Assessment of Structural Units Deletions in the Archaeal Oligosaccharyltransferase AglB

Running Title: Effects of structural deletions in AglB.

Conrado Pedebos^{a,1*}, Hugo Verli^a

^a Center for Biotechnology, Federal University of Rio Grande do Sul (UFRGS), CP 15005, Porto Alegre, RS, 91501-970, Brazil.

* To whom correspondence should be addressed: E-mail: C.Pedebos@soton.ac.uk (C. Pedebos). ORCID: <https://orcid.org/0000-0002-4796-8620> - Tel.: +55-51-3308-6068; Fax: +55-51-3308- 7309.

Acknowledgments

Research was supported by the Centro Nacional de Supercomputação of the Universidade Federal do Rio Grande do Sul (CESUP/UFRGS), CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), CAPES (Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior), and FAPERGS (Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul).

¹ *Present Address*: School of Chemistry, University of Southampton, Southampton, SO17 1BJ, U.K.

Abstract

Oligosaccharyltransferases (OSTs) are enzymes that catalyze the transfer of a glycan chain to an acceptor protein. Their structure is composed by a transmembrane domain and a periplasmic / C-terminal domain, which can be divided into structural units. The *Archaeoglobus fulgidus* OST, AfAgIB, has unique structural units with unknown functions. Here, we evaluate the stability role proposed for AfAgIB units by employing molecular modelling and molecular dynamics simulations, to examine the effect of single and double deletions in the enzyme structure. Our results show a strong effect on the dynamics of the C-terminal domain for the mutated systems with increased fluctuations near the deleted areas. Conformational profile and stability are deeply affected, mainly in the double unit deletion, modifying the enzyme behavior and binding interfaces. Coordination at the catalytic site was not disrupted, indicating that the mutated enzymes could retain activity at some level. Hotspots of variation were identified and rationalized with previous data. Our data shows that structural units may provide stabilization interactions, contributing for integrity of the wild-type enzyme at high temperatures. By correlating our findings to structural units mutagenesis experimental data available, it was observed that structural units deletion can interfere with OSTs stability and dynamics but it is not directly related to catalysis. Instead, they may influence the OST structural integrity, and, potentially, thermostability. This work offers a basis for future experiments involving OSTs structural and functional characterization, as well as for protein engineering.

Keywords: Oligosaccharyltransferase, AgIB, structural units, molecular dynamics,

Archaeoglobus fulgidus.

Introduction

N-glycosylation is a co-/post-translational modification widespread among all domains of life. This process is involved in many actions and activities throughout cells, such as stability, signalling, and immune response in *Eukarya*¹; flagella assembly, motility, protection against extreme environments, and integrity of structures in *Archaea*²; host adhesion, invasion, and colonization in *Bacteria*^{3,4}. Despite the intrinsic differences for each species, the overall N-glycosylation pathway share many similarities⁵. It is comprised by many enzymes that sequentially add monosaccharides to a lipid carrier which is, then, translocated by a flippase to the exterior side of the cell (in *Archaea*), to the periplasmic space (in *Bacteria*), or to the luminal face of the ER (in *Eukarya*). In *Archaea*, the final step of the pathway is catalyzed by the oligosaccharyltransferase AglB, an enzyme that transfers the lipid-linked oligosaccharide (LLO) to an acceptor protein carrying the sequon N-X-S/T (that is, Asn-any amino acid residue except Pro-Ser or Thr)⁶.

Recently, the full structure of two prokaryotic OSTs were elucidated: PglB from *Campylobacter lari* (CIPglB)^{7,8}, and the longest paralog AglB from *Archaeoglobus fulgidus* (AfAglB-L) (Figure 1)⁹. Before that, only partial structures for the C-terminal globular domains were obtained¹⁰⁻¹⁴, demonstrating that the archaeal domain of life possesses the greatest variability. As seen in these previous works, the structure of an OST can be divided in two large domains: a transmembrane domain (TM) consisting of

