed to solvent delimiting the active site cavity in the proposed fold (see Fig. 4A). It is well recognized that the active site cavity of enzymes with similar fold (such as sialidases, neuraminidases, methanol dehydrogenase, and galactose oxidase, with a β-propeller fold) is located in the same topological region.

Materials and methods: Protein sequences were retrieved from the current sequence databases using the SRS WWW service (Etzold et al., 1996). FTFs used in this work are summarized in Table 1. Comparison of the FTF proteins was generated using CLUSTALW (Thompson et al., 1994), MAXHOM (Sander & Schneider, 1991), and the secondary structural information predicted with the PHD program (Rost & Sander, 1994). The evolutionary tree was calculated using CLUSTALW. The sequence-structure compatibility search approach used in this work combined the TOPITS (Rost et al., 1997), SDP (Fischer & Eisenberg, 1996), 3D-PSSM (Kelley et al., 1999), and SAM-T98 (Karplus et al., 1998) programs. TOPITS, SDP, 3D-PSSM, SAM-T98, and PSI-BLAST (Altschul et al., 1997) are accessible via Internet using the URLs: http://dodo.cpmc.columbia. edu/pp/submit_adv.html, http://www.doe-mbi.ucla.edu/ people/ frsvr/frsvr.html, http://www.bmm.icnet.uk/~3dpssm, http:// www.cse.ucsc.edu/research/compbio/HMM-library-search.html and http://www.ncbi.nlm.nih.gov/blast, respectively. We used the MOTIF program (Cockwell & Giles, 1989) to search the SwissProt database with the new sequence pattern. The MACAW program (Schuler et al., 1991) was used to estimate the probabilities of the independent appearance of the regions of local similarity into the FTF family. All programs were used with default parameters.

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acid Res 25*:3389–3402.

Bairoch A, Bucher P, Hofmann K. 1997. The PROSITE database, its status in 1997. *Nucleic Acids Res 25*:217–221.

Batista FR, Hernández L, Fernández JR, Arrieta J, Menéndez C, Gómez R, Támbara Y, Pons T. 1999. Substitution of Asp-309 by Asn in the Arg-Asp-Pro (RDP) motif of *Acetobacter diazotrophicus* levansucrase affects sucrose hydrolysis, but not enzyme specificity. *Biochem J 337*:503–506.

Chambert R, Petit-Glatron MF. 1991. Polymerase and hydrolase activities of *Bacillus subtilis* levansucrase can be separately modulated by site-directed mutagenesis. *Biochem J 279*:35–41.

Cockwell KY, Giles IG. 1989. Software tools for motif and pattern scanning: Program descriptions including a universal sequence reading algorithm. *Comput Appl Biosci 5*:227–232.

Etzold T, Ulyanov A, Argos P. 1996. SRS: Information retrieval system for molecular biology data banks. *Methods Enzymol 266*:114–128.

Fischer D, Eisenberg D. 1996. Protein fold recognition using sequence-derived predictions. *Protein Sci 5*:947–955.

Hendry GAF, Wallace RK. 1993. The origin, distribution, and evolutionary significance of fructans. In: Suzuki M, Chatterton NJ, eds. *Science and technology of fructans*. Boca Raton, Florida: CRC Press. pp 119–139.

Henrissat B, Davies G. 1997. Structural and sequence-based classification of glycoside hydrolases. *Curr Opin Struct Biol 7*:637–644.

Kabsch W, Sander C. 1983. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers 22*:2577–2637.

Karplus K, Barret C, Hughey R. 1998. Hidden Markov models for detecting remote protein homologies. *Bioinformatics 14*:846–856.

Kelley LA, MacCallum RM, Sternberg MJE. 1999. Recognition of remote protein homologies using three-dimensional information to generate a position specific scoring matrix in the program 3D-PSSM. In: Istrail S, Pevzner P, Waterman M, eds. *RECOMB-99 proceeding of the 3ʳᵈ annual conference on computational biology*. New York: Association for Computing Machinery. pp 218–225.

Kraulis P. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr 24*:946–950.

Pons T, Olmea O, Chinea G, Beldarraín A, Márquez G, Acosta N, Rodríguez L, Valencia A. 1998. Structural model for family 32 of glycosyl-hydrolase enzymes. *Proteins 33*:383–395.

Reddy VA, Maley F. 1990. Identification of an active-site residue in yeast invertase by affinity labeling and site-directed mutagenesis. *J Biol Chem 265*:10817–10820.

Reddy VA, Maley F. 1996. Studies on identifying the catalytic role of Glu-204 in the active site of yeast invertase. *J Biol Chem 271*:13953–13958.

Reva BA, Skolnick J, Finkelstein AV. 1999. Averaging interaction energies over homologs improves protein fold recognition in gapless threading. *Proteins 35*:353–359.

Rost B, Sander C. 1994. Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins 19*:55–72.

Rost B, Schneider R, Sander C. 1997. Protein fold recognition by prediction-based threading. *J Mol Biol 270*:471–480.

Sander C, Schneider R. 1991. Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins 9*:52–68.

Schuler GD, Altschul SF, Lipman DJ. 1991. A workbench for multiple alignment construction and analysis. *Proteins 9*:180–190.

Song DD, Jacques NA. 1999. Mutation of aspartic acid residues in the fructosyltransferase of *Streptococcus salivarius* ATCC 25975. *Biochem J 344*:259–264.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTALW: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res 22*:4673–4680.

Vijn I, Smeekens S. 1999. Fructan: More than a reserve carbohydrate?. *Plant Physiol 120*:351–359.