5-2015

The Crystal Structure of Nitrosomonas Europaea Sucrose Synthase Reveals Critical Conformational Changes and Insights into the Sucrose Metabolism in Prokaryotes

Rui Wu

Matías D. Asención Diez

Carlos M. Figueroa
Loyola University Chicago

Matías Machtey

Alberto A. Iglesias

See next page for additional authors

Authors
Rui Wu, Matías D. Asención Diez, Carlos M. Figueroa, Matías Machtey, Alberto A. Iglesias, Miguel Ballicora, and Dali Liu

The crystal structure of *Nitrosomonas europaea* sucrose synthase reveals critical conformational changes and insights into the sucrose metabolism in prokaryotes

Rui Wu,[a] Matías D. Asención Diez,[a,b] Carlos M. Figueroa,[a,b] Matías Machtey,[b] Alberto A. Iglesias,[b] Miguel A. Ballicora,[a] and Dali Liu[a]#

Department of Chemistry and Biochemistry, Loyola University Chicago, Chicago, Illinois, USA[a]; Instituto de Agrobiotecnología del Litoral, Universidad Nacional del Litoral and Consejo Nacional de Investigaciones Científicas y Técnicas, Centro Científico Tecnológico Santa Fe, Santa Fe, Argentina[b]

Running Head: Crystal structure of a prokaryotic sucrose synthase

#Address correspondence to Dali Liu, dliu@luc.edu

R.W., M.D.A.D. and C.M.F. contributed equally to this work

ABSTRACT

In this paper we report the first crystal structure of a prokaryotic sucrose synthase from the non-photosynthetic bacterium *Nitrosomonas europaea*. The obtained structure was in an open form, whereas the only other available structure from the plant *Arabidopsis thaliana* was in a closed conformation. Comparative structural analysis revealed a "hinge-latch" combination, which is critical to transition between the open and closed forms of the enzyme. The *N. europaea* sucrose synthase shares the same fold as the GT-B family of the retaining glycosyltransferases. In addition, a triad of conserved homologous catalytic residues in the family showed to be functionally critical in the *N. europaea* sucrose synthase (Arg567, Lys572, Glu663). This implies that sucrose synthase shares not only a common origin with the GT-B family, but also a similar catalytic mechanism. The enzyme preferred transferring glucose from ADP-glucose rather than UDP-glucose like the eukaryotic counterparts. This predicts that these prokaryotic organisms have a different sucrose metabolic scenario from plants. Nucleotide preference determines where the glucose moiety is targeted after sucrose is degraded.

IMPORTANCE

We obtained biochemical and structural evidence of sucrose metabolism in non-photosynthetic bacteria. Until now, only sucrose synthases from photosynthetic organisms have been characterized. Here, we provide the crystal structure of the sucrose synthase from the chemo-litho-autotroph *N. europaea*. The structure supported that the enzyme functions with an open/close induced fit mechanism. The enzyme prefers as substrate adenine-based nucleotides rather than uridine-based like the eukaryotic counterparts, implying a strong connection between sucrose and glycogen metabolism in

these bacteria. Mutagenesis data showed that the catalytic mechanism must be conserved not only in sucrose synthases, but also in all other retaining GT-B glycosyltransferases.

INTRODUCTION

In plants, sucrose is a major photosynthetic product and plays a key role not only for carbon partition but also in sugar sensing, development, and regulation of gene expression (1-3). It was first thought that sucrose metabolism was a characteristic of plants but it was later found in other oxygenic photosynthetic organisms (4, 5). In the last decade, Salerno and coworkers demonstrated the importance of sucrose for carbon and nitrogen fixation in filamentous cyanobacteria (6, 7). More recently, genomic and phylogenetic analyses revealed the existence of sucrose-related genes in non-photosynthetic prokaryotes such as proteobacteria, firmicutes, and planctomycetes (4, 5, 8). It has been suggested that these organisms acquired the genes of sucrose metabolism by horizontal gene transfer (4, 5, 8). However, analysis of the enzymes encoded by such genes is currently lacking.

Nitrosomonas europaea is a chemo-litho-autotrophic bacterium that obtains energy by oxidizing ammonia to hydroxylamine and nitrite in presence of oxygen (9). It is a member of the β-proteobacteria group with a putative photosynthetic ancestor (10). *N. europaea* has potential for many biotechnological applications, including bioremediation of water contaminated with chlorinated aliphatic hydrocarbons (11) or ammonia, in combination with *Paracoccus denitrifi* (9). *N. europaea* displays some metabolic resemblance to photosynthetic organisms, but with marked differences. For instance, it possesses all the coding genes for enzymes of the Calvin-Benson cycle, but with two exceptions that could be replaced by other glycolytic enzymes (12). All the genes coding for enzymes from the tricarboxylic acid cycle were found in *N. europaea* (12); however, activity of α-ketoglutarate dehydrogenase is non-detectable (13).

Journal of Bacteriology

The evidence from genomic studies suggests that *N. europaea* can synthesize sucrose (12); however, the biochemical properties of enzymes from sucrose metabolism have not been characterized. Generally, in plants, sucrose is synthesized from UDP-glucose (UDP-Glc) and fructose-6-phosphate (Fru-6P) in a reaction catalyzed by sucrose-6-phosphate synthase (EC 2.4.1.14), followed by removal of the phosphate group by sucrose-6-phosphatase (EC 3.1.3.24). The disaccharide can be degraded to Glc and Fru by invertases (EC 3.2.1.26) or cleaved by UDP to form UDP-Glc and Fru by sucrose synthase (NDP-glucose:D-fructose 2-α-D-glucosyltransferase, EC 2.4.1.13, also abbreviated as SUS or SuSy) (2, 3). However, some plant sucrose synthases have a certain degree of substrate promiscuity (14-21) while the one from *Thermosynechococcus elongatus* prefers ADP (16). For that reason, a general reversible reaction could be written as:

$$\text{NDP} + \text{sucrose} \rightleftharpoons \text{NDP-Glc} + \text{Fru}$$

Besides its physiological role, sucrose synthase catalyzes a reversible reaction and its activity can be measured in both directions *in vitro*. In filamentous cyanobacteria, the products derived from sucrose cleavage contribute to other biological processes, such as polysaccharides synthesis (22). Therefore, understanding the catalysis and the regulation of sucrose synthase is of great significance. Recently, Zheng et al. (23) reported the crystal structure of the *Arabidopsis thaliana* sucrose synthase in complex with UDP and fructose in a closed conformation. This enzyme is a homotetramer composed of four identical subunits of ~90 kDa and belongs to group 4 of the GT-B retaining glycosyltransferase family (http://www.cazy.org/GlycosylTransferases.html) (24). A S_Ni-like reaction mechanism has been proposed for this enzyme family (23-25).

Although several cyanobacterial (8, 16, 19) and plant (14, 17, 26-29) sucrose synthases have been characterized, the enzyme from non-photosynthetic bacteria has never been studied and no structural information of any sucrose synthase from bacterial sources is available. In this work we report the recombinant expression and biochemical characterization of *N. europaea* sucrose synthase and its crystal structure. We also determined the catalytic implications of highly conserved residues and the specificity for nucleotide substrates.

MATERIALS AND METHODS

Materials

Chemicals and coupled enzymes used for activity assays were from Sigma-Aldrich (St. Louis, MO). *Escherichia coli* BL21 (DE3) cells were purchased from New England BioLabs (Ipswich, MA). Bacterial growth media and antibiotics were from Fisher Scientific (Pittsburgh, PA) and Sigma-Aldrich. Crystallization screen solutions and other supplies were purchased from Hampton Research (Aliso Viejo, CA) and Emerald Bio (Bedford, MA). All the other chemicals were of the highest quality available.

Cloning

The sequence coding for the sucrose synthase from *N. europaea* (gene ss2, accession: CAD85125.1) was amplified by PCR using genomic DNA from *N. europaea* ATCC 19718 as template, the specific oligonucleotides CATATGACCACGATTGACACACTCGCCACCTGTACCC (forward, *Nde*I site underlined) and GTCGACTCATATCTCATGGGCCAGCCTGTTTGCCAGCGGCC (reverse, *Sal*I site underlined) as primers, and Phusion HF DNA polymerase (Thermo

Fisher Scientific, Rockford, IL) following the manufacturer's instructions. The program used included an initial denaturation of 30 s at 98 °C; 30 cycles of 98 °C for 5 s, 50 °C for 20 s, and 72 °C for 2 min; and a final extension of 72 °C for 5 min. The PCR product was purified after agarose gel electrophoresis and inserted into the pSC-B vector using the StrataClone Blunt PCR cloning kit (Agilent Technologies, Santa Clara, CA). Sequence identity was checked by automated DNA sequencing at CRC (Comprehensive Cancer Center at University of Chicago, IL). Afterwards, the sequence was subcloned into the pET28c vector (Merck KGaA, Darmstadt, Germany) between *Nde*I and *Sal*I sites to obtain pNESS2, which is the plasmid that encodes the recombinant *N. europaea* sucrose synthase with an N-terminal His_6-tag.

Site-directed mutagenesis

Site-directed mutagenesis was performed by PCR overlap extension as previously described using Phusion DNA polymerase (30, 31). The plasmid encoding the *N. europaea* sucrose synthase (pNESS2) was used as a template for mutagenesis.

To introduce mutations in pNESS2 we used the following primers: TTTACCATGGCGgcgCTGGATCGGATC (forward) and GATCCGATCCAGcgcCGCCATGGTAAA (reverse) for mutant R567A; CTGGATCGGATCgcgAACATTACCGGC (forward) and GCCGGTAATGTTcgcGATCCGATCCAG (reverse) for mutant K572A; and CCAGCCCTGTTCgcgGCATTCGGCCTG (forward) and CAGGCCGAATGCcgcGAACAGGGCTGG (reverse) for mutant E663A. PCR conditions were the same as those described above. Flanking primers for the PCR overlap

Journal of Bacteriology

extension were the same used for cloning (described above). All mutations were confirmed by DNA sequencing.

Protein expression and purification

Transformed *E. coli* BL21 (DE3) cells with pNESS2 were grown in 4 x 1 L of LB supplemented with 100 µg/ml carbenicillin. This was performed in a 2.8 L Fernbach flask at 37 °C and 250 rpm until $OD_{600\ nm}$ reached ~0.6. Protein expression was induced by the addition of 0.5 mM isopropyl-β-D-1-thiogalactopyranoside. Cells were incubated at 25 °C and harvested after 16 h by centrifuging at 5000 x *g* and 4 °C for 15 min. The cell paste was resuspended in Buffer C [20 mM Tris-HCl pH 8.0, 200 mM NaCl, 5% (v/v) glycerol, 10 mM imidazole] and disrupted by sonication. The resulting suspension was centrifuged twice at 30000 x *g* and 4 °C for 15 min and the soluble fraction (crude extract) was loaded onto a 5 ml HisTrap column (GE Life Sciences, Piscataway, NJ) containing Ni^{2+} and previously equilibrated with Buffer C. Elusion of the retained proteins was achieved with a linear imidazole gradient (20 column volumes, 10-300 mM). Fractions containing sucrose synthase activity were pooled, concentrated to 2 ml, and loaded onto a 16/60 Superdex 200 column (GE Life Sciences) previously equilibrated with 50 mM HEPES-NaOH pH 8.0 and 300 mM NaCl. Fractions containing enzyme activity were pooled, concentrated, supplemented with 5% (v/v) glycerol, and stored at -80 °C until use. Under these conditions the enzyme remained stable and fully active for at least 3 months.