13 α -helices, while also having external loops (EL) that display important catalytic residues⁷, such as EL1, EL2, and EL5; and a C-terminal globular domain, which usually consists of an α/β mixed fold. This domain can be further subdivided into structural units. Firstly, the central core (CC) unit is the only substructure that is essential to every OST. It participates in the interface region with the TM domain, assembling the enzyme in two binding sites (protein acceptor site and LLO donor site). Accordingly, this unit holds the WWDYGY (WWDXGX) conserved motif^{10,15}, responsible for the recognition of the S/T (+2) of the N-glycosylation sequon, and the DK motif^{7,10}, which keeps the (+2) position residue in a tight binding by providing additional interactions. The second unit is the Insertion Sequence (IS), which receives this name because it seems to be inserted in the CC unit area¹⁴. Generally, it displays an organization of β -strands. For example, CIPglB and *P. furiosus* AglB (PfAglB) possess a distorted β -barrel fold, while AfAglB-L presents a reduced unit without the β -barrel organization. As a substituent for the barrel structure, a three-helices bundle (TH-Bundle) (originally classified as being part of the CC unit)¹⁴ is located at the same place and seems to play that role, a unique feature in the OSTs structures found until now, including the partial structures. The two final structural units are the Peripheries 1 (P1) and 2 (P2), only identified in Archaeal structures. These units are β -sheet rich, with slightly different organizations, and they are distributed around the CC unit, on the opposite direction of the IS unit. Archaeal OSTs can display either only P1 or both units at the same time, but no structure carrying only P2 was determined yet.

Despite all of this information, until today only one experimental study assessed

structural units properties¹¹. In this previous work by Matsumoto and co-workers, a mutant PfAglB with a deleted IS unit (PfAglB Δ IS) displayed the same activity than the wild type enzyme, indicating that this unit is not mandatory for catalysis. However, the assays were performed with detergent-solubilized PfAglB, in a temperature lower than optimal, and a short peptide as an acceptor, caveats that could interfere with proper evaluation of the enzyme function¹¹. Two hypotheses emerged for the functions of the IS unit: structural stability when the protein is embedded in a membrane, and support for the regulation or efficiency of the enzyme's activity when in the presence of larger polypeptide chains and when coupled to protein translation processes¹¹. As for P1 and P2, the only current hypothesis¹⁴ is that they could contribute for the thermostabilization of the enzyme under extreme conditions, since the extra units (and specifically P1 and P2) are only found in this domain of life. Another mentioned possibility is that these units could be equivalent to the subunits that constitute the OST multimeric complex from eukaryotic species (5-8 subunits). Recent structural characterizations of these eukaryotic OST complexes^{16,17} have demonstrated the presence of a β -sheet rich subunit (Wbp1) in areas similar to the ones occupied by P1 and P2 units in archaeal OSTs (Figure S1). However, this subunit presents distinct folding and organization compared to P1 and P2, along with a transmembrane portion not observed in those units. Although they might play similar roles in the process, the lack of similarity and the still unknown function of Wbp1 prevent us to make further inferences.

Investigations of OSTs mechanisms could lead to the production of tailor-made glycoproteins. In fact, recent data have demonstrated modifications enhancing OSTs

catalysis, such as mutations of residues to increase the recognition and binding of acceptors¹⁸⁻²⁰, change of the catalytic ion to increase the rate of product formation^{9,21,22}, as well as the production of different sizes of lipid carrier and interspecies glycan chains^{23,24}. However, there is still a lack of studies evaluating the role of IS, P1, and P2 structural units, which could bring new insights into the enzyme functionality. Therefore, in this work, we employed molecular dynamics (MD) simulations to observe the structural effects caused by full deletions of AfAgIB-L extra structural units. Specifically, we aimed to: I) evaluate the mutants stability after the deletion of P1, and P1 plus IS units; II) verify the behavior of the catalytic site in order to analyze if the requirements for catalysis are maintained upon deletion of the domains; III) identify important interactions between structural units that may govern the protein stability under hyperthermophilic conditions, allowing inferences regarding these units transferability between distinct OSTs. Our work seeks to contribute for the comprehension of the structural basis that governs OSTs dynamical behavior, providing insights for potential biotechnological applications of these enzymes.

Materials and methods

Systems Preparation

For this study, the structure from the RCSB Protein Data Bank under the PDB ID: 3WAK⁹ was retrieved, since it corresponds to a full apo structure containing EL5. Mutants were obtained by deleting the P1 region (residues 777-868), and the IS region plus the TH-Bundle (residues 638-776), including a short region from the CC unit that

connects IS to P1. Three systems were studied in this work: i) wild type AfAgIB-L in its apo state (AfAgIB-L-WT); ii) AfAgIB-L with the deletion of P1 unit (AfAgIB-L- Δ P1); iii) AfAgIB-L with deletion of IS and P1 (AfAgIB-L- Δ ISP1). In this work, we established that the Δ ISP1 included the TH-Bundle from the CC unit. The reason for that is because it makes close contacts with the IS unit and forms a structure separated from the enzyme core, resembling the common β -barrel IS seen in other OSTs. To better describe the environment where these proteins are found, they were embedded in a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE) membrane with 512 phospholipids. The methodology utilized for protein insertion was based on the tools LAMBADA and InflateGRO2²⁵.

