In Silico Analysis of the Dextransucrase Obtained From *Leuconostoc mesenteroides* Strain IBUN 91.2.98

Bioinformatics and Biology Insights
Volume 17: 1–10

DOI: 10.1177/11779322231212751

Luisa Alejandra García Galindo[1], Martha Margarita González[1], Jairo Alonso Cerón Salamanca[2] and Sonia Amparo Ospina Sánchez[3]

[1]PhD Student in Biotechnology, Universidad Nacional de Colombia, Bogota, Colombia. [2]Instituto de Biotecnología, Universidad Nacional de Colombia—Sede Bogotá, Bogota, Colombia.
[3]Departamento de Farmacia, Universidad Nacional de Colombia—Sede Bogota, Bogota, Colombia.

ABSTRACT: The DSR-IBUN dextransucrase produced by *Leuconostoc mesenteroides* strain IBUN 91.2.98 has a short production time (4.5 hours), an enzymatic activity of 24.8 U/mL, and a specific activity of purified enzyme 2 times higher (331.6 U/mg) than that reported for similar enzymes. The aim of this study was to generate a structural model that, from an in silico approach, allows a better understanding, from the structural point of view, of the activity obtained by the enzyme of interest, which is key to continue with its study and industry application. For this, we translated the nucleotide sequence of the *dsr_IBUN* gene. With the primary structure of DSR-IBUN, the in silico prediction of physicochemical parameters, the possible subcellular localization, the presence of signal peptide, and the location of domains and functional and structural motifs of the protein were established. Subsequently, its secondary and tertiary structure were predicted and a homology model of the dextransucrase under study was constructed using Swiss-Model, performing careful template selection. The values obtained for the model, Global Model Quality Estimation (0.63), Quality Mean (−1.49), and root-mean-square deviation (0.09), allow us to affirm that the model for the enzyme dextransucrase DSR-IBUN is of adequate quality and can be used as a source of information for this protein.

KEYWORDS: Structural bioinformatics, dextran, homology modeling

RECEIVED: April 25, 2023. **ACCEPTED:** October 21, 2023.

TYPE: Original Research Article

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was partially supported by a grant from Universidad Nacional de Colombia, Research Vice rectory (Grant number 47514) and a Grant for PhD student of the Science, Technology and Innovation Ministry. Besides the project is submerged in the Access to genetic resources contract No. 271 (from the Colombian Environment Ministry which covers the project "Obtaining and utilization of enzymatic activities of microbial origin for the production and utilization of biopolymers").

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Sonia Amparo Ospina Sánchez, Departamento de Farmacia, Universidad Nacional de Colombia—Sede Bogota, Bogota, Colombia. Email: saospinas@unal.edu.co

Introduction

The dextransucrase enzyme (DSR, EC 2.4.1.5) catalyzes the transfer of α-D-glucose units from sucrose to low-molecular-weight acceptors such as other sucrose molecules, primary alcohols, or a growing polymer, called glucan,[1,2] to synthesize oligo and polysaccharides while releasing fructose to the medium.[3,4] The main product of this reaction is a biopolymer of great economic importance called dextran, with a wide range of applications (1) in the food industry as an additive, viscosifier, and reducer of caloric content; (2) in the pharmaceutical industry as a volume expander of the blood plasma and promoter of blood flow; and (3) in the area of medicine due to the promotion of intestinal health due to its anti-inflammatory activity.[5-7]

The biopolymers and biofunctionals research group obtained a high-molecular-weight dextran-type biopolymer from the native strain *Leuconostoc mesenteroides* IBUN 91.2.98 culture, with promising properties for its use as a prebiotic and soluble fiber source in animals and humans. An enzyme responsible of dextran synthesis was then obtained by Flórez-Guzmán.[8] This enzyme, called DSR-IBUN, was purified and characterized. It was found that, under used experimental conditions, DSR-IBUN had a short production time (4.5 hours) and an enzymatic activity of 24.8 U/mL, which is higher than those reported for *L mesenteroides* NRRL B-512F,[9] *L mesenteroides* B512 FMC,[10] *L mesenteroides* NRRL B-1299,[11] and the alternansucrase of *L citreum* NRRL B-1355,[12] with values of 5.85, 6.33, 0.58, and 0.66 U/mL, respectively, as well as a specific activity up to 2 times higher (331.6 U/mg) than that observed to date for other strains such as *L mesenteroides* B-512FMC,[13] *L mesenteroides* 0326,[14] and *L mesenteroides* B-51[13] with values of 183, 72.54, and 53 U/mg, respectively. Although several research groups have researched about DSR-IBUN and the potential uses of the polymer,[8] the sequencing of the gene encoding dextransucrase had not been done nor had proposed a structural model that, from an in silico approach, allows a better understanding of protein activity.

Due to high activity of this enzyme, it is important to know the activity dynamics and the enzymatic kinetics from a structural point of view. In particular, dextransucrase DSR-IBUN and other large membrane protein structures are difficult to solve even with methods such as x-ray crystallography, nuclear magnetic resonance (NMR), or Cryo-EM.[15] However, the understanding of protein function generally requires understanding its 3-dimensional (3D) structure, to uncover biological process involved, and also to perform important modification that may lead to important activity dynamic changes.[15,16] A viable alternative approach is to predict the 3D structure in silico using homology modeling or comparative modeling,[17] in which the 3D model of a protein of interest (target) can be built from one or several related proteins of experimentally determined structure (template) that share a statistically global similarity in its sequence.[18-20]

The aim of this work was to generate a 3D model for the dextransucrase DSR-IBUN of *L mesenteroides* IBUN 91.2.98 based on gene sequence that encodes it, to study possible modifications that could imply changes that allow the production of polymers with different characteristics that could be used for different applications.

Methodology

Sequencing of dsr_IBUN *gene*

Applying Sambrook and Russell[21] criteria, the following primers were designed for amplification of the interest gene: dsr_IBUN_dir_HindIII: 5′-AAGCTTATGCCATTTACAGAAAAAAGT-3′ and dsr_IBUN_Rev_XbaI: 5′-TCTAGAAGAAAACTTATGCTGACACA-3′. A polymerase chain reaction (PCR) was performed using high-fidelity enzyme Phusion DNA polymerase (New England biolabs) (0.5 U), buffer (1X), $MgCl_2$ (1.5mM), dNTP mix (200 µM), direct primer (0.5 µM), reverse primer (0.5 µM), template DNA (150 ng), and dimethyl sulfoxide (3%). The PCR conditions were 98 °C for 30 seconds and 30 cycles of 98 °C for 5 seconds, 54 °C for 30 seconds, 72 °C for 2.5 minutes, and a final extension at 72 °C for 8 minutes. The amplified PCR product was analyzed by electrophoresis on a 1% agarose gel at 80 V. The product was purified, cloned in the vector pTOP TA (Takara) and sequenced by Macrogen. Gene sequence is available in GenBank under the accession number OQ731663.1.

DSR-IBUN sequence analysis

To obtain primary protein structure of DSR-IBUN dextransucrase, translation of the nucleotide sequence of *dsr_IBUN* gene was performed using Biomodel application.[22] Features such as protein molecular weight, theoretical Ip, amino acid composition, atomic composition, estimated half-life, instability index (II), aliphatic index, and average hydrophobicity index (GRAVY for Grand Average of Hydropathy) were estimated using the Expasy ProtParam tool.[23] Subcellular localization prediction was performed with the CELLO v2.5 server.[24,25] Signal peptide sequence was established using SignalP 5.0 server (www.cbs.dtu.dk/services/SignalP).[26] Subsequently, the amino acid sequence obtained was compared with those reported in the NCBI using the BLAST algorithm (https://www.ncbi.nlm.nih.gov/), and phylogenetic tree was constructed with Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/).[27]

To establish the presence of conserved functional domains and motifs in deduced DSR-IBUN sequence, Motif Scan program was used (https://myhits.sib.swiss/). Structural motifs were determined by sequence alignment with several GH70 family proteins using the algorithm for multiple sequence alignments MAFF (https://www.ebi.ac.uk/Tools/msa/clustalo/). To visualize these alignments, Jalview software was used (https://www.jalview.org/).[28]

Modeling of the DSR-IBUN

Primary DSR-IBUN structure was used to predict its secondary and tertiary structure to build a model. Homology with similar proteins, for which its 3D structure had been determined experimentally, was determined using 4 steps methodology proposed by Quinn[29]: (1) sequence preparation, (2) modeling, (3) validation, and (4) application.