Protein assay and detection

Protein concentration was determined by measuring the protein absorbance at 280 nm using a NanoDrop 1000 (Thermo Fisher Scientific) and an extinction coefficient of

1.153 ml mg^{-1} cm^{-1}, determined from the amino acid sequence using the ProtParam server (http://web.expasy.org/protparam/). Denaturing protein electrophoresis was performed as described by Laemmli (32).

Enzyme assays

Activity assays were performed as previously described (16), with minor modifications. In the direction of sucrose synthesis, the reaction medium contained 50 mM HEPPS pH 8.0, 10 mM $MgCl_2$, 5 mM UDP-Glc, 500 mM Fru, 0.3 mM phosphoenolpyruvate, 0.3 mM NADH, 1 U pyruvate kinase, 1 U lactate dehydrogenase, 0.2 mg ml^{-1} BSA, and enzyme in an appropriate dilution in a final volume of 50 µl. Alternatively, activity was measured with 1 mM ADP-Glc and 20 mM Fru. NADH oxidation was followed by measuring the absorbance at 340 nm in a Multiskan Ascent microplate reader (Thermo Fisher Scientific) at 37 °C. One unit of enzyme activity (U) is defined as the amount of protein necessary to produce 1 µmol of product in 1 min under the specified conditions.

Kinetic characterization

Since the saturation kinetics of the enzyme were slightly sigmoidal, data of initial velocity (v) versus substrate concentration (S) were plotted and fitted to a modified Hill equation: $v = V_{max}\ S^{nH} / (S_{0.5}{}^{nH} + S^{nH})$, where $S_{0.5}$ is the concentration of substrate necessary to obtain 50% of the maximal velocity (V_{max}) and n_H is the Hill coefficient. Fitting was performed by a non-linear least-squares algorithm provided by the software Origin 7.0 (OriginLab Corporation). Kinetic parameters were obtained using the averages of two independent datasets that were reproducible within errors of ± 10%.

Phylogenetic analysis

We searched for protein sequences using the term "sucrose synthase" and applied the RefSeq filter in the National Center for Biotechnology Information (NCBI) database. Afterwards, we manually curated them to discard some which were clearly wrongly annotated since they had higher identity to other glycosyltransferases. Sequences were analyzed with the program BioEdit 7.0.5.3 (33) and aligned using the ClustalW server (http://www.genome.jp/tools/clustalw/). Tree reconstruction was performed using the Neighbor-Joining algorithm with a bootstrap of 1000 in the program SeaView 4.4.0 (34). The tree figure was prepared using the FigTree 1.4.0 software (http://tree.bio.ed.ac.uk/software/figtree/).

Crystallization and data collection

After the initial crystallization screen and optimization, the recombinant protein was crystallized via the hanging drop method. The hanging drops were prepared with 1 µl of 15 mg ml^{-1} sucrose synthase and 1 µl of the reservoir solution, containing 5% Tacsimate pH 5.0, 5% (w/v) PEG 3350, and 0.1 M sodium citrate pH 5.6. The hanging drops were kept at 20 °C for crystallization. Crystals appeared in 3 days and were allowed to continue growing at 20 °C for 4 more days until they reached their maximum sizes. Crystals with good morphology and large sizes were transferred to a cryo-condition, which contained 25% glycerol in addition to the components of the reservoir solution, before being frozen in liquid nitrogen.

X-ray diffraction data sets were collected at the SBC19-ID beamline at the Advanced Photon Source (Argonne National Laboratory, Chicago, IL). The wavelength used in the monochromatic data collection was 1.008 Å. All the collected data sets were indexed and integrated using iMosflm and scaled with Scala in the CCP4 program suite

Journal of Bacteriology

(Collaborative Computational Project Number 4) (35). After investigating all statistic values indicating data quality, especially I/σ<I>, and $CC_{1/2}$ (36), we decided to cut the data resolution at 3.05 Å, where I/σ<I> = 2 while $CC_{1/2}$ = 0.561 indicating good data quality (Table 2).

Phasing, model building, and refinement

Molecular replacement was carried out using the program Phaser (37) from the CCP4 program suite. The starting search model in molecular replacement was modified from the known *A. thaliana* sucrose synthase structural model (PDB ID: 3S29) (23). The molecular replacement using the full-length *A. thaliana* Sucrose synthase as a search model did not yield any solutions using Phaser. Suspecting that and inter-domain movement may have been the problem; we tried to virtually isolate some domains based on homology. Then, when we truncated the GT-B(D) domain (cyan domain in Fig. 1B) and used the rest of the molecule as the search model for molecular replacement in Phaser, a solution was finally obtained. Afterwards, model building was conducted in COOT (38). The GT-B(D) domain was built according to the electron density maps. Rigid body refinement and restrained refinement were conducted in refmac5 (39). In order to remove model bias and achieve the best refinement results possible, simulated annealing refinement and ordered solvent identification were conducted using PHENIX.refine (40). Final model and the structure factor have been deposited in the RCSB Protein Data Bank with the accession code 4RBN.

Homology modeling

A model of the monomeric closed form of the *N. europaea* sucrose synthase (residues 16 to 788) was constructed with the program Modeller 9.11

(http://salilab.org/modeller/) (41). As template we used the atomic coordinates of the *A. thaliana* sucrose synthase (3S27) with the ligands UDP and fructose (23). Before the modeling process, sequence alignment was performed manually to match functionally conserved residues and secondary structures. An identity of 50.3% ensured a high confidence alignment since we only had to introduce four one-residue indels. The accuracy of the models was assessed with the Verify3D Structure Evaluation Server (http://nihserver.mbi.ucla.edu/Verify_3D/) (42).

Difference distance matrix map

We used an *ad hoc* program written in C applying previously developed concepts to detect domain motion and identify regions that move closer upon conformational changes (43). Distances were calculated between all pair of C_α of one reference structure (open), and a second pairwise distance matrix was calculated for the target (closed) structure. Afterwards, the target matrix was subtracted from the reference matrix to calculate the Δdistance plot (https://github.com/ballicoragroup/didimama).

Hinge analysis

In order to detect possible local conformations or hinges, we performed an analysis with the ad hoc program "hingescan" (https://github.com/ballicoragroup/hingescan). We compared the crystal structure of the open form of the *N. europaea* sucrose synthase with a closed form homology model of the same enzyme. To detect if there is a significant local conformational change around a given residue ("hinge"), we extracted the coordinates of a given number (n) of C_α before the putative hinge and the same given number (n) of residues after (window size = 2n+1). This was done for both the open and closed forms and obtained two fragments to

compare. After optimal rigid body superposition of only these two set of coordinates, an average distance was calculated (root-mean-square deviation, RMSD). This RMSD calculated in these conditions was called the "hinge score". When this score is at a peak, the "flanking" n number of C_α at both sides display a maximum change between the two structures. For that reason, a hinge is detected. The bigger the window, the bigger the domain movement is detected surrounding the hinge. To identify hinges that link small and bigger domains, different window sizes were scanned. A flowchart illustrating the process is in Fig. S1.

RESULTS AND DISCUSSION

Sequence analysis

To know how the sucrose synthase from *N. europaea* relates to others from divergent organisms we constructed a phylogenetic tree using 117 amino acid sequences retrieved from the NCBI database (Fig. 2, Table S1, and Fig. S2). The tree comprised seven major branches, containing the sequences from cyanobacteria (group I; 21 sequences), proteobacteria (groups II and III; 17 sequences), the moss *Physcomitrella patens* subsp. patens (group IV; 4 sequences), and vascular plants (groups V, VI, and VII; 75 sequences) (Fig. 2). The shape of the tree shown in Fig. 2 is similar to the one published by Kolman et al. (8). Group I is subdivided in two branches, containing the sequences encoded by the *susA* (cyan) and *susB* (orange) genes (Fig. 2) (8). Most sequences from proteobacteria are included in group III (including β-, γ- and δ-proteobacteria); though, the sequences from δ-proteobacterium MLMS-1 and *Desulfurivibrio alkaliphilus* AHT2 are in a diverging branch (group II) (Fig. 2). Sucrose

synthase sequences from *P. patens* subsp. patens (group IV) are clearly separated from those of vascular plants (Fig. 2). Interestingly, groups V and VII are further divided in two major branches, containing the sequences from dicots (green) and monocots (blue), respectively. This separation is less clear in group VI (Fig. 2). The sucrose synthase from *N. europaea* is in a small branch with other β-proteobacteria in group III (proteobacteria). Clearly, it is well separated from plant and cyanobacterial enzymes, although they share a significant similarity. For instance, the identity between sequences from *N. europaea* and those from *Nostoc* sp. PCC 7120 susA, *P. patens*, *Zea mays* sucrose synthase 1, and *A. thaliana* sucrose synthase 1 were 45.3, 49.3, 50.4, and 50.3%, respectively (Fig. S2). These values are indicative of a high structural conservation among enzymes from very divergent organisms.

Protein expression and characterization

The gene of the putative sucrose synthase in *N. europaea* (NCBI Protein ID NP_841269) codes for 794 amino acids. To shed light on sucrose metabolism of group III (Fig. 2), we amplified this sequence and expressed the recombinant protein in *E. coli* cells. The enzyme was purified to homogeneity by HisTrap column and gel filtration chromatography as mentioned in "Materials and Methods". The recombinant protein migrated in SDS-PAGE as a single band of ~95 kDa (data not shown), which is in good agreement with the predicted molecular mass of 93 kDa (including the His-tag provided by the pET28c vector). The enzyme eluted from the Superdex 200 (size exclusion) column as a protein of ~360 kDa (data not shown), suggesting a tetrameric quaternary structure, as it was reported for cyanobacterial and plant sucrose synthases (16, 19, 23).