Molecular Dynamics Simulations

The triclinic simulation boxes received SPC water solvation²⁶, in the presence of periodic boundary conditions. In the solvation step, Van der Waals radii of C atoms were raised from 0.150 to 0.375 to avoid water molecules filling spaces between phospholipids and protein. This parameter was returned to the default value after this step. PME²⁷ was chosen for the electrostatic treatment of the system. The LINCS method²⁸ was applied to constrain covalent bonds, allowing an integration step of 2 fs. GROMACS simulation suite²⁹ with GROMOS54A7 force field³⁰ was used for the molecular dynamics (MD) simulations. POPE membranes were described by GROMOS53A6 parameters (compatible with GROMOS54A7) following a previous report³¹. Cl⁻ counter ions were added to neutralize the residual charge of the systems,

when needed.

Firstly, energy minimizations were performed using the Steepest Descent algorithm, stopping when force on the system was below $100 \text{ kJ}\cdot\text{mol}^{-1}\cdot\text{nm}^{-1}$ as a convergence criterion. Then, MD simulations were performed in two phases: an equilibration and a production phase. For the equilibration phase, the systems were simulated for 1 ns in a canonical ensemble (NVT) and for 20 ns in an isothermal-isobaric ensemble (NPT). Parrinello-Rahman barostat^{32,33} was applied with a coupling constant of 2.0 ps. As thermostats, V-rescale (NVT step)³⁴ and Nosé-Hoover (NPT step - equilibration and production phases)^{35,36} were employed with coupling constants of $\sigma = 0.1$ and 1.0 , respectively. Also, semiisotropic pressure was applied due to the presence of the membrane. A constant temperature of 356 K was implemented, since this is the optimal growth temperature for *A. fulgidus*³⁷. During the equilibration phase, the protein was subjected to position restraints with a force constant of $1000 \text{ kJ}\cdot\text{mol}^{-1}\cdot\text{nm}^{-2}$. Subsequently, in the production phase, unrestrained MD simulations were performed for 300 ns for each system. Simulations were performed in triplicates (3 x 300 ns), starting with different velocities, aiming to enhance the sampling robustness of the conformational ensemble and remove low probability conformational events. Data was collected during the production phase and analyzed using GROMACS suite tools. VMD³⁸ and PyMOL³⁹ were used for trajectories visualization and for manipulation of structures.

Results

Global Structural Stability

We began by performing a global evaluation of the enzymes observing the deviation from the initial structures. The root mean square deviation (RMSD) (Figure 2A) plot for the whole structure indicates that both mutated systems demonstrate higher divergence in their structures when compared to the wild-type system. Particularly, the AfAgIB-L- Δ ISP1 mutated system appears to suffer a dramatic increase on its average RMSD, showing that the double deletion causes a much deeper impact in the protein structure. AfAgIB-L- Δ P1 does not display such high variations, although there is a clear tendency of higher magnitude change compared to AfAgIB-L-WT.

Analyzing the difference between the fluctuation of the backbone atoms as an average per residue (Figure 3), we observed that the TM domain (residues 1 to 500) maintain a generally stable behavior in its helices for all systems, with some minor variations in flexibility for the intracellular loops. EL5 shown its well-known flexibility, depicted by the large standard deviations observed in its residues.

As expected, we perceived a deeper influence from the structural units deletion in the C-terminal domain. Each mutated system displays a unique pattern of higher fluctuations when compared to AfAgIB-L-WT. The single unit deletion (AfAgIB-L- Δ P1) impacts areas previously interfaced with the P1 unit (Figure 3B), specially near the IS unit and TH-Bundle region: the loop (residues 713-717) connecting helices α B and α C (TH-Bundle), the loop (residues 756-762) that covalently attaches helix α A (TH-Bundle) with the core of the enzyme, and a loop (residues 685-692) that provided non-bonded interactions between the CC unit and the P1 unit. The double unit deletion (AfAgIB-L-

Δ ISP1) deeply affected (Figure 3C) the CC unit region: the first helix of the C-terminal domain (residues 501-516), both the β -strands and the kinked-helix (residues 546-549 and residues 604-618) that carry important recognition motifs (WWDXGX and DK), as well as two other small helical areas (residues 587-594 and 631-637).