Initial prediction of DSR-IBUN secondary structure was performed with SOPMA server (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.htm).[30] The amino acid sequence of DSR-IBUN was compared with those available in the Protein Data Bank (PDB) using Swiss-Model (https://swissmodel.expasy.org/),[31] Phyre[2] (http://www.sbg.bio.ic.ac.uk/phyre2/),[32] I-TASSER (https://zhanglab.ccmb.med.umich.edu/I-TASSER/),[33] and the NCBI servers (https://www.ncbi.nlm.nih.gov/) allowing comparison between them to select those structures with a higher identity percentage and greater coverage as possible candidates to be used as templates for structure modeling.

To select template structure, quality of preselected crystal structures was addressed, in terms of their resolution and *R* values, as these indicate the quality obtained from crystallographic data. Metrics structure summary graphs validation reports were also considered, in which the overall quality of each structure was compared with the ranges of structures previously reported in the PDB.

Once template structure was selected, Swiss-Model, Phyre,[2] and I-TASSER servers were used to corroborate secondary structure prediction and structural domains distribution, as well as to validate the different models proposed for DSR-IBUN. Protein structure models were compared using Ramachandran plots data, concerning to energetically allowed regions for the dihedral angles of models' skeletons, as well as the obtained GMQE (Global Model Quality Estimation) and QMEANDisCo Global (Quality Mean) values. With these data, a 3D in silico model structure for DSR-IBUN was selected.

Interactions in the catalytic domain of DSR-IBUN proposed model

Active site catalytic triad (Asp, Glu, Asp) previously reported by other authors was identified using Pymol.[34-37] The validated in silico structure was used to establish the location of the active site and the catalytic triad involved in the function of the enzyme by superimposing template structure used above and structural DSR-IBUN protein model using the Pymol software.[38] Interacting amino acids of DSR-IBUN validated model and sucrose molecule were analyzed via Discovery Studio Visualizer software.[39]

Molecular docking between isomaltose and DSR-IBUN was directed to glucan-binding pockets as the interaction sites using the HADDOCK 2.4 platform (https://wenmr.science.uu.nl/haddock2.4/).[40]

Table 1. Predicted physicochemical properties for the DSR-IBUN protein.

PARAMETER	WORTH
Half-life (mammalian reticulocytes in vitro)	30 PM
Half-life in yeast in vivo	>20 hours
Half-life in *Escherichia coli* in vivo	>10 AM.
Instability index (II)	22.16
Aliphatic index	68.22
GRAVY (Grand Average of Hydropathy)	−0.599

Source: Protparam-obtained data.

Results and Discussion

DSR-IBUN primary and secondary sequence analysis

Translation of obtained nucleotide sequence for *dsr_IBUN* gene resulted in a 1527 residues protein. Its estimated molecular weight and isoelectric point (Ip) were 169.8 kDa and 4.29, respectively. Data are consistent with those reported experimentally by Flórez-Guzmán et al,[8] in an electrophoretic analysis performed to DSR-IBUN, which showed a molecular mass of 170.1 kDa and a Ip of 4.2. On the contrary, the predicted molecular formula for this dextransucrase was $C_{7487}H_{11242}N_{1966}O_{2519}S_{21}$.

Most abundant amino acid is aspartate (10% of the residues found), whereas the least abundant is cysteine with only 0.1% (1 residue). Total number of negatively charged residues (Asp+Glu: 217) was greater than those observed for positively charged residues (Arg+Lys: 101). Although amino acids such as Asp and Gly do not appear preferentially in extracellular or intracellular proteins, it has been observed that polar amino acids such as Thr, Asn, Gln, and Ser are characteristic of extracellular enzymes,[41] which coincides with observed for DSR-IBUN, in which these 4 amino acids are abundant, representing 30.6% of total residues. Even though extracellular eukaryotic proteins frequently have a high percentage of cysteines, prokaryotic proteins rarely present disulfide bridges, even in the case of extracellular ones.[41] These results agree with the prediction of the subcellular localization done, which indicates that DSR-IBUN is an extracellular protein which presents a fraction anchored to cell wall and membrane, a characteristic that is confirmed with subsequent bioinformatic analysis.

Estimated DSR-IBUN half-life values in different biological media such as mammals, yeast, and *Escherichia coli* (in vivo and in vitro) for all cases were greater than 10 hours, indicating that it is a stable protein (Table 1). This stability over time was corroborated by the II, which estimates the protein stability under laboratory conditions[42] and by the aliphatic index, which refers to the relative volume occupied by aliphatic chains (alanine, valine, isoleucine, and leucine).[43] Any protein that, like DSR-IBUN, presents an II value <40[42] and an aliphatic index >50 is considered stable,[43] as aliphatic chains provide heat resistance and therefore stability in different conditions.[44] As the value of GRAVY determined for the studied dextransucrase was negative (−0.599), which indicates that it is a hydrophilic protein,[45] this is consistent with the proportion of polar groups present in the side chains located on the protein surface, which promotes solubility.[43]

Using SignalP 5.0 server, a typical secretion signal peptide sequence in the N-terminal region was predicted with a peptidase cleavage site between amino acids at positions 42 and 43 (**G-D**), which is consistent with the reported data for other DSR such as DexYG[33] and DsrD[46] and supports the already observed fact of dextransucrase secretion of from *L mesenteroides* IBUN 91.2.98 to growth media.[47]

DSR-IBUN showed 100% coverage and an identity percentage greater than 98% compared with other *L mesenteroides* strain sequences, such as DsrN from KIBGE-IB-22,[48,49] DexYG from *L mesenteroides* 0326,[50] DEX protein from strain NRRL B-512F,[51] and DsrD from the Lcc4 strain.[46] Compared with other proteins, dsrV (AHC31982.1) from *L citreum* M-3[52] displayed 100% coverage and 94.31% identity percentage, whereas sequences such as dsrb742 (AAG38021.1) from *L mesenteroides* B-742CB[53] and dsrB from *L mesenteroides* NRRL B-1299 (AAB95453.1)[54] showed a 99% coverage and 66.25% and 66.45% identity percentages, respectively. Lower identity percentages were observed comparing with strains of other species such as *Lactococcus lactis* (around 95% coverage and 60% of identity percentage). Phylogenetic tree constructed using 14 sequences of dextransucrases deposited in NCBI database showed that DsrD is the closest to DSR-IBUN protein as Flórez-Guzmán et al[8] reported. Phylogenetic analysis also showed that glucansucrases (GS) of the same species grouped together; as Martínez et al[55] mentioned, in its simplest form, the genetic diversity that drives adaptation is based on the accumulation of point mutations that are established in the same species.

By comparing the deduced amino acid sequence for DSR-IBUN with the databases available information, it was possible to corroborate the presence of a functional domain conserved in GH70 family, as well as to observe the functional domain of α-amylase super family, to which dextransucrase belongs, which is organized under a structure (α/β).[8,14,34,54,56-60] In addition, repeated motifs were observed both at the start and at the end of the sequence related to the binding site to cell wall binding (cell wall motifs), which agrees with that mentioned by Vuillemin et al,[61] who point out that it has been determined that most of the active GH70 enzymes are strongly bound to the cell wall, making their purification difficult.