Substrate specificity of the sucrose synthase from *N. europaea*

Sucrose synthases from plants have shown a certain degree of promiscuity to transfer glucoses from ADP-Glc and UDP-Glc, though UDP-Glc is generally preferred. We tested the substrate specificity of sucrose synthase from *N. europaea* in the sucrose synthesis direction (Table 1), and observed that ADP-Glc is a more efficient substrate than UDP-Glc. The main difference is not given by V_{max}, but by a higher apparent affinity towards ADP-Glc. The $S_{0.5}$ for ADP-Glc is 0.044 mM in presence of optimal concentrations of Fru (20 mM); whereas the $S_{0.5}$ for UDP-Glc is 0.98 mM in presence of optimal concentrations of Fru (500 mM). On the other hand, the apparent affinity for Fru is higher in presence of ADP-Glc rather than UDP-Glc. The $S_{0.5}$ for Fru at saturated concentrations of ADP-Glc is 5.6 mM whereas the $S_{0.5}$ for Fru in presence of UDP-Glc is significantly higher. Because of the high concentrations of Fru needed to reach saturation, it is not possible to measure the $S_{0.5}$ for Fru with high precision; but it is at least ~20-fold higher (120 mM). The catalytic efficiencies calculated for ADP-Glc and $Fru_{(ADP\text{-}Glc)}$ were 17- and 37-fold higher than those obtained for UDP-Glc and $Fru_{(UDP\text{-}Glc)}$, respectively (Table 1). These results indicate that the sucrose synthase from *N. europaea* prefers ADP-Glc over UDP-Glc as substrate. Similar conclusions were obtained for the enzyme from the cyanobacterium *T. elongatus*, which showed a 26-fold higher catalytic efficiency for ADP-Glc than UDP-Glc (16). As it was stated for *T. elongatus* (16), this suggests that the metabolism of sucrose could be linked to the synthesis of glycogen, since ADP-Glc is the donor for its polymerization.

X-ray diffraction, data processing, model building, and refinement

The best data set collected at synchrotron beamline was processed to 3.05 Å and indexed as space group P65. It was integrated and scaled producing good statistics (Table

2). After the molecular replacement search, four copies of the starting model described in "Materials and Methods" were found in one asymmetric unit. Iterative cycles of model building and refinement were conducted yielding a well-defined structure with R_{work} and R_{free} values of 17.37 % and 21.75 %, respectively (Table 2). The truncated GT-B(D) domain was built according to the electron density map. The final structural model contains all the residues except the first three at the N-terminus and the last two at the C-terminus of the amino acid sequence (Fig. 1).

Structural analysis of the sucrose synthase from *N. europaea*

Overall structure. Although the resolution of the data set was 3.05 Å, the backbone of the protein and some of the key residues side chains were well defined by the electron density (Fig. S3 and Fig. S4). This allowed us to conduct detailed structural analysis on sucrose synthase's conformational changes involving backbone movement, which are relevant to the catalytic cycle. The crystal structure displayed a similar fold to the previously reported structural model from the *A. thaliana* enzyme (PDB ID 3S29) (23). The sucrose synthase from *N. europaea* is a tetramer composed of four identical subunits (Fig. 1A), where each monomer contains four domains (Fig. 1B).

The first domain designated as "Sucrose Synthase N-terminal-1" (SSN-1) included residues 1-112 (Fig. 1B, red) and contained five α-helices and four β-strands. The second domain, which included residues 142-264, is the "Sucrose Synthase N-terminal-2" (SSN-2) domain (Fig. 1B, green) with five α-helices. Domain SSN-1 and SSN-2 correspond to domains CTD and EPBD in the enzyme from *A. thaliana* (23). CTD and EPBD stand for "cellular targeting domain" and "ENOD40 peptide-binding domain", which indicate the domain functions for the plant enzyme. In the case of the bacterial

form, the roles for these domains are not known, thus the nomenclature is only based on structure. Both the third and fourth domains constitute a typical GT-B fold of glycosyltransferases (24). The third is a domain that typically binds the nucleotide donor for the glycosyl group in that family (23, 25, 44). For this reason, we refer to it as GT-B(D) domain (Fig. 1B, cyan), although in sucrose synthase the transfer of glucosyl group is reversible. This nomenclature also matches the systematic name of the sucrose synthase (NDP-glucose:D-fructose 2-α-D-glucosyltransferase). The GT-B(D) domain includes residues 514-742 with eight α-helices and three β-strands. The fourth domain is the GT-B(A) domain (Fig. 1B, blue and yellow), which consists of residues from three separate regions. These separate regions are encompassed by the SSN-1, SSN-2 and GT-B(D) domains in the center of the monomer. The first region is a linker (residues 113-141) that joins SSN-1 and SSN-2 but structurally integrated to GT-B(A). The other two regions are 265-513, and 743-794. The GT-B(A) domain included nine α-helices and eight β-strands and functions in the GT-B family as the sugar acceptor (A) in catalysis (23, 25, 44).

As mentioned above, the identity between sequences from *N. europaea* and *A. thaliana* is considerably high (50.3%). When the different domains were analyzed separately, we found identity values of 26.2% for SSN-1 (CTD), 40.2% for SSN-2 (EPBD), 52.4% for GT-B(D), and 61.9% for GT-B(A), suggesting a high structural conservation. A comparison between the *A. thaliana* and *N. europaea* x-ray structures confirms it. With the exception of conformational changes, each of the folds for their respective domains is identical. The fact that the structure is so conserved, even for the domains that are not related to catalysis, would suggest that certain non-catalytic

functional roles have been preserved or adapted. On the other hand, SSN-1 (CTD in *A. thaliana*) does not have the Ser that is phosphorylated in plants, indicating that it is a role acquired in eukaryotes. Therefore, it is not certain whether *N. europaea* sucrose synthase is regulated for binding macromolecular structures such as actin or membranes as plant enzymes do (26, 45). Prokaryotes do not have cytoskeleton, although actin related proteins have been detected in *Anabaena* species (46). Whether sucrose synthase from bacteria can actually interact with actin or similar structures is a matter of further studies.

In *N. europaea*, SSN-2 is involved in the oligomerization forming one of the contacts between subunits. It is not clear if it has any other physiological role. In *A. thaliana*, EPBD (SSN-2 in *N. europaea*), together with the CTD domain (SSN-1 in *N. europaea*), forms a groove hypothesized to bind actin (23). In our structure, the same structural arrangement is present (data not shown) highlighting the possibility that a similar role has been conserved. However, this needs to be investigated.

The obtained *N. europaea* sucrose synthase structure with no substrates bound has a clearly different overall conformation when compared to the *A. thaliana* structure with UDP and Fru (23). This implies that substrate binding induces significant conformational changes (Fig. 3), and correlates with similar conformational changes that occur upon binding of substrates in other GT-B retaining glycosyltransferases (25, 47). After superimposition of only the GT-B(A) domains of the *A. thaliana* and *N. europaea* structures (using the least squares function in COOT), the SSN-1, SSN-2, and GT-B(A) domains overlapped well while the GT-B(D) domains were in a different relative position. The angle between the GT-B(A) and GT-B(D) domains in the obtained structure was about 23.5 degrees wider than in *A. thaliana*. Based on such comparison, we suggest

that the *N. europaea* structure in this work was in an "open" conformation whereas the *A. thaliana* form was "closed" (23). We have identified some distinct structural determinants (hinges and latches) related to the movements of the sugar (GT-B(A)) and nucleotide (GT-B(D)) binding domains.

Sugar-binding GT-B(A) domain. In this analysis, we compared the open structure crystal structure of *N. europaea* enzyme with the homology model in a closed conformation built as described in "Materials and Methods". Considering how modeling works, and that the closed structure template (*A. thaliana*) has no gaps with the *N. europaea* target in the sites of interest, the backbone comparison with the model is as reliable as comparing the backbones of both structures directly. The RMSD of backbone between the model obtained and the closed *A. thaliana* template was 0.29 Å. However, the use of the model is more convenient since the number is not shifted, which would be really confusing in the following analysis. One of the important assumptions we make is that the closed structure of the *A. thaliana* enzyme is a fair representation of the closed structure of that from *N. europaea*. We believe that this is a reasonable assumption, at least in the critical areas. Otherwise, the backbone of critical residues may not align properly for catalysis.

Analysis of a difference distant matrix map of the Fru-binding GT-B(A) domain as described in "Materials and Methods" highlights three main regions that move closer upon sugar binding (Fig. 4). These are 325-375 to 280-290 (~5 Å), 425-435 to 280-290 (~4 Å), and 425-435 to 325-375 (~3 Å) (Fig. 4). Other pair of regions that move towards each other are 280-290 to 490-505 (~3 Å) and 280-290 to 450-460 (~2 Å) (Fig. 4). From this analysis, the area 280-290 is the most involved in an induced fit interaction with Fru.

Further inspection of these areas reveals that Fru induces local conformational changes via superimposition of the GT-B(A) domains of the *A. thaliana* (closed) and the *N. europaea* (open) sucrose synthase (Fig. 5). These include the side chain of K431 and the backbone of residues 288-290. The re-shaping of the Fru binding site facilitates the closing via a set of inter-domain hydrogen bonds (Fig. 5 in green). These local conformational changes along with the presence of Fru further promote the interactions between the GT-B(A) and GT-(D) domains. Thus, we propose that Fru binding contributes to stabilizing the closed structure.

Nucleotide-binding GT-B(D) domain. The GT-B(D) domain binds to sugar nucleotide (synthesis direction) or nucleotide (cleavage direction) substrates. When overlapping GT-B(D) domains from both *A. thaliana* and *N. europaea* structures, the residues interacting with the phosphate and ribose moieties of the nucleotides are not only conserved, but also at the same positions (Fig. 6). On the other hand, two residues interacting with the nucleotide base are not conserved. Residues Q648 and N654 from *A. thaliana* are replaced by R636 and A642 in the *N. europaea* sucrose synthase, respectively. This difference creates a more spacious binding site in *N. europaea*, which may accommodate bulkier adenosine nucleotide substrates. Modeling an ADP ligand into the *N. europaea* structure shows that the site may have a deeper pocket, which would be needed not to clash with the adenine ring (Fig. 6 and Fig. 7). Similar sequence differences were observed in the sucrose synthase from *T. elongatus* (16). Based on sequence analysis and homology modeling it was suggested that these two residues could be responsible for the preference towards ADP/ADP-Glc over other nucleotides such as UDP/UDP-Glc in the cyanobacterial enzyme (16). It is important to notice that the side

chains in R636 and A642 in the *N. europaea* sucrose synthase are not conserved in the *E. coli* glycogen synthase, which is another glycosyltransferase that binds ADP-Glc (48). *E. coli* glycogen synthase has a different motif in that position with a Tyr and Ser instead of Arg and Ala (25, 47, 48), implying a different structural arrangement for accommodating ADP-Glc. Overall, the nucleotide binding to the GT-B(D) domain does not seem to trigger significant local conformational changes (Fig. 6). The direct interactions with the nucleotide do not make major contributions to the induced fit mechanism.