The IS unit seems to be naturally flexible, since it demonstrated some flexibility also in the AfAgIB-L-WT system. However, the deletion of P1 caused an enhancement of this feature, more specifically at the TH-Bundle, which interacts with the β -hairpin structure, substituting the distorted β -barrel seen in ClPgIB. To further evaluate the extent of these instabilities, we analyzed the secondary structure content of each system (Figure 2B), considering only the residues in common between all systems. All proteins had a decrease in their content of secondary structure of at least 20 residues. Both mutations appear to have a higher decrease trend on the secondary structure content, especially in AfAgIB-L- Δ ISP1, exhibiting the most pronounced diminishment (around 35 residues).

Structural Units Deletion Affects the Conformational Behavior of AfAgIB

To better comprehend how the mutations impact the sampling of AfAgIB conformational landscape, we employed principal component analysis (PCA) on the MD simulations data. By plotting the two-dimensional subspace defined by the top two principal components (PCs) (Figure 4), we were able to evaluate the structural relaxation of the proteins individually. For the WT system, a similar behavior occurred for all replicates considering the whole protein structure, as observed by plotting the top two

eigenvectors projections: an initial fluctuation, jumping among low-energy basins, as well as an apparent accommodation by the end of each simulation. Replicates 1 and 2 sample mostly one specific conformational state, while replicate 3 takes some time jumping among different regions before finding a more stable area. Despite that, the replicates explored multiple distinct areas from the subspace, indicating that our simulations possibly did not reach convergence in the simulated time. As for the mutated systems, the replicas behaviors are identical: large drifts through the conformational space, as well as no overlap between systems, indicating that these systems are even farther from convergence than the apo state simulations, which may be related to the appearance of new energy minima upon domains deletion.

To better characterize the influence of the deletions on the protein structure, we employed the same analysis for the isolated TM domain, and for the C-terminal domain structural units (CC, IS, and P1) of AfAgIB-L (Figure 5). Firstly, all trajectories are concatenated into one, so that the results for different systems may be compared for a common subspace. In agreement with our observations concerning the structural divergences and fluctuations, the AfAgIB-L-WT and the AfAgIB-L- Δ P1 systems demonstrated a less divergent behavior, including an apparent level of overlap between some of their replicates (indicating similar conformational sampling), specifically in the TM domain and in the CC unit. In addition, these units movements are more contained at specific portions of the subspace spanned by the two PCs analyzed, a behavior which suggests that there is less conformational variation in this domain. Contrastingly, AfAgIB-L- Δ ISP1 is largely affected by the dual deletion, demonstrating the same drift behavior observed for the global structure. In fact, the structural adjustment suffered by

AfAgIB-L- Δ ISP1 is so intense that subtle conformational variations from the other two systems are probably concealed.

The CC unit repeated the same pattern seen for the TM domain: AfAgIB-L-WT and AfAgIB-L- Δ P1 displayed a similar shape on the projections, while AfAgIB-L- Δ ISP1 maintained its dominance in the analysis with a high drifting pattern, even larger than for the TM domain, confirming that this mutation affected more the C-terminal area.

Subsequently, when we analyzed the IS unit, the deletion in AfAgIB-L- Δ P1 demonstrated a considerable amount of influence on this unit behavior. Replicate 2 mainly sampled the same regions as the replicate 2 of the WT system, while replicates 1 and 3 from AfAgIB-L- Δ P1 drifted apart from the others, accounting for most of the variance constituting both PCs. AfAgIB-L-WT replica 3 samples of the same area as replicate 1 from the same system, while replicate 2 generates a new basin. Finally, P1 unit (from AfAgIB-L-WT) shown a less diffuse behavior, populating one well-defined area of the subspace where the three replicates occupied for some time during the simulations, and two other areas that were occupied separately by replicates 2 and 3, while replicate 1 stayed trapped near its initial region.