Likewise, repeated motifs related to sugar-binding sites were found at the N and C terminus of the protein. That is why domain V is known as the glucan-binding domain (GBD), as Claverie et al[59] reported that these motifs are made up of "pockets" in which the glucoses can interact. Importance of this

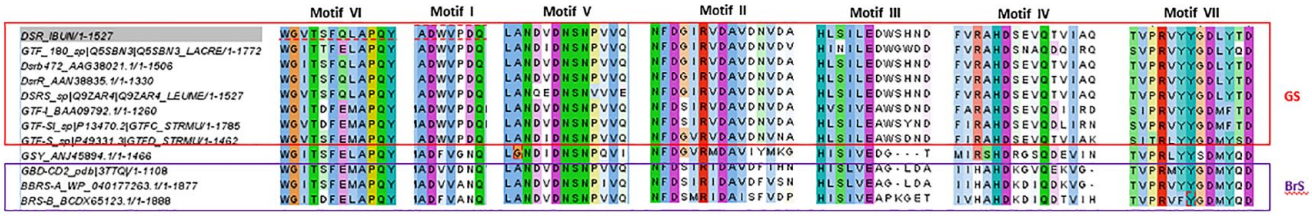

Figure 1. Structural motifs conserved in DSR-IBUN.

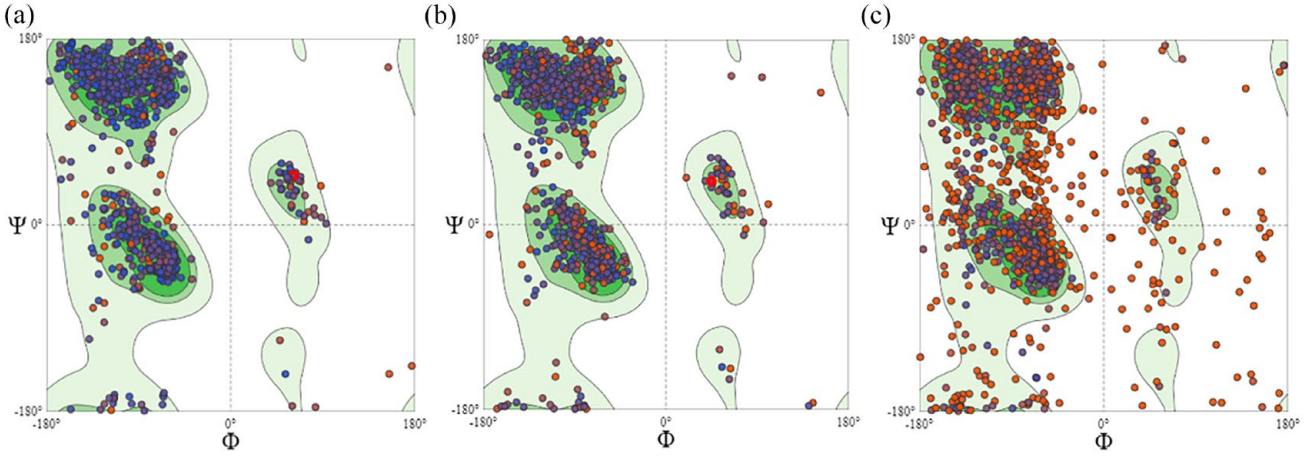

Figure 2. Ramachandran plots for DSR-IBUN models, obtained by (A) Swiss-Model, (B) Phyre², and (C) I-TASSER platforms.

domain has been suggested to be related to the polymer size and, therefore, the enzyme activity, which varies between dextransucrases.[14]

The 7 structural motifs (I to VII) of the GH70 indicated by Yan et al,[62] were identified in the DSR-IBUN primary sequence that has 22 strictly conserved residues in this family. Vuillemin et al[61] and Venkatachalam et al[63] have shown that GH70 enzymes can be grouped into 2 subfamilies, (1) GS and (2) branching enzymes or branching sucrase enzymes, sharing the enzyme called Gys as an evolutionary intermediate between them. In the alignment (Figure 1), it can be observed that DSR-IBUN is part of the GS subfamily, which indicates that the synthesized dextran has mainly α1,6 bonds, accompanied by α1,3 bonds.[60]

DSR-IBUN secondary structure prediction showed 39.03% random coils, 28.55% β-sheets, 20.89% α-helices, and 11.53% β-turns, which is comparable with what was obtained by Du et al[64] for *L mesenteroides* DRP105 dextransucrase.

DSR-IBUN prediction structure

When comparing DSR-IBUN amino acid sequence with the experimental data of solved structures reported in the PDB using 4 different platforms, it was found that, in all of them, the first 5 suggested templates correspond with the same ones (5NGY, 5LFC, 6HVG, 6SYQ, and 3HZ3), but the order proposed on each platform for the templates was different. Both NCBI and Swiss-Model match in suggesting 5NGY as the highest value for GMQE, followed by 5LFC, while Phyre² and I-TASSER suggested 6HVG as the first option. This is explained by the fact that the first 2 platforms use modeling algorithms in which the sequence prevails over the structure, whereas Phyre² and I-TASSER use threading algorithms, in which structural templates of PDB are first identified and later the models are built by iterative simulations that assemble or weave the fragments.[65]

Models that showed the highest quality indices on each platform were evaluated to compare them and select the most appropriated one. Comparison was made using Ramachandran plots (Figure 2). Swiss-Model (2A) showed a Ramachandran favorability of 94.14%, followed by the model generated using Phyre² (2B) with a favorability of 93.35%, while in the model suggested by I-TASSER (2 C), this parameter had only 71.90%. This is observed in the plots where it is clear that the outliers for the evaluated structure in Figure 2C had the highest number of outliers (11.37%), whereas in the proposed models by Phyre² and Swiss-Model, these values did not exceed 2% (1.17% and 1.44%, respectively). Taking these results into account, the Swiss-Model platform was used to create the final model.

Coverage and the identity percentage values were taken as referents for the template structure selection. In all cases, the highest coverage (89%) and identity percentage (51%) were turned up by the resolved structures for the dextransucrase

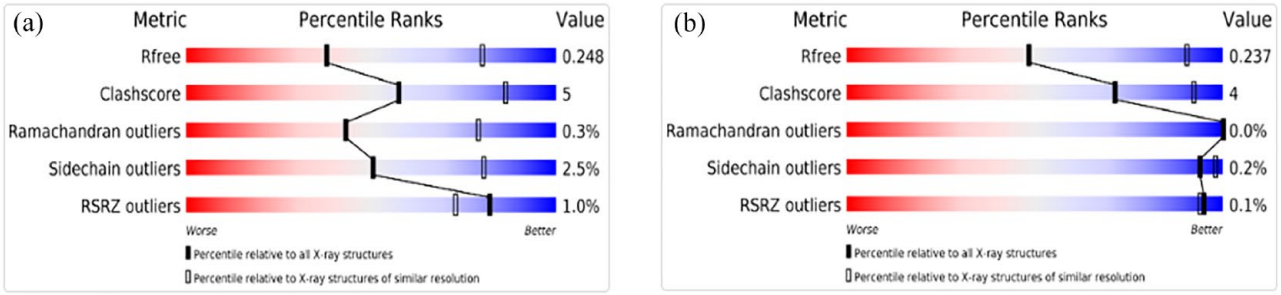

Figure 3. Validation report DSR-M dextransucrase from *Leuconostoc citreum* NRRL B-1299 solved structures. (A) 5LFC structure and (B) 5NGY structure. The values presented in the blue range are considered good, whereas those observed toward the red range are considered of low quality. PDB indicates Protein Data Bank.
Source: Taken from PDB.[19]

DSR-M from *Leuconostoc citreum* and NRRL B-1299 with PDB codes 5LFC and 5NGY, respectively.[58] The DSR sequences reported for *L mesenteroides* such as DSR-E and its mutants had a coverage of 79% and an identity percentage of 44%. Finally, alternansucrase corresponding to the structure 6HVG showed coverage of 73% and presented identity percentage of 51%.