Hinge Analysis. We scanned the structures of the acceptor and donor domains for hinges and subtle conformational changes that could be functionally important in catalysis. We used the in-house program *hingescan* described in "Materials and Methods". Using several window sizes we detected several local conformational changes (Fig. S5). For a window size of 51, we detected two clear hinge elements near residues ~515 and ~744 (Fig. 8). These two elements actually form a single "hinge" that comprises a hydrogen bond between two conserved residues (G514 and W743) in a flexible area (Fig. 3, Fig. S4, and Fig. S5). These two residues remain at the same position in both the open and the closed conformations of the enzyme. For smaller windows, we found other significant local secondary structure rearrangements between the open and closed structures (Fig. S5, Fig. S6, and Fig. 8). Upon closing, two α-helices in the GT-B(D) domain are extended, and a β-strand replaces a previously coiled stretch. The outcome is a more ordered structure of the GT-B(D) domain. We propose that this secondary structure rearrangement, despite the local entropy decrease, would release extra energy to close the conformation facilitating the binding of substrates.

458 There are also differences between the conformations of the SSN-1 and SSN-2
459 domains from the open structure of the *N. europaea* enzyme and the model of the closed
460 form (Fig. 8). The analysis detected hinges because of local differences, and there are
461 four major regions with scores above 2 (Fig. 8, Fig S5, Fig. S7). This predicts that some
462 of the loops in these two domains are quite flexible, but we cannot assign a functional
463 role to them (Fig. S7). In *A. thaliana*, the flexibility of the CTD domain (SSN-1) is
464 hypothesized to have a role in actin binding (23).

465 *Latches*. A feature that contributes to the stabilization of the closed form is a
466 "latch", E609, which comprises the highly conserved E609 residue located at the
467 periphery of the GT-B(D) domain (Fig. S2). Going from an open to a closed
468 conformation, this glutamate residue moves ~11 Å towards the GT-B(A) domain and
469 ends up hydrogen-bonded to two tyrosine residues (Y432 and Y446) stabilizing the
470 closing (Fig. 3 and Fig. S8). Interestingly, there were small secondary structure
471 rearrangements in the vicinity of this latch, which could facilitate the interaction between
472 E609 and the two tyrosines (Fig. S6). On the other side of the active site, opposite to the
473 latch described, there is a hydrophobic patch that also contributes to the closing (Fig. 9).
474 Two hydrophobic residues (M635, L637) in the GT-B(D) domain get in contact with a
475 hydrophobic cluster (V281, L282 and L284) in the GT-B(A) domain, upon closing. The
476 side chain of N280 also provides a methylene to build a non-polar pocket that latches on
477 to M635 and L637 (Fig. 9). On the other hand, the amide polar group is exposed to the
478 solvent.

479 The closed structure seems to induce stronger interactions with the nucleotide and
480 vice versa. In the *N. europaea* sucrose synthase, the conserved E671 is in the same

position as E369 in the *E. coli* trehalose-6-phosphate synthase (OtsA) (48, 49). In OtsA, as well as in the close conformation of the *A. thaliana* sucrose synthase (E683), the carboxylate of this side chain forms two hydrogen bonds with the hydroxyl groups of the ribose of the nucleotide. In these enzymes, the carboxylate is surrounded by hydrophobic residues (Y520, Y646, L667 and T668 in *N. europaea* sucrose synthase; Y533, Y658, L679 and T680 in the *A. thaliana* enzyme), which makes the hydrogen bonds stronger in the non-polar environment. In the open form of the sucrose synthase structure, V291 (V306 in *A. thaliana*) moves away from the side chain of the glutamate residue (Fig. S9). This implies that the closing recruits a non-polar side chain to completely surround the carboxylate. Consequently, nucleotide binding stabilizes the carboxylate charge and facilitates the interaction with V291 upon closing. Interestingly, in another glycosyltransferase such as bacterial glycogen synthase, E671 has been replaced by a Tyr, and V291 was replaced by Asp, thus, switching their roles (48). Therefore, this ligand-dependent interaction may be a common feature in this family of enzymes.

Site directed mutagenesis of critical residues

Previously, important residues for catalysis were identified in other retaining GT-B glycosyltransferases. A triad of critical residues has been found in the active site of maltodextrin phosphorylase (50, 51) and glycogen synthase from *E. coli* (52). Based on X-ray structures, these residues were also predicted to be important for catalysis in OtsA (49) and the *A. thaliana* sucrose synthase (23). The homologue residues in *N. europaea* are R567, K572, and E663 (Fig. 10). When any of those residues were replaced by alanine the activity in the direction of sucrose synthesis severely decreased, either in presence of ADP-Glc or UDP-Glc (Table 3). The most active of all these mutants was

E663A in presence of ADP-Glc, but it was still 200-fold less active than the wild type. These results indicate that this triad is critical in sucrose synthases. Consequently, it also suggests that sucrose synthases together with all other retaining glycosyltransferases with a GT-B fold share the same reaction mechanism (Fig. 10).

Structural and mechanistic consequences of the open/closed conformational change

Large conformational movements may have a large impact on the architecture of the active site. For that reason, it is important to analyze how critical catalytic residues are arranged in the closed and open structure of sucrose synthase. We have identified and confirmed by mutagenesis three critical side chains in *N. europaea* sucrose synthase (R567, K572, E663, corresponding to R580, K485, E675 in *A. thaliana*). In addition, the comparison with other glycosyltransferases predicts another interaction (protein backbone-substrate) that stabilizes the transition state (24), which is not possible to be replaced by mutagenesis. The comparison between the open and closed forms of *N. europaea* and *A. thaliana* sucrose synthases, respectively, provide important information. But, the arrangement of the critical residues needs to be put in context of the reaction mechanism.

It has been proposed that retaining glycosyltransferases either have a S_Ni-like mechanism with a oxocarbenium-phosphate short lived ion pair intermediate, or a S_Ni mechanism forming an oxocarbenium ion-like transition state that is not totally dissociated from the donor and acceptor (24, 53, 54). In either of those two cases, an important stabilization of the transition state would be based on the interaction between the anomeric carbon (C1) of the sugar being transferred and the oxygen of the main chain of a His residue (24, 55). Recently, an alternative elimination/addition mechanism has

been proposed for the *Pyrococcus abyssi* glycogen synthase (56), which was argued to be compatible with the current available data. In this mechanism, a general base is needed to extract a proton from C2 of the glycosyl group. The authors proposed this base is the same main chain oxygen from the His residue mentioned above. Interestingly, in the crystal structure of the *A. thaliana* sucrose synthase, the oxygen of the main chain of H438 (H425 in *N. europaea*) is at a close distance of both C1 and C2 of a proposed 1,5-anhydro-D-arabino-hex-1-enitol (Fig. S10). This ligand mimics the planar structure of the transition state in either the S_Ni, S_Ni-like, or the elimination/addition mechanism (56). This type of *in situ* generated intermediate was also observed in the *E. coli* and *P. abyssi* glycogen synthase (25, 56). Regardless of these alternative mechanisms, it must be critical that the main chain oxygen of H425 in *N. europaea* sucrose synthase is near C1/C2 in the transition state. Since this residue is located in the GT-B(A) domain, and there are other critical residues in GT-B(D) (Table 3), a precise arrangement between these two domains is necessary for a proper architecture of the catalytic site. Only in the closed structure all these functional groups would be at the right distance for catalysis (Fig. S10). Therefore, one of the roles of the closing is to bring the critical residues to a proper position.

It is tempting to argue that UDP-Glc/ADP-Glc can induce the closing, based on the fact that the base and the ribose have numerous contacts with the GT-B(D) domain, and the glycosyl group with the opposing GT-B(A) domain (23). However, it is not clear how stable the closed form would be in presence of the donor without the acceptor. During the crystallization process, the *A. thaliana* enzyme cleaves UDP-Glc to generate UDP and possibly 1,5-anhydro-D-arabino-hex-1-enitol (or its tautomer 1,5-anhydro-D-

fructose) yielding a closed structure (23). But, this structure, which could occur transiently, may have been driven and stabilized by the extremely slow generation of 1,5-anhydro-D-arabino-hex-1-enitol. It is expected that this putative transition state analog binds favorably to the most active form of the enzyme, which in this case, is the closed one.

The rationale for a conformational change induced by substrates ("induced fit") was first described by Koshland to explain why a specific (good) substrate reacts faster than smaller (poor) alternatives that can also fit into the active site (57). If the changes that lead to a precise orientation of catalytic groups occur only upon binding of the (good) substrate, another (poor) substrate that does not trigger those conformational changes will not react effectively, even at high concentrations. This concept or at least its interpretation has been controversial (58, 59). On the other hand, Fersht stated that an induced fit mechanism does not increase substrate specificity *per se* (59), and the only contribution that matters is the relative binding affinities to the transition states of the competing reactions. According to this, to increase the sucrose synthase specificity for the acceptor Fru against water, selective interactions seems to be maximized by surrounding Fru completely by different functional groups from the enzyme. Consequently, the active site is isolated from the solution as it is shown in Fig. S11. For that reason, an induced fit mechanism becomes an indirect necessity to allow substrates and products to enter and leave, while maximizing a selective interaction with Fru.

For retaining glycosyltransferases, another key issue is the stabilization of the β-phosphate to make it a better leaving group (54, 60). A hydroxyl group from the acceptor (Fru in this case) participates in a hydrogen bond with the β-phosphate. Consequently, the

oxygen of this group becomes a better nucleophile to attack the C1 of the forming oxocarbenium ion (54). Water could in theory compete with the acceptor (Fru) for this role, but is a poor substrate, probably because it does not stabilize the closed structure as well as Fru does. Fru not only interact with the phosphate leaving group, but also interacts more tightly with the closed form. Not only the distances between the residues in the GT-B(A) domain that contact Fru gets closer upon closing, but also networks of interactions of Fru with the GT-B(D) domain are established (Fig. 6 and Fig. S11). Noteworthy are the interaction of Fru with R580 (*A. thaliana*) and the hydrogen bond with K444 that brings the Y445 closer to E621, forming the latch (Fig. S11, panels D and E). Interestingly, the closed structure of the *A. thaliana* enzyme with the cleaved products of UDP-Glc seems to shape the active site to readily accommodate Fru. Even if Fru is absent, the site is nearly identical to the structure with Fru bound (Fig. S11, RMSD 0.27 Å). On the other hand, the structure of the open form of the *N. europaea* enzyme does not have all these residues at a proper distance to bind Fru (Fig. S11, C). This indicates that Fru would preferentially bind to the closed form, stabilizing it.

This mechanism in which the catalytic residues get into places upon closing may explain why it is not trivial to obtain a closed structure with an intact sugar-nucleotide. For instance, crystal structures of the closed forms were obtained for the *E. coli* glycogen synthase and the *A. thaliana* sucrose synthase grown in presence the sugar nucleotide, but the glycosyl group was slowly cleaved (23, 25, 56). There are other retaining GT-B structures with a sugar nucleotide bound, but those were described as “semi-closed” (44). A mechanism with a domain movement that allows the exchange of ligands to the solution is not unique for sucrose synthase and may be general among retaining GT-B

enzymes. However, not all of them may require such a large conformational change. The glycogen synthase was another case with closed and wide open structures described (25, 47, 56). The sucrose-6-phosphate synthase must also have the same type of behavior, but only an open structure is available (61). In other cases, open/closed structures have been obtained, but the most significant movements were local rearrangement of loops (rather than a large domain rearrangement) such as in OtsA (44) and VldE (62, 63).