Binding Site Coordination and Interface Stability

An important feature of OSTs, as metalloenzymes, is the maintenance of an organized network of interactions between the catalytic residues, as well as the metal ion coordination geometry. Consequently, aiming to assess the impact of the deletions on this network, we evaluated these interactions during the MD simulations by

calculating the distances concerning the atoms from the side-chain of essential residues (Table 1). From that, it was possible for us to observe that all systems maintained a mostly similar coordination geometry to the one found on the crystallographic structure (PDB ID: 3WAK), although with a couple of differences: His163, which showed an increase on its distance relative to Zn^{2+} , and the presence of more water molecules surrounding the metal ion (3:1), similarly to the coordination proposed on crystal form 1 in the same work by Matsumoto *et al.* (PDB ID: 3WAJ)⁹. Glu360, a presumed critical residue for the catalysis mechanism (equivalent to Glu319 in CIPglB), did not interacted with the metal ion, thus not forming the complete active site, as detected in the crystallographic structure. The metal ion remained in a tight binding at the enzyme core, with very rigid distances between the metal ligands (max. standard deviation of 0.4 Å), and one of the water molecules bridging an interaction involving His163 and Zn^{2+} . We identified fluctuating coordination numbers, between 6 and 7, depending on the denticity of the carboxylates from Asp and the number of water molecules involved (Figure 6), exhibiting either a distorted octahedral shape or a face-capped octahedron.

Discussion

The deletion of structural units from the C-terminal domain of AfAglB-L impacts on many aspects of the enzyme. The TM domain appears to be the least influenced region, as we expected, since it is far from the mutations area, besides being embedded within the lipid environment, which provides stabilizing interactions. As seen by analyzing the structural fluctuations and divergence to the crystal structure, the CC unit

of AfAgIB-L- Δ ISP1 was the most affected by the deletions, showing larger fluctuations throughout this unit, including a modification in the behavior of the kinked helix that carries the DK motif and in the WWDXGX motif area. As tested by previous reports^{12,40}, this modification could negatively influence sequon recognition by the OST, rendering a poor protein binding and, ultimately, reducing the enzyme activity. AfAgIB-L- Δ P1 is also influenced by the CC unit, but to a lesser extent; most of its destabilization is observed in the TH-Bundle, interfaced with the IS unit. The increased flexibility of both of these areas may pose as a drawback because not only it modifies the protein binding site interface with the acceptor protein, but it might also propagate to the CC unit, further destabilizing it in other areas, including the DK motif and the area near the LLO cavity. The presence of P1 in AfAgIB-L-WT stabilizes the IS unit by the interface between the two units, interacting with helix α C of the TH-Bundle and the loop connecting it to the CC unit.

PCA technique is reported as a valuable tool to assess conformational sampling and convergence of simulations⁴¹⁻⁴³. In our work, by using PCA we identified important aspects involving the conformational ensembles of the studied systems: 1) In spite of none of the systems reaching convergence on the PCA metric during our simulations, AfAgIB-L-WT was the most dynamically stable system, since it was able to achieve stable energy basins, sample this areas, and had more overlap between its replicates, as well as a smaller drift along the PCs. The mutations affected the behavior of AfAgIB-L, leading to huge drifts along the conformational space with just a few clearly defined basins being explored. 2) When evaluated in the same subspace, it is evident that the

double unit deletion has a heavier impact on protein structure than the single deletion. In this case, none of the domains from AfAgIB-L-ΔISP1 demonstrated a stable behavior, keeping the same projection patterns observed for the whole protein. AfAgIB-L-WT and AfAgIB-L-ΔP1, contrarily, exhibited some superimposition, which is probably due to the dominance of AfAgIB-L-ΔISP1 largely fluctuating motions. Individual analyses for AfAgIB-L-WT and AfAgIB-L-ΔP1 systems confirmed that the overlap between their replicates is low, as well as highlighted the instabilities in the conformational profile of AfAgIB-L-ΔP1 when compared to WT. Therefore, we conclude that the mutated systems are mostly unable to keep a similar conformational profile to AfAgIB-L-WT, possibly interfering with the enzyme action, since the specific conformations that allow the binding of substrates may not be available on these systems.