To select the template, structure quality of 5LFC and 5NGY structures was checked. Both structures had adequate resolution values (around 4Å), but *R* values, which indicate the measure of the atomic model quality obtained from the crystallographic data, were better for 5NGY, since the closer to 0 the value of *R* is, the greater the fit of the structure to the model, with a typical value of 0.2.[19] This, together with the summary graphs metrics in which the global quality of each structure (black bars) is compared, with respect to the ranges of similar structures previously deposited in the PDB (white bars), allows us to observe that the quality of 5NGY is superior as it does not present outliers in the Ramachandran plots and fewer outliers in the resulting side chains, which is why it was used as a template for the construction of the model (Figure 3).

Once 5NGY was selected as a template, SOPMA was used to obtain secondary structure prediction, while using Swiss-Model, DSR-IBUN model was constructed and validated. As expected, it corresponds to a U-shaped monomer (Figure 4A), according to that reported for the glycosylhydrolases of the GH-H Clan, which corresponds to the GH70 family.[34,66]

Global Model Quality Estimation value which estimates quality and reflects the expected precision of a model created from the alignment between query sequence and template sequence displayed a value of 0.56, which is considered positive as higher values at 0.5 indicate acceptable overall quality (PDB, nd). The Mean Quality indicator (QMEANDisCo Global) score, based on different geometric properties of the model that provides both global (ie, for the whole structure) and local (ie, by residue) quality estimates and distance evaluation pairwise in a model and a set of constraints derived from experimentally determined protein structures that are homologous to the model being tested, was 0.72 ± 0.005. As scores between 0 and 1 indicate that the model is comparable to what would be expected from experimental structures of similar size, the calculated value is considered adequate.

DSR-IBUN structure model displays a typical organization of dextransucrase consisting of 5 structural domains called A, B, C, IV, and V, with domain C being the only one continuous throughout the structure (Figure 5). The U-shaped folding of DSR-IBUN tallies with reported structures of other GH70 enzymes.[34] Figure 5 shows the organization of the 5 DSR domains: domain A comprising a barrel (α/β)8; domain B (from ß3 to α3); and domains C, IV, and V which form the bottom of the "U."

Figure 5 shows a domain A which consists of 8 α-helices and 8 β-strands alternated with each other, linked by irregular loops of variable sizes that engage the C-terminal ends of the β-strands with the N-terminal ends of the α-helices and the C-terminal ends of the α-helices joined with the N-terminal ends of the β-chains.[34,58] The beta barrel structure (β/α)8 is organized with parallel β strands organized almost parallel to each other, as if they were around a cylinder and located toward within the protein. On the contrary, the 8 α-helices are located toward the outside of the β strands, on the protein surface.[67]

Domain B, located close to the catalytic domain (A), is formed by 5 β-sheets (Figure 5). It has been suggested that this domain is essential for the enzyme function as it provides residues that are part of the protein donor and acceptor sites and some of this domain loops contribute to give the groove shape near the catalytic site.[68] Domain C, also located close to the catalytic domain, has 8 β-sheets and, as seen in the model, joins sections of domains A and B so that they can perform their function.

Domain IV structure reveals a novel folding that is characteristic of GH70 enzymes. Domain IV connects domain B to V, but other functions remain unknown, although it has been proposed that it may provide a "hinge" that would lead domain V to attached glycans near or far from the catalytic site. In fact, the connection between IV and V domains consists of 2 relatively long polypeptide stretches with disordered structure, and the relative position of different dextransucrases V domain

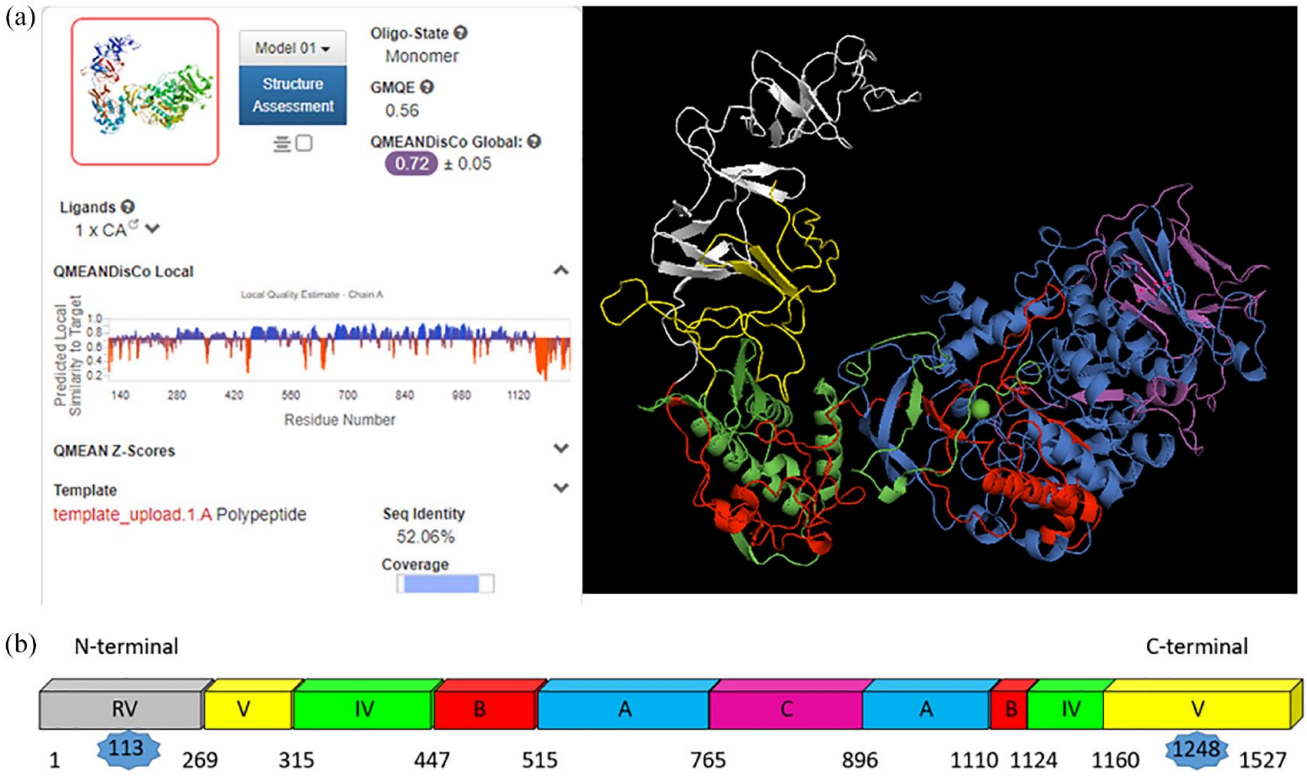

Figure 4. (A) Model of the dextransucrase of *Leuconostoc mesenteroides* IBUN 91.2.98 tertiary structure built and validated with Swiss-Model. (B) Scheme of predicted organizational domains for DSR-IBUN. Residue number at which each domain in the primary structure begins and ends is shown and the initial and final amino acids present in the model are highlighted.