CONCLUSIONS

In this manuscript, we observed an "open" conformation for sucrose synthases. Based on the comparison with a previously published "closed" sucrose synthase structure (23), a "hinge-latch" combination was identified as a critical feature responsible for the open-close enzyme actions.

We identified three highly conserved amino acids proposed to be critical for catalysis. We concluded that the triad composed of residues R567, K572, and E663 (numbers according to the *N. europaea* enzyme) plays a key role not only in sucrose synthases, but also in all the retaining GT-B glycosyltransferases (23, 49-52).

With both structural and kinetic results we propose that the sucrose synthase from *N. europaea* has a substrate preference in favor of ADP/ADP-Glc over UDP/UDP-Glc. This behavior is similar to the one observed for *T. elongatus* sucrose synthase (16).

The evolutionary origin of enzymes from sucrose metabolism in proteobacteria has been previously discussed (4, 5, 8, 64). The evolution of sucrose synthases in cyanobacteria, proteobacteria, and plants is not yet fully understood, but most likely it involved horizontal gene transfers. On one hand, the sucrose synthase from *N. europaea*

619 is closer to the plant enzymes in the phylogenetic tree (Fig. 2), but on the other hand, the
620 specificity for nucleotides is similar to several cyanobacterial enzymes examined (8, 16).
621 It is possible that the enzyme from *N. europaea* evolved from a protein already present in
622 the common ancestor of proteobacteria and cyanobacteria (10).
623
624 **AKNOWLEDGEMENTS**
625 This work was supported by grants from ANPCyT (PICT'12 2439) and
626 CONICET (PIP 112-201101-00438) to A.A.I.; UNL (CAI+D 2011) to C.M.F. and
627 A.A.I.; NSF (MCB-1024945) to M.A.B.; and NSF (CHE-1308672) to D.L. M.D.A.D.,
628 C.M.F. and A.A.I. are researchers from CONICET.
629

REFERENCES

1. **Koch K.** 2004. Sucrose metabolism: regulatory mechanisms and pivotal roles in sugar sensing and plant development. Curr Opin Plant Biol **7:**235-246.
2. **Lunn JE.** 2008. Sucrose Metabolism, Encyclopedia of Life Sciences. John Wiley & Sons.
3. **Winter H, Huber SC.** 2000. Regulation of sucrose metabolism in higher plants: localization and regulation of activity of key enzymes. Crit Rev Biochem Mol Biol **35:**253-289.
4. **MacRae E, Lunn J.** 2012. Photosynthetic Sucrose Biosynthesis: An Evolutionary Perspective, p. 675-702. *In* Eaton-Rye JJ, Tripathy BC, Sharkey TD (ed.), Photosynthesis, vol. 34. Springer Netherlands.
5. **Salerno GL, Curatti L.** 2003. Origin of sucrose metabolism in higher plants: when, how and why? Trends Plant Sci **8:**63-69.
6. **Cumino AC, Marcozzi C, Barreiro R, Salerno GL.** 2007. Carbon cycling in *Anabaena* sp. PCC 7120. Sucrose synthesis in the heterocysts and possible role in nitrogen fixation. Plant Physiol **143:**1385-1397.
7. **Curatti L, Giarrocco L, Salerno GL.** 2006. Sucrose synthase and RuBisCo expression is similarly regulated by the nitrogen source in the nitrogen-fixing cyanobacterium *Anabaena* sp. Planta **223:**891-900.
8. **Kolman MA, Torres LL, Martin ML, Salerno GL.** 2012. Sucrose synthase in unicellular cyanobacteria and its relationship with salt and hypoxic stress. Planta **235:**955-964.

9. **Yamanaka T.** 2008. Chemolithoautotrophic Bacteria: Biochemistry and Environmental Biology. Springer, Japan.

10. **Teske A, Alm E, Regan JM, Toze S, Rittmann BE, Stahl DA.** 1994. Evolutionary relationships among ammonia- and nitrite-oxidizing bacteria. J Bacteriol **176:**6623-6630.

11. **Vannelli T, Logan M, Arciero DM, Hooper AB.** 1990. Degradation of halogenated aliphatic compounds by the ammonia- oxidizing bacterium *Nitrosomonas europaea*. Appl Environ Microbiol **56:**1169-1171.

12. **Chain P, Kurtz S, Ohlebusch E, Slezak T.** 2003. An applications-focused review of comparative genomics tools: capabilities, limitations and future challenges. Brief Bioinform **4:**105-123.

13. **Hooper AB.** 1969. Biochemical basis of obligate autotrophy in *Nitrosomonas europaea*. J Bacteriol **97:**776-779.

14. **Baroja-Fernandez E, Munoz FJ, Saikusa T, Rodriguez-Lopez M, Akazawa T, Pozueta-Romero J.** 2003. Sucrose synthase catalyzes the de novo production of ADPglucose linked to starch biosynthesis in heterotrophic tissues of plants. Plant Cell Physiol **44:**500-509.

15. **Delmer DP.** 1972. The purification and properties of sucrose synthetase from etiolated *Phaseolus aureus* seedlings. J Biol Chem **247:**3822-3828.

16. **Figueroa CM, Asencion Diez MD, Kuhn ML, McEwen S, Salerno GL, Iglesias AA, Ballicora MA.** 2013. The unique nucleotide specificity of the sucrose synthase from *Thermosynechococcus elongatus*. FEBS Lett **587:**165-169.

17. **Grimes WJ, Jones BL, Albersheim P.** 1970. Sucrose synthetase from *Phaseolus aureus* seedlings. J Biol Chem **245:**188-197.

18. **Morell M, Copeland L.** 1985. Sucrose synthase of soybean nodules. Plant Physiol **78:**149-154.

19. **Porchia AC, Curatti L, Salerno GL.** 1999. Sucrose metabolism in cyanobacteria: sucrose synthase from *Anabaena* sp. strain PCC 7119 is remarkably different from the plant enzymes with respect to substrate affinity and amino-terminal sequence. Planta **210:**34-40.

20. **Ross HA, Davies HV.** 1992. Purification and Characterization of Sucrose Synthase from the Cotyledons of *Vicia faba* L. Plant Physiol **100:**1008-1013.

21. **Tanase K, Yamaki S.** 2000. Purification and characterization of two sucrose synthase isoforms from Japanese pear fruit. Plant Cell Physiol **41:**408-414.

22. **Curatti L, Giarrocco LE, Cumino AC, Salerno GL.** 2008. Sucrose synthase is involved in the conversion of sucrose to polysaccharides in filamentous nitrogen-fixing cyanobacteria. Planta **228:**617-625.

23. **Zheng Y, Anderson S, Zhang Y, Garavito RM.** 2011. The structure of sucrose synthase-1 from *Arabidopsis thaliana* and its functional implications. J Biol Chem **286:**36108-36118.

24. **Lairson LL, Henrissat B, Davies GJ, Withers SG.** 2008. Glycosyltransferases: structures, functions, and mechanisms. Annu Rev Biochem **77:**521-555.

25. **Sheng F, Jia X, Yep A, Preiss J, Geiger JH.** 2009. The crystal structures of the open and catalytically competent closed conformation of *Escherichia coli* glycogen synthase. J Biol Chem **284:**17796-17807.

26. **Hardin SC, Winter H, Huber SC.** 2004. Phosphorylation of the amino terminus of maize sucrose synthase in relation to membrane association and enzyme activity. Plant Physiol **134:**1427-1438.

27. **Klotz KL, Finger FL, Shelver WL.** 2003. Characterization of two sucrose synthase isoforms in sugarbeet root. Plant Physiol Biochem **41:**107-115.

28. **Schäfer WE, Rohwer JM, Botha FC.** 2004. A kinetic study of sugarcane sucrose synthase. Eur J Biochem **271:**3971-3977.

29. **Wolosiuk RA, Pontis HG.** 1974. Studies on sucrose synthetase. Kinetic mechanism. Arch Biochem Biophys **165:**140-145.

30. **Kuhn ML, Falaschetti CA, Ballicora MA.** 2009. *Ostreococcus tauri* ADP-glucose pyrophosphorylase reveals alternative paths for the evolution of subunit roles. J Biol Chem **284:**34092-34102.

31. **Sambrook J, Russell DW.** 2001. Molecular Cloning: A laboratory Manual, 3rd ed, vol. 2. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

32. **Laemmli UK.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227:**680-685.

33. **Hall TA.** 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser **41:**95-98.

34. **Gouy M, Guindon S, Gascuel O.** 2010. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol **27:**221-224.

35. **Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov**

GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS. 2011. Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr **67:**235-242.

36. **Diederichs K, Karplus PA.** 2013. Better models by discarding data? Acta Crystallogr D Biol Crystallogr **69:**1215-1222.

37. **McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ.** 2007. Phaser crystallographic software. J Appl Crystallogr **40:**658-674.

38. **Emsley P, Cowtan K.** 2004. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr **60:**2126-2132.

39. **Murshudov GN, Vagin AA, Dodson EJ.** 1997. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr **53:**240-255.

40. **Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH.** 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr **66:**213-221.

41. **Sali A, Blundell TL.** 1993. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol **234:**779-815.

42. **Luthy R, Bowie JU, Eisenberg D.** 1992. Assessment of protein models with three-dimensional profiles. Nature **356:**83-85.

741 43. **Richards FM, Kundrot CE.** 1988. Identification of structural motifs from
742 protein coordinate data: secondary structure and first-level supersecondary
743 structure. Proteins **3:**71-84.

744 44. **Gibson RP, Tarling CA, Roberts S, Withers SG, Davies GJ.** 2004. The donor
745 subsite of trehalose-6-phosphate synthase: binary complexes with UDP-glucose
746 and UDP-2-deoxy-2-fluoro-glucose at 2 A resolution. J Biol Chem **279:**1950-
747 1955.

748 45. **Duncan KA, Huber SC.** 2007. Sucrose synthase oligomerization and F-actin
749 association are regulated by sucrose concentration and phosphorylation. Plant Cell
750 Physiol **48:**1612-1623.

751 46. **Guerrero-Barrera AL, Garcia-Cuellar CM, Villalba JD, Segura-Nieto M,**
752 **Gomez-Lojero C, Reyes ME, Hernandez JM, Garcia RM, de la Garza M.**
753 1996. Actin-related proteins in *Anabaena spp*. and *Escherichia coli*. Microbiology
754 **142 (Pt 5):**1133-1140.

755 47. **Buschiazzo A, Ugalde JE, Guerin ME, Shepard W, Ugalde RA, Alzari PM.**
756 2004. Crystal structure of glycogen synthase: homologous enzymes catalyze
757 glycogen synthesis and degradation. EMBO J **23:**3196-3205.