Coordination of the metal ion at the binding site was very stable, indicating that, despite all the disturbance in the protein structure caused by the deletions, the enzyme could still be catalytic competent at some level, as long as the two substrates manage to bind in the cavities. A puzzling aspect of the coordination at the catalytic site involves His163. This residue is coordinating the metal ion in the crystal structure, displaying a distance of 2 Å from the nitrogen atom Nε2 to Zn²⁺. It is part of the DXD motif (DXH in AfAgIB-L) of glycosyltransferases, equivalent to Asp156 in CIPgIB OST, which has not been identified as one of the coordinating residues in its crystallographic structure⁷. These structures were obtained in different states, the apo (AfAgIB-L) and the cocrystallized with the peptide acceptor (CIPgIB), therefore one could assume that Asp156 and His163 may have a coordinating role only when the protein is in its apo

state. However, both in this work and in a previous report from CIPglB simulations⁴⁴, we did not observe this coordination participation during our simulations. Instead, a bidentate coordination by Asp47 (Asp56 in CIPglB) and water molecules mediated this interaction, maintaining the octahedral coordination geometry. Differently, a recent work⁴⁵ also involving MD simulations of PglB, observed a coordination participation from Asp156 in all states studied. This way, the role of these residues is not clear yet, despite being widespread in this enzyme family⁴⁶, as well as being fundamental to the enzyme activity⁴⁷. One possibility is that His163 (and Asp156 in CIPglB) serves as a “backup” coordination ligand, preventing the metal ion from escaping the catalytic site by bridging water molecules interactions or by acting in situations where the protein substrate is absent. Another possibility would be that this residue is involved in catalysis by supporting the binding or the stabilization of the LLO substrate, since the monosaccharide at the reducing-end of the glycan is a glucose with three sulfates in positions C2, C3 and C6 and could be accommodated in a pocket consisting of the conserved residues His81, His162, His163, and Trp215. Of course, we have to consider that MD simulations do not handle electronic properties directly, which could be a caveat that prevented us from observing a correct description of the His163 interactions. Glu360 never reached the metal ion during any of the simulations, hence, the binding site was never fully complete. This strengthens the hypothesis that acceptor protein binding is important for the formation of the complete binding site^{7,9}.

The data obtained for the catalytic site coordination corroborates previous data from PfAgIB, where the catalytic activity of PfAgIB-ΔIS was not affected by the deletion¹¹. Contrastingly, the global motions and the binding interface, as seen by the

PCs motions on its extreme projections (Figure S2), disagree with these observations. In AfAgIB-L-WT, motions are dominated by EL5 from the TM unit, and by a rearrangement of the IS unit, similarly to motions identified in CIPgIB dynamics⁴⁴, reinforcing the observations that the IS unit is intrinsically flexible. In AfAgIB-L-ΔP1, there is a twisting motion between TM and PP domains, whereas, in AfAgIB-L-ΔISP1, we identified huge collective motions, especially on the PP domain, disorganizing the protein binding site. Comparing the PP domain of the wild-type structures from both species (Figure 7), we identified possible explanations for these discrepancies: the IS unit from PfAgIB is a stable β -barrel structure, and the deletion (residues 603-678) allows residues 602 and 679 to be easily connected, thus causing less perturbations in the enzyme core than our deletion in AfAgIB-L-WT (which included the TH-Bundle and one conserved β -sheet at the enzyme core). Indeed, the deletion in PfAgIB permits a structural organization of the CC unit that resemble the other two paralogs from *A. fulgidus*, AfAgIB-S1 and AfAgIB-S2, which do not portray IS units. These observations indicate that structural units from PfAgIB could act more independently when compared to AfAgIB. Furthermore, PfAgIB possesses two periphery units (P1 and P2), one that occupies the same place as P1 in AfAgIB-L-WT, and another that encircles the area above the LLO cavity. As seen in our RMSF plots (Figure 3), this area has many unstable regions with fluctuations that are largely increased as a side-effect of the deletions. Flexibility of the same nature and at the same regions was also seen in CIPgIB MD simulations⁴⁴, confirming the disordering of these areas. Possibly, the presence of the β -strand rich P2 unit provides additional stabilization for PfAgIB, protecting CC unit from unfolding, as well as allowing the

deletion of IS to be tolerable. Perhaps, if AfAgIB-L-WT possessed both P1 and P2, it could also tolerate the deletion of IS, as the single deletion of P1 was less aggressive to the structural organization of the enzyme core than the double deletion. Besides, the large-amplitude motions observed after both deletions in AfAgIB-L would be contained by the presence of P2, promoting diminished fluctuations at the PP domain.