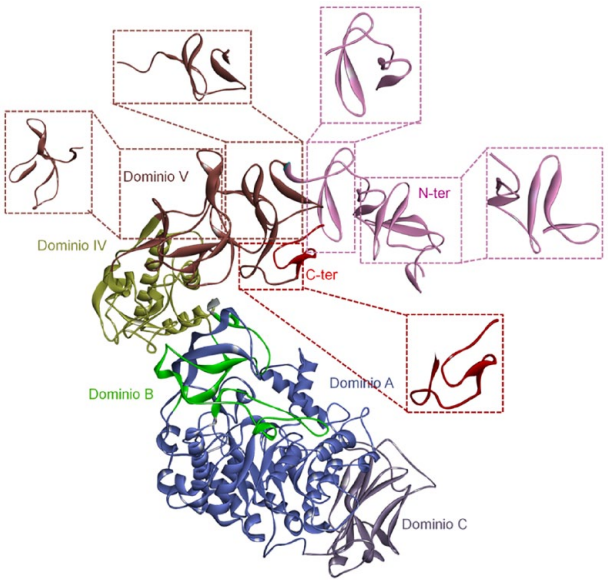

Figure 5. Typical domains of dextransucrase A, B, C, IV, V, C-terminal, and N-terminal. The identified glucan-binding sites for each of the domains are shown.

with respect to the rest of the protein indicates some flexibility in these stretches.[69]

Finally, domain V is made up of the N-terminal and C-terminal segments.[69] This agrees with that reported by Vujicic-Zagar et al,[34] who point out that in family 70 glycosyl-hydrolases, except for the C domain that forms the base of the U-shaped fold, the domains are formed by 2 noncontiguous segments of amino acids from the N-terminal and C-terminal ends, where the amino and carboxy termini meet to form the V domain of the framework. Therefore, it has been proposed that the GH70 enzymes originated from the GH13 enzymes,[70] by an evolutionary pathway based on duplication permutation, as the GH70 enzymes share the domains with the GH13 enzymes (A, B, and C), but in GH70 enzymes, 2 of these domains, as previously discussed, are made up of noncontiguous sections in the structure. Despite these rearrangements, there are several conserved motifs (eg, motifs I-VII) in members of the 2 families, and these conserved motifs provide crucial information for analyzing the origin and evolution of GH70 enzymes.[62]

Domain V, known as GBD, has YG repeat motifs[71] and some pockets that can bind glycan residues which were identified in the model (Figure 5). Several studies suggest that domain V, located at the N-terminus and C-terminus of enzymes, determines (1) the size of the polymer, (2) enzymatic activity, (3) specificity, and (4) processivity. In addition, it has been shown that in some *Streptococcus* and *Leuconostoc* species these amino acids allow the binding of the enzyme to the cell wall.[34,54,59,72,73] However, its precise role is not yet fully understood.[35] Unfortunately, the structures reported for GH70

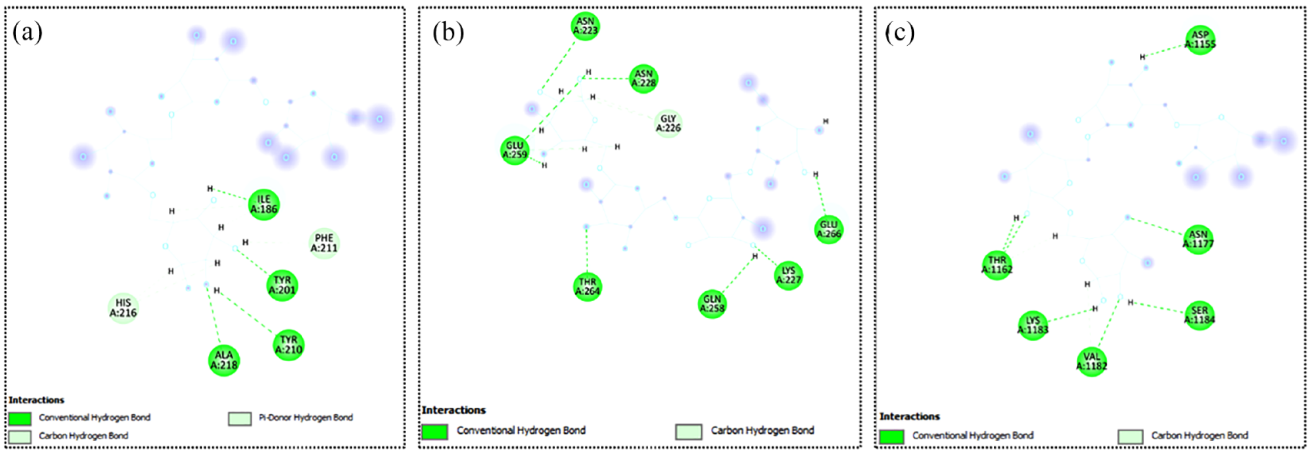

Figure 6. Hydrogen bonds present in glucan-binding pockets with isomaltose: (A) N-terminal domain binding pocket, (B) glucan-binding pocket domain V, and (C) glucan-binding pocket domain IV.

enzymes are truncated and the complete structure is yet unknown.

Some structures showing interacting residues with ligands such as sucrase, glucose, isomaltose, and isomaltriose provide structural evidence of common topology of glucan-binding sites in V-domain repeats.[37] The importance of these repeats in the V domain lies in the fact that deleting some of them produces a progressive loss of activity and a decrease in the size of the polymer.[72] In Figure 6, the hydrogen bond type interactions between isomaltose and some residues of the glucan-binding pockets of the IV, V, and N-terminal domains are shown (Figure 6).

Identification of active site and calcium ion pocket

Taking into account that the data quality obtained for DSR-IBUN model and its validation supported its reliability to be used for subsequent in silico studies, deeper analyses were done using modeled structure.

Localization of the active site and the catalytic triad involved in enzyme function was determined in the structural model, by making an alignment and superimposition between the model and the template. This allowed to determine the root-mean-square deviation (RMSD) value of the atomic positions (RMSD), which is a measure of the average distance between the carbon skeleton atoms of 2 superimposed proteins[29] and the value of which allows us to affirm whether the prediction made for the structural model is adequate with respect to the available information available on the template. For this case, the RMSD was 0.09, which is considered positive, since the closer the value is to 0, it is inferred that the positions in the 3D structure are similar. After reviewing the interaction interface using sucrose as ligand, it was observed that the substrate in its most stable conformation shows interactions with the catalytic triad Asp509, Glu547, and Asp620 established mainly by the formation of hydrogen bonds. It is also possible to find other interactions between substrate and amino acids close to the catalytic site, such as Arg 507, His 619, and Asp 936 (Figure 7).

To complement it was found that the model suggests the presence of a conserved site for the binding of $Ca+^{2}$ ions, interacting with the residues Glu 483, Asp 469, Asn 513, and Asp 986 (Figure 8), it was found that interaction occurs near to nucleophilic aspartate of the active site, which has been shown to be essential for the enzyme catalytic activity.[36]

Calcium dependence has been reported for some GS,[34,37,59] showing a 14% decrease in enzymatic activity in the presence of chelating agents such as EDTA,[37] as well as increases of up to 48% of the activity in presence of these ions.[73] However, this aspect must be carefully considered, as according to what was reported by Flórez-Guzmán[47] for the dextransucrase of *L mesenteroides* IBUN 91.2.98 strain, the calcium ion had a slight inhibitory effect of 8% of this enzyme activity.

With the structural model obtained, it is possible to generate a methodological approach that allows to propose strategies to carry out mutations in the gene that codes for DSR-IBUN, first by establishing that in silico mutations do not have destabilizing effects on the structure of mutant proteins, seeking to guarantee that they are functional in vitro.