758 48. **Yep A, Ballicora MA, Preiss J.** 2006. The ADP-glucose binding site of the
759 Escherichia coli glycogen synthase. Arch Biochem Biophys **453:**188-196.

760 49. **Gibson RP, Turkenburg JP, Charnock SJ, Lloyd R, Davies GJ.** 2002. Insights
761 into trehalose synthesis provided by the structure of the retaining
762 glucosyltransferase OtsA. Chem Biol **9:**1337-1346.

763 50. **Schinzel R, Drueckes P.** 1991. The phosphate recognition site of *Escherichia*
764 *coli* maltodextrin phosphorylase. FEBS Lett **286:**125-128.
765 51. **Schinzel R, Palm D.** 1990. *Escherichia coli* maltodextrin phosphorylase:
766 contribution of active site residues glutamate-637 and tyrosine-538 to the
767 phosphorolytic cleavage of alpha-glucans. Biochemistry **29:**9956-9962.
768 52. **Yep A, Ballicora MA, Preiss J.** 2004. The active site of the *Escherichia coli*
769 glycogen synthase is similar to the active site of retaining GT-B
770 glycosyltransferases. Biochem Biophys Res Commun **316:**960-966.
771 53. **Lee SS, Hong SY, Errey JC, Izumi A, Davies GJ, Davis BG.** 2011.
772 Mechanistic evidence for a front-side, SNi-type reaction in a retaining
773 glycosyltransferase. Nat Chem Biol **7:**631-638.
774 54. **Gomez H, Lluch JM, Masgrau L.** 2013. Substrate-assisted and nucleophilically
775 assisted catalysis in bovine alpha1,3-galactosyltransferase. Mechanistic
776 implications for retaining glycosyltransferases. J Am Chem Soc **135:**7053-7063.
777 55. **Mitchell EP, Withers SG, Ermert P, Vasella AT, Garman EF, Oikonomakos**
778 **NG, Johnson LN.** 1996. Ternary complex crystal structures of glycogen
779 phosphorylase with the transition state analogue nojirimycin tetrazole and
780 phosphate in the T and R states. Biochemistry **35:**7341-7355.
781 56. **Diaz A, Diaz-Lobo M, Grados E, Guinovart JJ, Fita I, Ferrer JC.** 2012. Lyase
782 activity of glycogen synthase: Is an elimination/addition mechanism a possible
783 reaction pathway for retaining glycosyltransferases? IUBMB Life **64:**649-658.
784 57. **Koshland DE.** 1958. Application of a Theory of Enzyme Specificity to Protein
785 Synthesis. Proc Natl Acad Sci U S A **44:**98-104.

786 58. **Johnson KA.** 2008. Role of induced fit in enzyme specificity: a molecular
787 forward/reverse switch. J Biol Chem **283:**26297-26301.
788 59. **Fersht A.** 1999. Structure and mechanism in protein science : a guide to enzyme
789 catalysis and protein folding. W.H. Freeman, New York.
790 60. **Gomez H, Polyak I, Thiel W, Lluch JM, Masgrau L.** 2012. Retaining
791 glycosyltransferase mechanism studied by QM/MM methods: lipopolysaccharyl-
792 alpha-1,4-galactosyltransferase C transfers alpha-galactose via an oxocarbenium
793 ion-like transition state. J Am Chem Soc **134:**4743-4752.
794 61. **Chua TK, Bujnicki JM, Tan TC, Huynh F, Patel BK, Sivaraman J.** 2008. The
795 structure of sucrose phosphate synthase from *Halothermothrix orenii* reveals its
796 mechanism of action and binding mode. Plant Cell **20:**1059-1072.
797 62. **Zheng L, Zhou X, Zhang H, Ji X, Li L, Huang L, Bai L, Zhang H.** 2012.
798 Structural and functional analysis of validoxylamine A 7'-phosphate synthase
799 ValL involved in validamycin A biosynthesis. PLoS One **7:**e32033.
800 63. **Cavalier MC, Yim YS, Asamizu S, Neau D, Almabruk KH, Mahmud T, Lee**
801 **YH.** 2012. Mechanistic insights into validoxylamine A 7'-phosphate synthesis by
802 VldE using the structure of the entire product complex. PLoS One **7:**e44934.
803 64. **Kolman MA, Nishi CN, Perez-Cenci M, Salerno GL.** 2015. Sucrose in
804 cyanobacteria: from a salt-response molecule to play a key role in nitrogen
805 fixation. Life (Basel) **5:**102-126.
806
807

808 **FIGURE LEGENDS**

809 FIGURE 1. **The crystal structure of the sucrose synthase from *N. europaea*.** A.
810 Tetrameric structure of the enzyme. B. Monomeric structure and its different domains:
811 SSN-1, SSN-2, GT-B(A), and GT-B(D), and a linker between SSN-1 and SSN-2.
812
813 FIGURE 2. **Phylogenetic tree of sucrose synthases.** Group I contains sequences from
814 cyanobacteria and is divided in two major branches, susA (cyan) and susB (orange)
815 proteins; groups II and III (pink) contain sequences from proteobacteria; group IV
816 contains the sequences from the moss *P. patens* (violet); and groups V, VI, and VII,
817 which contain the sequences from vascular plants are further divided in two branches,
818 dicots (green) and monocots (blue). Numbers for major branches are the bootstrap values
819 obtained during tree reconstruction, as described in the "Materials and Methods" section.
820 *Neu*, *N. europaea*; *Ath*, *A. thaliana* sucrose synthase 1; *Nos*, *Nostoc* sp. PCC 7120 susA;
821 *Tel*, *T. elongatus*.
822
823 FIGURE 3. **Comparison of the open and closed monomeric forms.** The open form
824 structure is represented by the *N. europaea* sucrose synthase structure reported in this
825 paper; the closed form is represented by the *A. thaliana* enzyme (PDB ID 3S29). The
826 SNN-1, SNN-2 and GT-B(A) domains are shown in blue for the open form structure and
827 in cyan for the closed form structure. The GT-B(D) domain is shown in magenta for the
828 open form and in green for the closed form. The "hinge-latch" features of the domain
829 movement are shown in blown-up views.
830

FIGURE 4. **Difference distance matrix map of the GT-B(A) domain.** Distances were calculated between all pair of Cα carbon of the open structure (*N. europaea* sucrose synthase). A second pairwise distance matrix was calculated for the closed structure (homology model as described in "Materials and Methods"). Afterwards, these two matrices were subtracted, and the Δdistance was color coded. The negative and zero values are represented in white. Red colors (higher Δdistance values) are pairs of Cα carbon that are getting closer upon closing of the enzyme. Only residues from 260 to 510 are shown, which correspond to the GT-B(A) domain.

FIGURE 5. **Overlap comparison of the fructose binding sites of the open (*N. europaea*) and closed (*A. thaliana*, PDB ID 3S29) sucrose synthase structures.** The carbon atoms in the closed form structure are in pale yellow. The carbon atoms in the open form structure are in cyan (GT-B(A) domain) and pink (GT-B(D) domain). Conserved residues between two structures are labeled with respective residue numbers; the residue numbers of the open form structure are in parenthesis. The hydrogen bonds in the GT-B(D) domain are shown in black. The hydrogen bonds in between GT-B(A) and GT-B(D) domains and the ones in GT-B(A) domains are shown in green.

FIGURE 6. **Overlap comparison of the nucleotide binding sites of the open (*N. europaea*) and closed (*A. thaliana*, PDB ID 3S29) sucrose synthase structures.** The carbon atoms in the closed form structure are in pale yellow. The carbon atoms in the open form structure are in cyan (GT-B(A) domain) and pink (GT-B(D) domain). Conserved residues between two structures are labeled with respective residue numbers;

Journal of Bacteriology

the residue numbers of the open form structure are in parenthesis. The asterisks indicate the non-conserved binding residues with the closed form residues labeled in front of the ones in the open form structure.

FIGURE 7. **Modeling of the ADP-Glc binding site.** Panel A shows the GT-B(D) domain of the *N. europaea* sucrose synthase, in which ADP has been modeled with Modeller. For that purpose, the closed structure of glycogen synthase with ADP bound (PDB code: 2QZS) was manually aligned to the closed structures of *A. thaliana*, (PDB code: 3S27 and 3S28) and the structure from *N. europaea* (this paper). All those alignments were used as templates. Loops that did not structurally align well were not used for the modelling and the backbone structure was inherited from the *A. thaliana* structures. The rest of the modelling and validation proceeded as described in "Materials and Methods". Panel B shows the GT-B(D) domain from *A. thaliana* (PDB code: 3S27) and the UDP bound.

FIGURE 8. **Hinge analysis by comparison of the open v. close conformations**. The blue and red show the "hinge score" using 51 and 9 windows, respectively. The magmata dots (also point with black arrows) shows the two distinct hinges: G514 and W743. The purple arrows points at the region displaying the secondary structure rearrangements.

FIGURE 9. **Hydrophobic residues contribute to the latch action.** Panel A depicts the open (*N. europaea* sucrose synthase crystal structure) and B a homology model of a closed structure, which was built as described in the "Materials and Methods" section.

Upon closing, the hydrophobic residues: M635, L637 in the GT-B (D) domain and N280, V281, L282 and L284 in GT-B(A) domain generate a hydrophobic environment that stabilize the close action.

FIGURE 10. **Three highly conserved catalytic residues in different members of the retaining GT-B glycosyltransferase family.** The structures analysed in this figure are maltodextrin phosphorylase (PDB ID 1E4O), trehalose-6-phosphate synthase (PDB ID 1GZ5), and glycogen synthase (PDB ID 2ZQS) from *E. coli*, sucrose synthase from *A. thaliana* (PDB ID 3S29), and *N. europaea* (this work).

TABLE 1. Kinetic parameters of substrates of the *N. europaea* sucrose synthase in the synthesis direction. Assays were performed using the conditions described in the "Materials and Methods" section. Analogous values to catalytic efficiency ($k_{cat}/S_{0.5}$) were calculated using the predicted molecular mass of 93 kDa.

Substrate	$S_{0.5}$ (mM)	V_{max} (U mg^{-1})	n_H	$k_{cat}/S_{0.5}$ (mM^{-1} s^{-1})
UDP-Glc	0.89 ± 0.05	4.3 ± 0.1	1.1	7.5
ADP-Glc	0.044 ± 0.006	3.7 ± 0.1	1.3	130.3
$Fru_{(UDP\text{-}Glc)}$	120 ± 10	2.8 ± 0.2	1.3	0.036
$Fru_{(ADP\text{-}Glc)}$	5.6 ± 0.4	4.8 ± 0.2	1.6	1.33

TABLE 2. Data collection and refinement statistics.