Conclusions

Structural studies of OSTs have been constantly advancing, enabling a better understanding of the molecular basis involving N-glycosylation modifications, which, in turn, may provide new tools for the glycoengineering of biomolecules. Also, many experiments have been performed regarding these enzymes catalytic mechanism, usually evaluating residues from CC unit and TM domain, as well as studies designed to enhance their activities. However, to this date, only one previous work has assessed the role of other structural units from OSTs¹¹. In this work, we try to contribute for the structural comprehension of OSTs and increase the knowledge about structural units dynamics and function. By employing MD simulations, we could verify that both P1 and IS units provide fundamental stabilizing interactions that permits both structural integrity of AfAgIB-L in extreme environments, such as the high temperatures faced by *A. fulgidus*, along with the preservation of the enzyme conformational profile and binding interface, important features for its proper functioning. Combining our findings with previous data from literature, we identified that there might be some interchangeability between structural units. For example, one could transfer a P1 unit from AfAgIB or P1

and P2 from PfAglB to another OST, aiming to achieve a new class of engineered chimeric thermostable OSTs. PglB from the *Campylobacter* or *Desulfovibrio* genera would be good candidates for these approaches, since they are found in mesophilic species and have a well-known promiscuity²⁰. We hope that this work may support the development of more assessments regarding OSTs structural units.

Conflict of interest statement

None declared.

References

- [1] Varki A. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology*. 1993;3(2):97–130.
- [2] Eichler J, Koomey M. Sweet New Roles for Protein Glycosylation in Prokaryotes. *Trends Microbiol*. 2017;5(8):662-672.
- [3] Szymanski CM, Burr DH, Guerry P, Campylobacter protein glycosylation affects host cell interactions. *Infect Immun*. 2002;70(4):2242–2244.
- [4] Nothaft H, Szymanski CM. Bacterial protein N-glycosylation: new perspectives and applications. *J Biol Chem*. 2013;288(10):6912–6920.
- [5] Schwarz F, Aebl M. Mechanisms and principles of N-linked protein glycosylation. *Curr Opin Struct Biol*. 2011;21(5):576-582.
- [6] Chaban B, Voisin S, Kelly J, Logan SM, Jarrell KF. Identification of genes involved in

the biosynthesis and attachment of *Methanococcus voltae* N-linked glycans: insight into N-linked glycosylation pathways in Archaea. *Mol Microbiol.* 2006;61(1):259-268.

[7] Lizak C, Gerber S, Numao S, Aebi M, Locher KP. X-ray structure of a bacterial oligosaccharyltransferase. *Nature.* 2011;474:350-355.

[8] Napiórkowska M, Boilevin J, Sovdat T, Darbre T, Reymond JL, Aebi M, Locher KP. Molecular basis of lipid-linked oligosaccharide recognition and processing by bacterial oligosaccharyltransferase. *Nat Struct Mol Biol.* 2017;24:1100-1106.

[9] Matsumoto S, Shimada A, Nyirenda J, Igura M, Kawano Y, Kohda D. Crystal structures of an archaeal oligosaccharyltransferase provide insights into the catalytic cycle of N-linked protein glycosylation. *Proc Natl Acad Sci U S A.* 2013;110(44):17868-17873.

[10] Maita N, Nyirenda J, Igura M, Kamishikiryo J, Kohda D. Comparative structural biology of eubacterial and archaeal oligosaccharyltransferases. *J Biol Chem.* 2010;285(7):4941-4950.

[11] Matsumoto S, Igura M, Nyirenda J, Matsumoto M, Yuzawa S, Noda N, Inagaki F, Kohda D. Crystal structure of the C-terminal globular domain of oligosaccharyltransferase from *Archaeoglobus fulgidus* at 1.75 Å resolution. *Biochemistry.* 2012;51(20):4157-4166.

[12] Igura M, Maita N, Kamishikiryo J, Yamada M, Obita T, Maenaka K, Kohda D. Structure-guided identification of a new catalytic motif of oligosaccharyltransferase. *EMBO J.* 2008;27(1):234-243.