Taking into account the benefits of homology modeling applied in the methodology carried out in this study, it is emphasized that the template used had 52.06% identity, which is considered optimum according to what has been proposed by other authors. Comparative modeling of the 3D structure, if the target presents at least 30% to 40% identity with an empirically determined structure, can be applied with reasonable accuracy to 10 times as many protein sequences, relative to the number of experimentally determined protein structures,[74] result in a model with a precision close to that of a structure obtained by low-resolution x-rays or one studied by medium-resolution NMR.[16]

Within the limitations of homology modeling, it is found that similarity values less than 30% decrease the accuracy of comparative models and therefore their applications, mainly because of the intrinsic errors of the alignment algorithms.[75] This can be overcome if the quality parameters of the template and the model are considered, as was done in this work. Errors

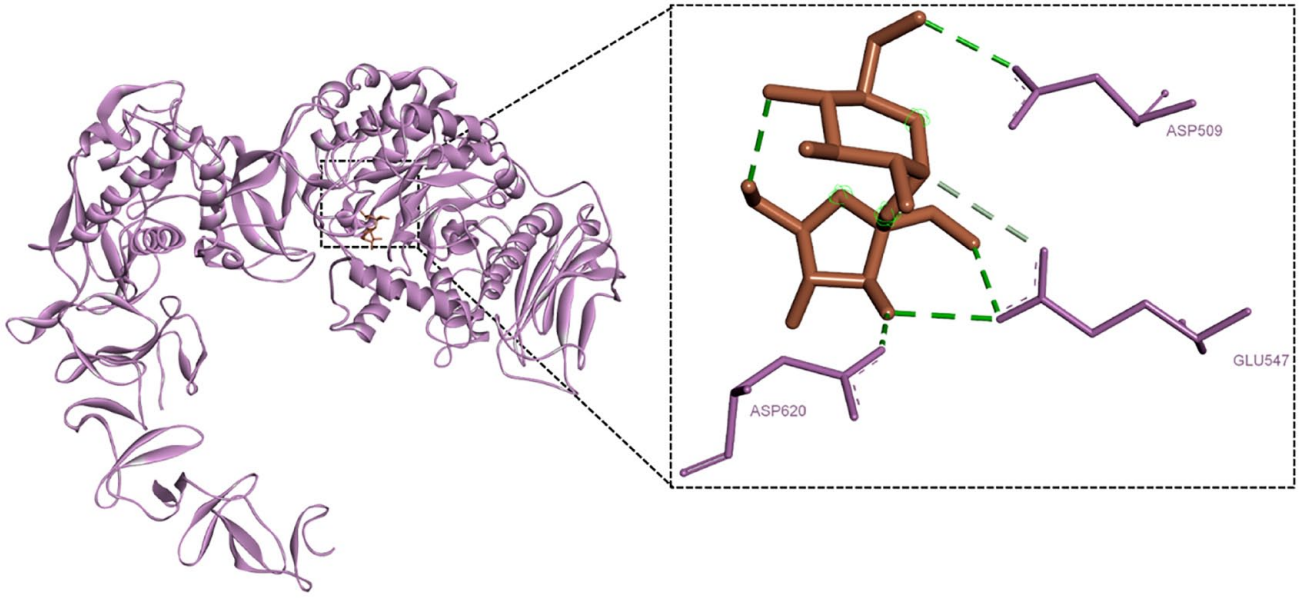

Figure 7. Surface interaction between sucrase (brown) and the active site of the catalytic domain Asp509, Glu547, and Asp620 of dextransucrase (lavender).

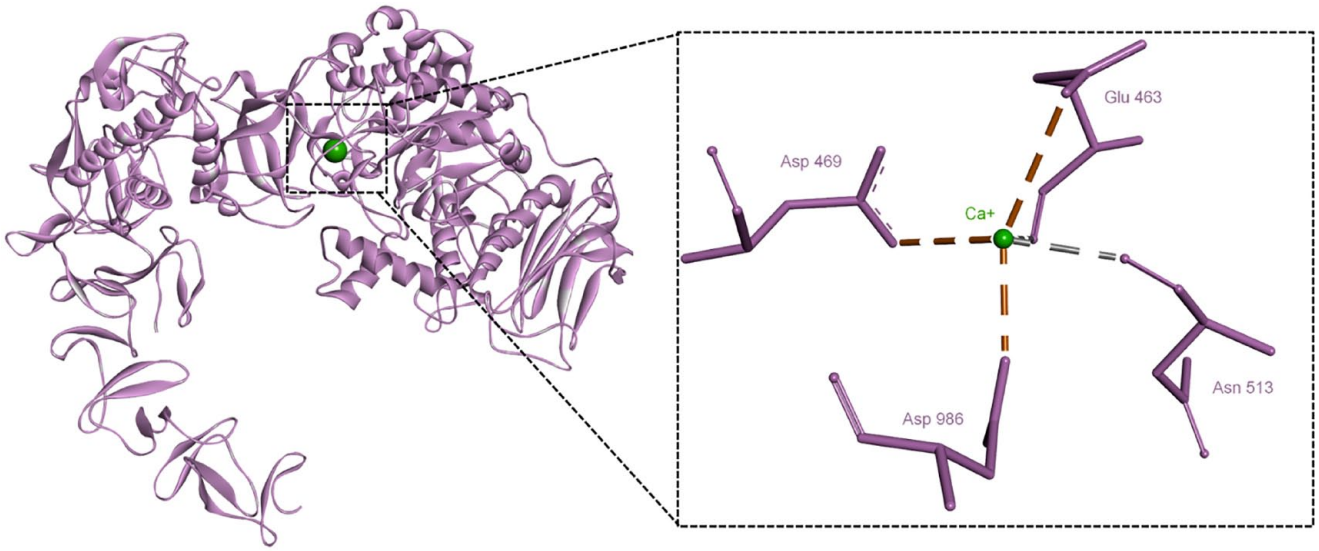

Figure 8. Surface interaction between calcium (green) and the residues Glu 483, Asp 469, Asn 513, Asp 986 of the DSR-IBUN (lavender).

in the packaging of the side chains and in the conformation of the central segments and loops have also been observed in the comparative models,[16] as well as the transmembrane sequences of the proteins,[76,77] as these are the ones that tend to present the greatest experimental error, which, in turn, is reflected in the models. These drawbacks have been minimized by development of new algorithms for both sequence comparison and structure comparison,[32] and because other techniques such as single-particle cryoelectron microscopy (Cryo-EM) and small-angle x-ray scattering have complemented existing x-ray information and provided new information, obtaining more and better experimentally determined protein structures.[78] For this, the number of applications in which comparative modeling has been shown to be useful has grown rapidly.

It is important to emphasize that, as is known, knowledge of the 3D structure of a protein is crucial to answer many biological questions; however, obtaining protein structures resolved by experimental methods is up to 7 times smaller, compared with the number of sequences of genes and genomes reported,[79] a difference that increases day by day, thanks to the next-generation sequencing (NGS) systems. Finally, comparative modeling methods, such as the one presented in this work, reduce the gap in knowledge of a specific structure, by providing more reliable and accurate protein structural models, compared with the experimental data provided. That is the reason why the model obtained is going to be used to study possible modifications that could imply changes that allow the production of polymers with different characteristics that could be used for

different applications. Molecular dynamics tests will also be carried out to establish more precisely what has been determined with the bioinformatics analyses carried out to date.

Conclusions

The values obtained for DSR-IBUN structural model for the GMQE and the QMEAN (Quality Mean) of 0.63 and −1.49, respectively, as well as the value obtained for the mean square deviation of the atomic positions (RMSD), corresponding to 0.09, allow us to affirm that the model obtained for the enzyme dextransucrase DSR_IBUN has adequate quality to be used as a source of information to perform different analyses directed to answer biological questions related to this enzyme.

Author Contributions

Luisa Garcia conceived the article and performed the general computational analysis and wrote the manuscript with the support of Sonia Ospina and Jairo Cerón. Margarita Gonzalez performed docking analysis and its graphical presentation. Sonia Ospina and Jairo Ceron supervised the project. All authors discussed the results and contributed to the final manuscript.