Data Processing		
Space group	P65	
Cell dimensions		
a; b; c (Å)	236.90; 236.90; 231.44	
α; β; γ (°)	90.00; 90.00; 120.00	
Resolution (Å)	3.05	
Mosaicity (°)	0.47	
[a]R_{merge}	0.169 (0.963)	
$CC_{1/2}$	0.993(0.561)	0.993(0.561)
I/sigma	9.6 (2.1)	
Completeness	97.9 (98.6)	
Multiplicity	6.3 (6.3)	
Refinement		
Resolution (Å)	3.05	
No. reflections	794715	
No. unique reflections	126170	
[b]R_{work}/[c]R_{free}	17.37/21.75	
[d]RMSD Bond length (Å)	0.009	
RMSD Bond angle (°)	1.439	

The values for the highest resolution bin are in parentheses.

[a]Linear $R_{merge} = \Sigma|I_{obs}-I_{avg}|/ \Sigma I_{avg}$

[b]$R = \Sigma|F_{obs}-F_{calc}|/ \Sigma F_{obs}$.

[c]Five percent of the reflection data were selected at random as a test set and only these data were used to calculate R_{free}.

[d]RMSD, root mean square deviation.

Journal of Bacteriology

903 **TABLE 3. Activity of wild type and mutants of the *N. europaea* sucrose synthase.**

904 Assays were performed using the conditions described in the "Materials and Methods"

905 section.

Substrate	V_{max} (U mg^{-1})			
	WT	R567A	K572A	E663A
UDP-Glc	4.3 ± 0.1	< 0.0017	< 0.0019	< 0.01
ADP-Glc	3.7 ± 0.1	< 0.0014	< 0.0016	0.020 ± 0.02

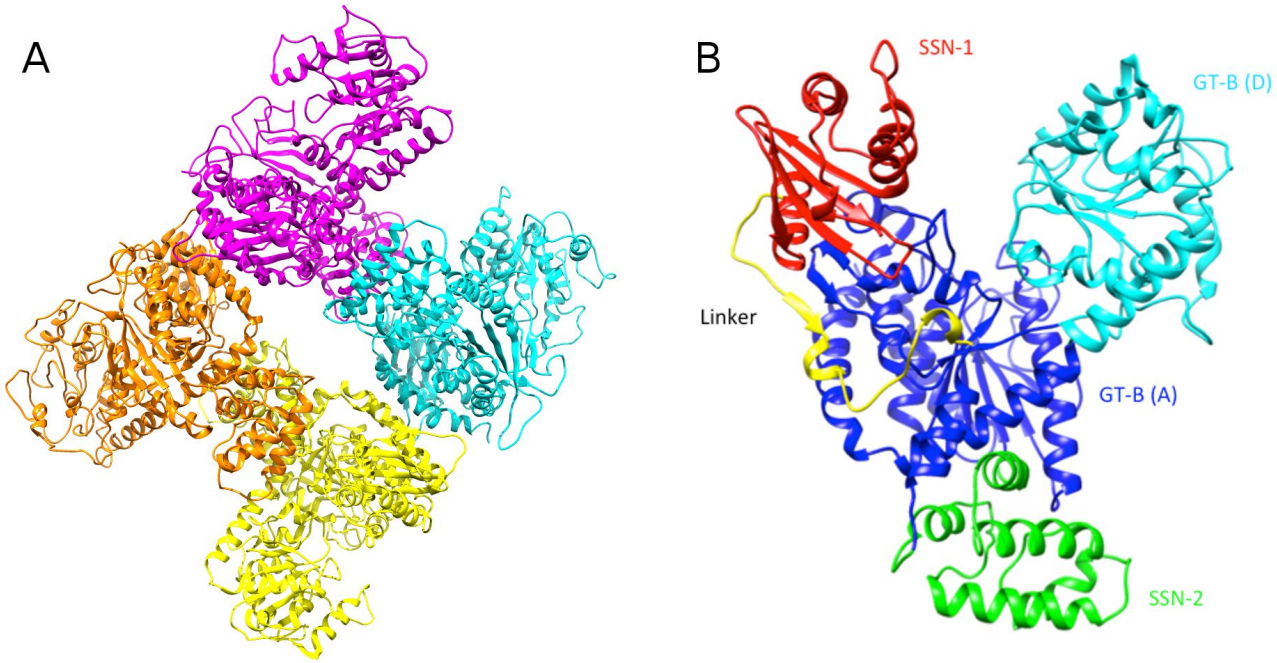
906

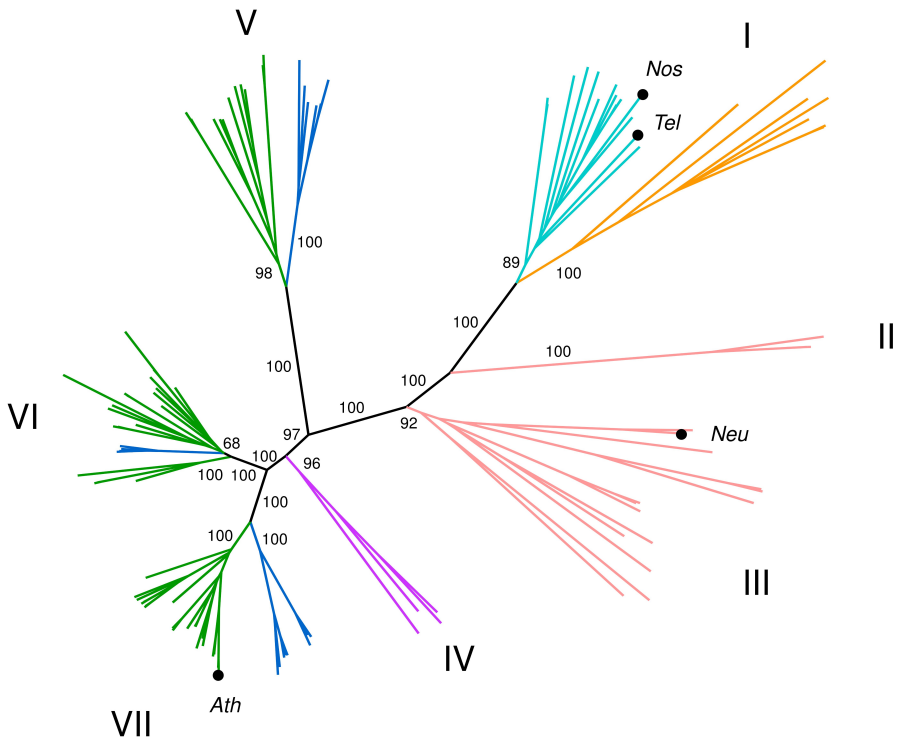
907

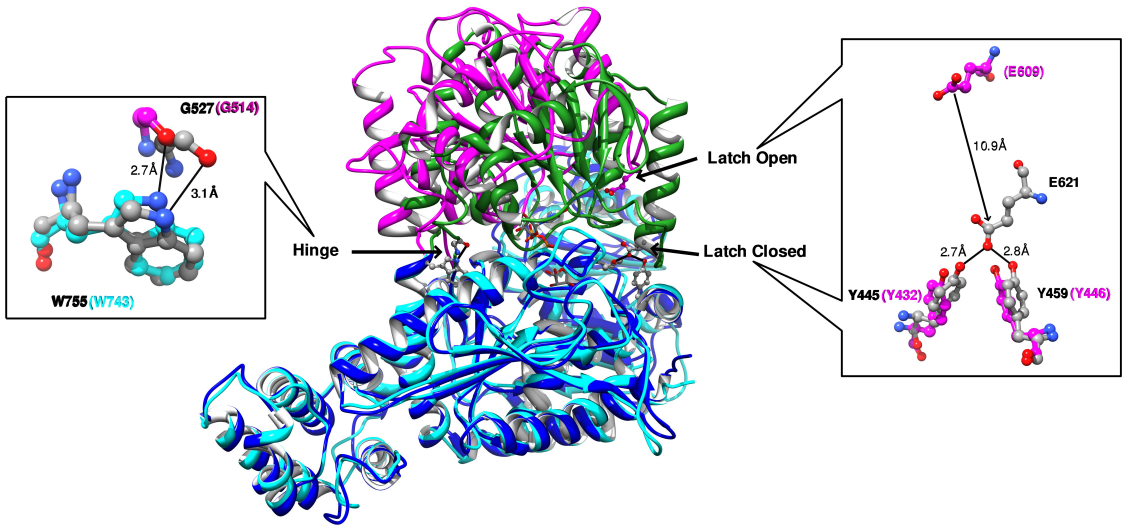
908

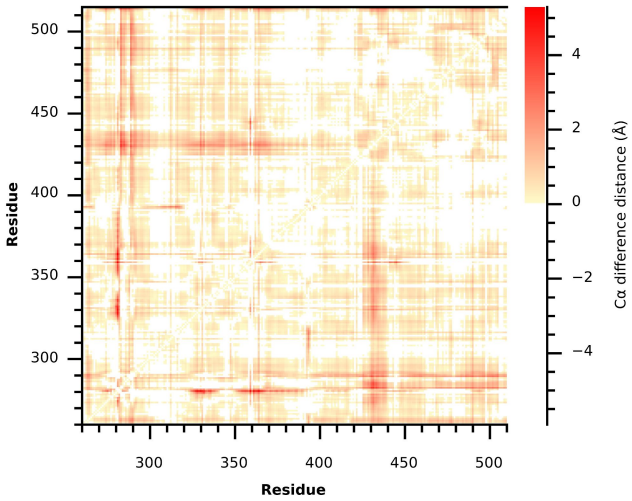
909

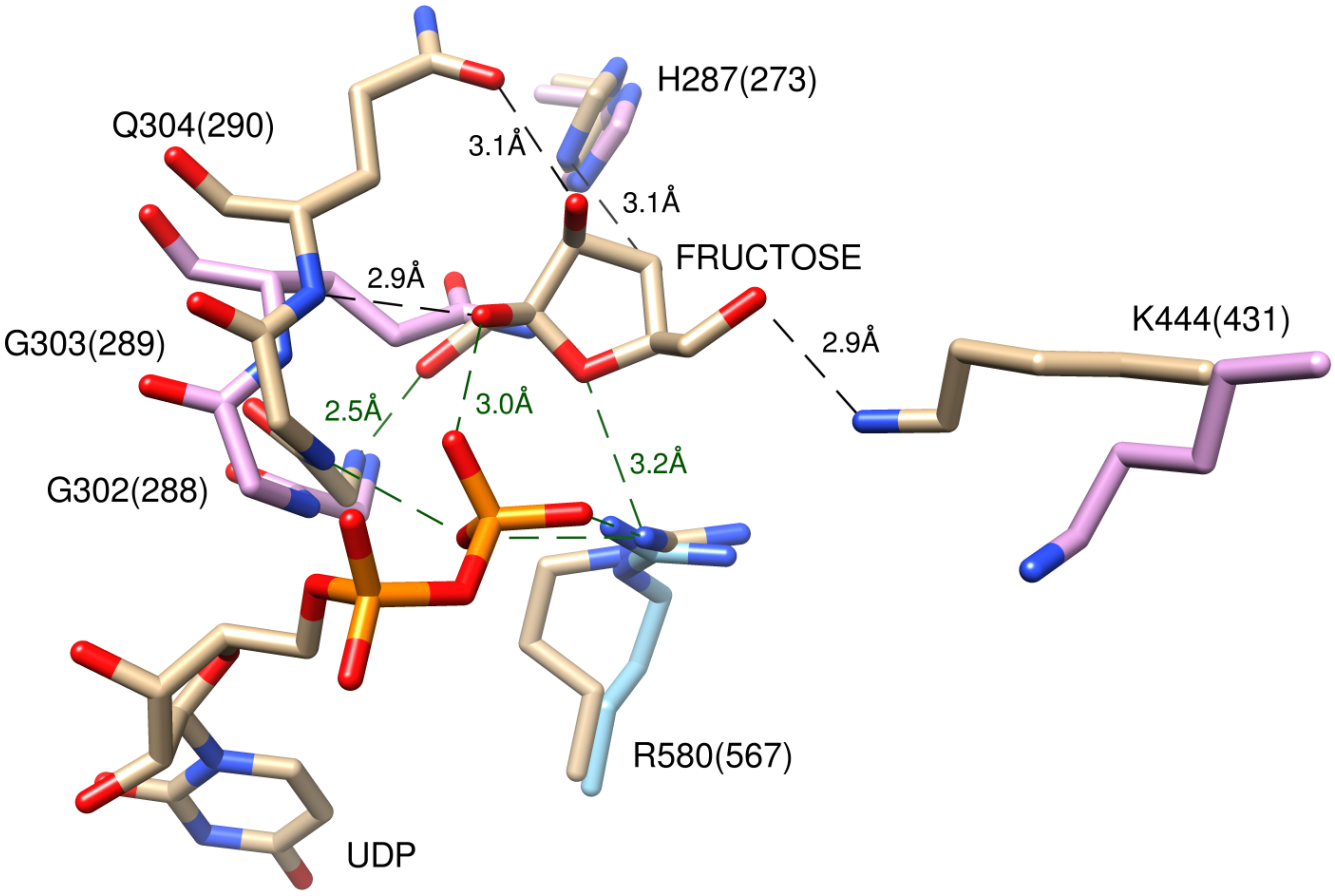
Journal of Bacteriology

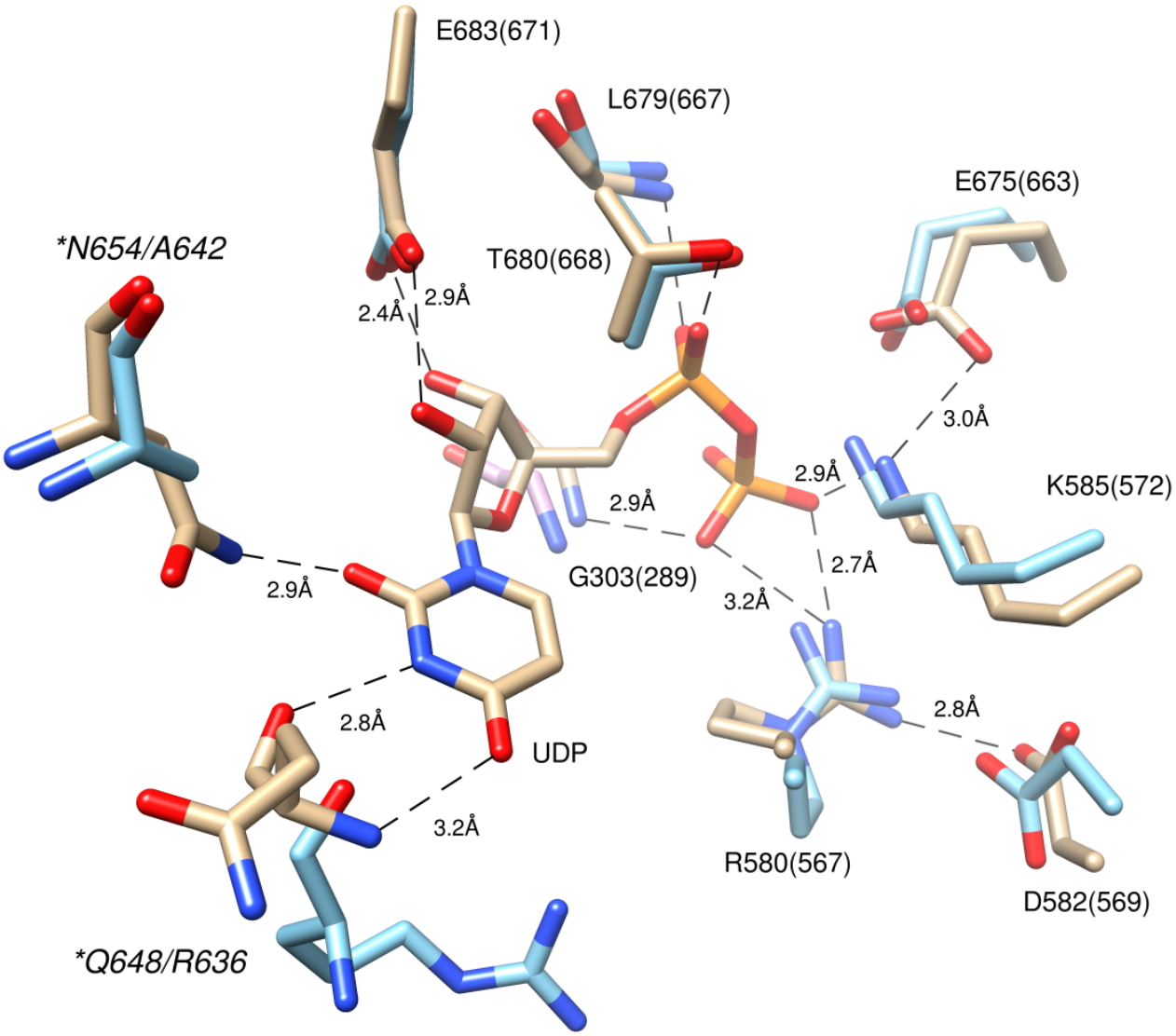


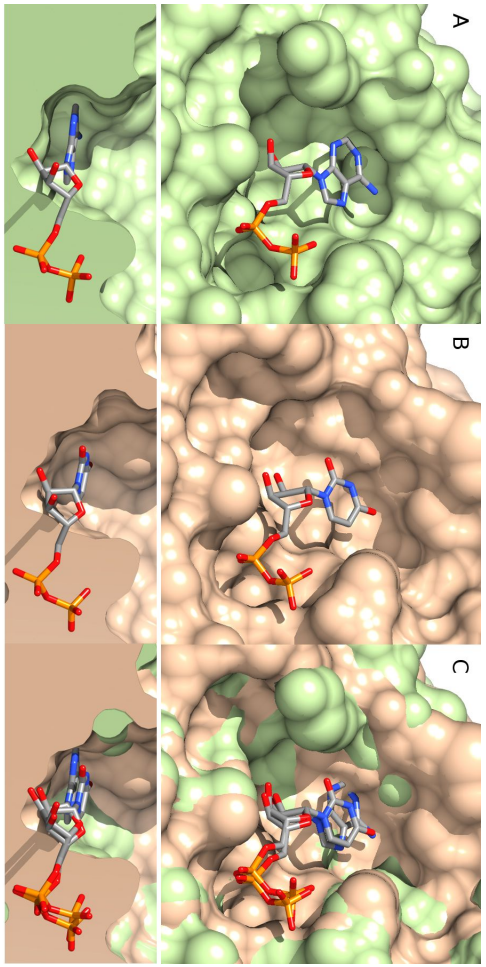


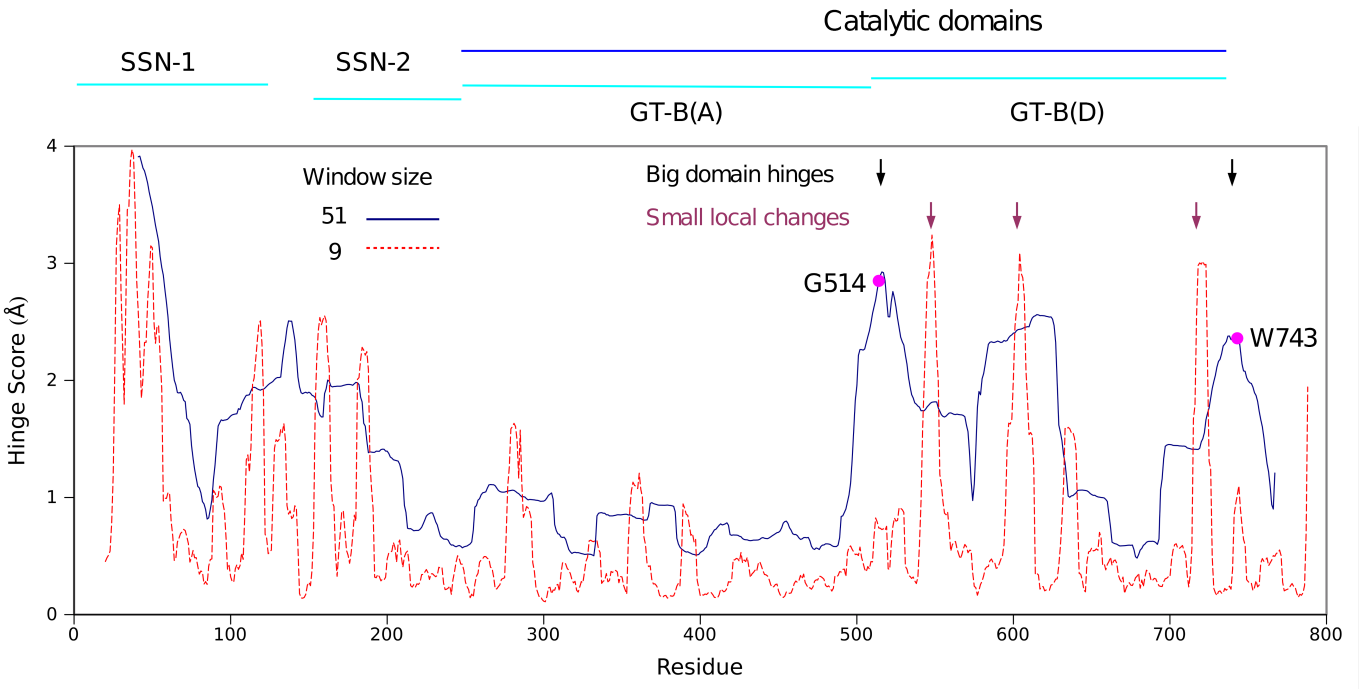









A
B
C



Journal of Bacteriology