REFERENCES

1. Edward JH. Glycosyl transfer: a history of the concept's development and view of its major contributions to biochemistry. *Carbohydr Res*. 2001;331:347-368.
2. Quirasco M, López-Munguía A, Remaud-Simeon M, Monsan P, Farrés A. Induction and transcription studies of the dextransucrase gene in *Leuconostoc mesenteroides* NRRL B-512F. *Appl Environ Microbiol*. 1999;65:5504-5509.
3. Dols M, Remaud-Simeon M, Willemot RM, Vignon M, Monsan O. Characterization of the different dextransucrase activities excreted in glucose, fructose, or sucrose medium by *Leuconostoc mesenteroides* NRRL B-1299. *Appl Environ Microbiol*. 1998;64:1298-1302. doi:10.1128/AEM.64.4.1298-1302.1998
4. Vasileva T, Kirilov A, Bivolarski V, et al. Characterization of glycansucrase activities from *Leuconostoc Mesenteroides* LM17 and URE 13 strains. *Biotechnol Biotechnol Equip*. 2009;23:698-701. doi:10.1080/13102818.2009.10818520
5. Belder AN. *Dextran Handbook*. Amersham Biosciences; 2012. http://pro.unibz.it/staff2/sbenini/documents/protein%20purification%20handbooks/don't%20move/18116612aa.pdf
6. Kothari D, Das D, Patel S, Goyal A. Dextran and food application. In: Ramawat, K, Merillon, J eds. *Polysaccharides. Use*. Springer; 2014. https://www.researchgate.net/publication/272171914
7. Naessens M, Cerdobbel A, Soetaert W, Vandamme EK. Leuconostoc dextransucrase and dextran: production, properties and applications. *J Chem Technol Biotechnol*. 2005;80:845-860. doi:10.1002/jctb.1322
8. Flórez-Guzmán GY, Buitrago-Hurtado G, Ospina SA. New dextransucrase purification process of the enzyme produced by *Leuconostoc mesenteroides* IBUN 91.2.98 based on binding product and dextranase hydrolysis. *J Biotechnol*. 2018;265:8-14.
9. Fabre E, Bozonnet S, Arcache A, Vignon M, Monsan P, Remaud-Simeon M. Role of the two catalytic domains of dsr-e dextransucrase and their involvement in the formation of highly α -1, 2 branched dextran. *J Bacteriol*. 2005;187:296-303.
10. Parlak M, Ustek D, Tanriseven A. Designing of a novel dextransucrase efficient in acceptor reactions. *Carbohydr Res. C*. 2014;386:41-47.
11. Moulis C, Joucla G, Harrison D, et al. Understanding the polymerization mechanism of glycoside-hydrolase family 70 glucansucrases. *J Biol Chem*. 2006;281:31254-31267.
12. Joucla G, Pizzut S, Monsan P, Remaud-Simeon M. Construction of a fully active truncated alternansucrase partially deleted of its carboxy-terminal domain. *Febs Lett*. 2006;580:763-768.
13. Kitaoka M, Robyt JF. Large-scale preparation of highly purified dextransucrase from a high-producing constitutive mutant of mesenteroides B-FMC. *Leuconostoc Enzyme Microb Technol*. 1998;512:386-391.
14. Wang C, Zhang HB, Li MQ, Hu XQ, Li Y. Functional analysis of truncated and site-directed mutagenesis dextransucrases to produce different type dextrans. *Enzyme Microb Technol*. 2017;102:26-34.
15. Santhoshkumar R, Yusuf A. In silico structural modeling and analysis of physicochemical properties of curcumin synthase (CURS1, CURS2, and CURS3) proteins of Curcuma longa. *J Genet Eng Biotechnol*. 2020;18:24. doi:10.1186/s43141-020-00041-x
16. Sánchez R, Sali A. Advances in comparative protein-structure modelling. *Curr Opin Struct Biol*. 1997;7:206-214.
17. Fiser A. Template-based protein structure modelling. *Methods Mol Biol*. 2010;673:73-94.
18. Ginalsky K. Comparative modeling for protein structure prediction. *Curr Opin Struct Biol*. 2006;16:172-177. doi:10.1016/j.sbi.2006.02.003
19. Protein Data Bank (PBD). Guide to understanding PDB data. Updated 2017. https://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/r-value-and-r-free
20. Kuhlman B, Bradley P. Advances in protein structure prediction and design. *Nat Rev Mol Cell Biol*. 2019;20:681-697.
21. Sambrook J, Russell DW. *Molecular Cloning a Laboratory Manual*. Cold Spring Harbor Laboratory Press; 2001.
22. Herráez A. Biomodel: complementary pages to the study of biochemistry and molecular biology. Updated 2020. http://biomodel.uah.es/lab/cibertorio/analysis/trans.htm
23. Expasy ProtParam tool. https://web.expasy.org/cgi-bin/protparam/protparam
24. Yu CS, Lin CJ, Hwang JK. Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. *Protein Sci*. 2004;13:1402-1406.
25. Yu CS, Chen YC, Lu CH, Hwang JK. Prediction of protein subcellular localization. *Prot Struct Func Bioinf*. 2006;64:643-651.
26. Almagro Armenteros JJ, Tsirigos KD, Kaae Sønderby C, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol*. 2019;37:420-423. doi:10.1038/s41587-019-0036-z
27. Sievers F, Wilm A, Dineen D, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*. 2011;7:539. doi:10.1038/msb.2011.75
28. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview version 2: a multiple sequence alignment and analysis workbench. *Bioinformatics*. 2009;25:1189-1191. doi:10.1093/bioinformatics/btp033
29. Quinn M. *Beginners Guide to Protein Modelling* [Video File]; 2017. https://www.youtube.com/watch?v=UAF6ggMHoGE
30. Combet C, Blanchet C, Greourion C, Deléage G. Network protein sequence analysis. *TIBS*. 2000;25:147-150. https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html
31. Waterhouse A, Bertoni M, Bienert S, et al. SWISS-MODEL: homology modeling of protein structures and complexes. *Nucleic Acids Res*. 2018;46:W296-W303.
32. Kelly LA, Sternberg MJE. Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc*. 2009;4:363-371. doi:10.1038/nprot.2009.2
33. Zhang Y. I-TASSER server for protein 3D structure prediction. *BCM Bioinform*. 2008;9:40. doi:10.1186/1471-2105-9-40
34. Vujicic-Zagar A, Pijning T, Kralj S, Dijkstra BW. Crystal structure of a 117 KDa glucansucrase fragment provides insight into evolution and product specificity of GH70 enzymes. *Proc Natl Acad Sci*. 2010;107:21406-21411.
35. Ito K, Ito S, Shimamura T, et al. Crystal structure of glucansucrase from the dental caries pathogen Streptococcus mutans. *J Mol Biol*. 2011;408:177-186.
36. Leemhuis H, Pijining T, Dobruchowska JM, et al. Glucansucrases: three-dimensional structures, reactions, mechanisms,α-grucan analysis and their applications in biotechnology and food applications. *J Biotechnol*. 2013;163:250-272.
37. Brison Y, Pijning T, Malbert Y, et al. Functional and structural characterization of α-(1-2) branching sucrase derived from DSR-E glucansucrase. *J Biol Chem*. 2012;128:7915-7924.
38. DeLano WL. Pymol: an open-source molecular graphics tool. *CCP4 Newslett Prot Crystallgraph*. 2002;40:82-92.
39. BIOVIA. 2022. https://discover.3ds.com/discovery-studio-visualizer-download
40. Van Zundert GCP, Rodrigues JPGL, Trellet M, et al. The haddock2.2 Web server: user-friendly integrative modeling of biomolecular complexes. *J Mol Biol*. 2016;4218:720-725.
41. Nakashima H, Nishikawa K. Discrimination of intracellular and extracellular proteins using amino acid composition and residue-pair frequencies. *J Mol Biol*. 1994;238:54-61. doi:10.1006/jmbi.1994.1267
42. Gasteiger E, Hoogland C, Gattiker A, et al. Protein identification and analysis tools on the ExPASy server. In: Walker, JM, ed. *The Proteomics Protocols Handbook*. Humana Press; 2005:571-607.
43. Jaimes M. *In-Silico Prediction of the Structure and Function of the Hypothetical Protein P284 of Trypanosoma Cruzi* [Undergraduate thesis biology]. Bogotá, Colombia: Pontificia Universidad Javeriana; 2012.
44. Cozzone AJ. Fundamental *chemical properties*. In: Maccarrone M, ed. *Encyclopedia of life sciences*. MacMillan Publishers Ltd, Nature Publishing Group; 2002: 1-10.

45. Filiz E, Koç I. In silico sequence analysis and homology modeling of predicted beta-amylase 7-like protein in Brachypodium distachyon L. *J Biosci Biotech*. 2014;3:61-67.
46. Neubauer H, Bauche A, Mollet B. Molecular characterization and expression analysis of the dextransucrase DsrD of *Leuconostoc mesenteroides* Lcc4 in homologous and heterologous Lactococcus lactis cultures. *Microbiology (Reading)*. 2003;149:973-982.
47. Flórez-Guzmán GY. *Study of the Dextransucrase (DS) Enzyme Produced by Leuconostoc Mesenteroides Strain IBUN 91.2.98* [PhD thesis in biotechnology]. Bogota, Colombia: National University of Colombia, Institute of Biotechnology; 2014.
48. Siddiqui NN, Aman A, Qader SA. Molecular characterization of dextransucrase DsrN (dsrN) gene from Leuconostoc mesenteroides KIBGE IB-2.Genebank. Access Number AFP53921.1. https://www.ncbi.nlm.nih.gov/protein/399893018.
49. Siddiqui MS, Thodey K, Trenchard I, Smolke CD. Advancing secondary metabolite biosynthesis in yeast with synthetic biology tools. *FEMS Yeast Res* 2012;12:144-170.
50. Zhang H, Hu Y, Zhu C, Zhu B, Wang Y. GeneBank ABC75033.1. Cloning, sequencing and expression of a dextransucrase gene (dexYG) from *Leuconostoc mesenteroides*. *Biotechnol Lett*. 2008;30:1441-1446. doi:10.1007/s10529-008-9711-8
51. Bhatnagar R, Singh DKS. Cloning and molecular characterization of dextransucrase gene from *Leuconostoc mesenteroides* NRRL B-F.512 GenBank: AAD10952.1. 1999. https://www.ncbi.nlm.nih.gov/protein/4205088
52. Fraga R, Aristicas RC, Martinez L, Gonzalez T, Garcia R. A novel dextransucrase encoding gene from Leuconostoc citreum M-3. GeneBank AHC31982.1. 2013. https://www.ncbi.nlm.nih.gov/protein/AHC31982.1
53. Kim H-S, Kim D, Ryu H-J, Robyt JF. *Leuconostoc mesenteroides* B-742CB, a dextransucrase gene. GeneBank AAG38021.1. https://www.ncbi.nlm.nih.gov/protein/AAG38021.1
54. Monchois V, Remaud-Simeon M, Monsan P, Willemot RM. Genebank AAB95453. Cloning and sequencing of a gene coding for an extracellular dextransucrase (DSRB) from *Leuconostoc mesenteroides* NRRL B-1299 synthesising only a α-1, 6 glucan. *Fems Microbiol Lett*. 1998;159:307-315.
55. Martínez S, Asad S, Furnham N, Thornton JM. The classification and evolution of enzyme function. *Biophys J*. 2015;2015:1082-1086. doi:10.1016/j.bpj.2015.04.020
56. Kato C, Kuramitsu HK. Carboxyl-terminal deletion analysis of the streptococcus mutans glucosyltransferase-I enzyme. *FEMS Microbiol Lett*. 1990;60:299-302.
57. Monchois V, Willemot R, Monsan P. Glucansucrases: mechanism of action and structure–function relationships. *FEMS Microbiol Rev*. 1999;23:131-151.
58. van Hijum SA, Kralj S, Ozimek LK, Dijkhuizen L, van Geel-Schutten IG. Structure-function relationships of glucansucrase and fructansucrase enzymes from lactic acid bacteria. *Microbiol Mol Biol Rev*. 2006;70:157-176.
59. Claverie M, Cioci G, Vuillemin M, et al. Investigations on the determinants responsible for low molar mass dextran formation by DSR-M dextransucrase. *ACS Catal*. 2017;7:7106-7119. doi:10.1021/acscatal.7b02182
60. Molina M, Cioci G, Moulis C, Séverac E, Remaud-Siméon M. Bacterial α-glucan and branching sucrases from GH70 family: discovery, structure-function relationship studies and engineering. *Microorganisms*. 2021;2021:91607. doi:10.3390/microorganisms9081607
61. Vuillemin M, Claverie M, Brison Y, et al. Characterization of the first α-(1→3) branching sucrases of the GH70 family. *J Biol Chem*. 2016;291:7687-7702. doi:10.1074/jbc.M115.688044
62. Yan M, Wang BH, Xu X, et al. Molecular and functional study of a branching sucrase- like glucansucrase reveals an evolutionary intermediate between two subfamilies of the GH70 enzymes. *Appl Environ Microbiol*. 2018;84:e02810-02817. doi:10.1128/AEM.02810-17
63. Venkatachalam G, Arumugam S, Doble M. Industrial production and applications of α/β linear and branched glucans. *Indian Chem Eng*. 2021;63:533-547. doi:10.1080/00194506.2020.1798820
64. Du R, Zhou Z, Han Y. Functional identification of the dextransucrase gene of *Leuconostoc mesenteroides* DRP105. *Int J Mol Sci*. 2020;21:6596. doi:10.3390/ijms21186596
65. Yang J, Zhang Y. Protein structure and function prediction using I-TASSER. *Curr Prot Bioinform*. 2015;52:5.8.1-5.815. doi:10.1002/0471250953.bi0508s5216
66. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. Carbohydrate-active enzymes database (cazy): an expert resource for glycogenomics. *Nucleic Acids Res*. 2009;37:D233-D238. doi:10.1093/Nar/Gkn663
67. Macgregor EA. An overview of clan GH-H and distantly related families. *Biology*. 2005;60(Suppl 16): 5-12.
68. Banas JA, Vickerman MM. Glucan-binding proteins of the oral Streptococci. *Crit Revoral Biol Med*. 2003;14:89-99.
69. Leemhuis H, Pijining T, Dobruchowska JM, et al. Glucansucrases: three-dimensional structures, reactions, mechanisms, α-grucan analysis and their applications in biotechnology and food applications. *J Biotechnol*. 2012;163:250-257.
70. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res*. 2014;42:D490-D495.
71. Giffard PM, Jacques NA. Definition of a fundamental repeating unit in streptococcal glucosyltransferase glucan-binding regions and related sequence. *J Dent Res*. 1994;73:1133-1141. doi:10.1177/00220345940730060201
72. Moulis C, Joucla G, Harrison D, et al. Understanding the polymerization mechanism of glycoside-hydrolase family 70 glucansucrases. *J Biol Chem*. 2006;281:31254-31267.
73. Wang C, Zhang HB, Li MQ, Hu XQ, Li Y. Functional analysis of truncated and site-directed mutagenesis dextransucrases to produce different type dextrans. *Enzyme Microb Technol*. 2017;102:26-34.
74. Rhee SH, Lee CH. Properties of dextransucrase from *Leuconostoc mesenteroides* isolated from Sikhae. *J Microbiol Biotechnol*. 1991;1:176-181.
75. Baker D, Sali A. Protein structure prediction and structural genomics. *Science*. 2001;294:93-96. doi:10.1126/science.1065659
76. Vitkup D, Melamud E, Moult J, Sander C. Completeness in structural genomics. *Nat Struct Biol*. 2001;8:559-566.
77. Hopf TA, Colwell LJ, Sheridan R, Rost B, Sander C, Marks DS. Three-dimensional structures of membrane proteins from genomic sequencing. *Cell*. 2012;149:1607-1621. doi:10.1016/j.cell.2012.04.012
78. Brito JA, Archer M. Structural biology techniques: X-ray crystallography, cryo-electron microscopy, and small-angle X-ray scattering. In: Crichton, RR, Louro, RO eds. *Practical Approaches to Biological Inorganic Chemistry*. 2nd ed. Elsevier; 2020. doi:10.1016/B978-0-444-64225-7.00010-9
79. Ginalsky K. Comparative modeling for protein structure prediction. *Curr Opin Struct Biol*. 2006;2006:16172-16177. doi:10.1016/j.sbi.2006.02.003