- [13] Nyirenda J, Matsumoto S, Saitoh T, Maita N, Noda NN, Inagaki F, Kohda D. Crystallographic and NMR evidence for flexibility in oligosaccharyltransferases and its catalytic significance. *Structure*. 2013;21(1):32-41.
- [14] Matsumoto S, Shimada A, Kohda D. Crystal structure of the C-terminal globular domain of the third paralog of the *Archaeoglobus fulgidus* oligosaccharyltransferases. *BMC Struct Biol*. 2013;13:11-20.
- [15] Yan Q, Lennarz WJ. Studies on the function of oligosaccharyl transferase subunits. Stt3p is directly involved in the glycosylation process. *J Biol Chem*. 2002;277(49):47692-47700.
- [16] Wild R, Kowal J, Eyring J, Ngwa EM, Aepli M, Locher KP. Structure of the yeast oligosaccharyltransferase complex gives insight into eukaryotic N-glycosylation. *Science*. 2018;359(6375):545-550.
- [17] Bai L, Wang T, Zhao G, Kovach A, Li H. The atomic structure of a eukaryotic oligosaccharyltransferase complex. *Nature*. 2018;555:328-333.
- [18] Ollis AA, Zhang S, Fisher AC, DeLisa MP. Engineered oligosaccharyltransferases with greatly relaxed acceptor-site specificity. *Nat Chem Biol*. 2014;10(10):816-822.
- [19] Ihssen J, Haas J, Kowarik M, Wiesli L, Wacker M, Schwede T, Thöny-Meyer L. Increased efficiency of *Campylobacter jejuni* N-oligosaccharyltransferase PglB by structure-guided engineering. *Open Biol*. 2015;5(4):140227.
- [20] Ollis AA, Chai Y, Natarajan A, Perregaux E, Jaroentomeechai T, Guarino C, Smith J, Zhang S, DeLisa MP. Substitute sweeteners: diverse bacterial

oligosaccharyltransferases with unique N-glycosylation site preferences. *Sci Rep*. 2015;20:15237.

[21] Gerber S, Lizak C, Michaud G, Bucher M, Darbre T, Aebi M, Reymond JL, Locher KP. Mechanism of bacterial oligosaccharyltransferase: in vitro quantification of sequon binding and catalysis. *J Biol Chem*. 2013;288(13):8849-8861.

[22] Lizak C, Gerber S, Michaud G, Schubert M, Fan YY, Bucher M, Darbre T, Aebi M, Reymond JL, Locher KP. Unexpected reactivity and mechanism of carboxamide activation in bacterial N-linked protein glycosylation. *Nat Commun*. 2013;4:2627.

[23] Liu F, Vijayakrishnan B, Faridmoayer A, Taylor TA, Parsons TB, Bernardes GJ, Kowarik M, Davis BG. Rationally designed short polyisoprenol-linked PglB substrates for engineered polypeptide and protein N-glycosylation. *J Am Chem Soc*. 2014;136(2):566-569.

[24] Srichaisupakit A, Ohashi T, Misaki R, Fujiyama K. Production of initial-stage eukaryotic N-glycan and its protein glycosylation in *Escherichia coli*. *J Biosci Bioeng*. 2015;119(4):399-405.

[25] Schmidt TH, Kandt C. LAMBADA and InflateGRO2: efficient membrane alignment and insertion of membrane proteins for molecular dynamics simulations. *J Chem Inf Model*. 2012;52(10):2657-2666.

[26] Berendsen HJC, Grigera JR, Straatsma TPJ. The missing term in effective pair potentials. *Phys Chem*. 1987;91:6269–6271.

[27] Darden T, York D, Pedersen L. Particle mesh Ewald: an N·log (N) method

for Ewald sums in large systems. *J Chem Phys.* 1993;98(24):10089–10092.

[28] Hess B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J Chem Theory Comput.* 2008;4(1):116-122.

[29] Abraham MJ, Murtolad T, Schulzb R, Páll S, Smith JC, Hess B, Lindahl E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX.* 2015;1-2:19-25.

[30] Schmid N, Eichenberger AP, Choutko A, Riniker S, Winger M, Mark AE, van Gunsteren WF. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur Biophys J.* 2011;40(7):843-856.

[31] Piggot TJ, Holdbrook DA, Khalid S. Electroporation of the E. coli and S. Aureus membranes: molecular dynamics simulations of complex bacterial membranes. *J Phys Chem B.* 2011;115(45):13381-13388.

[32] Parrinello M, Rahman A. Polymorphic transitions in single crystals: a new molecular dynamics method. *J Appl Phys.* 1981;52:7182–7190.

[33] Nosé S, Klein ML. Constant pressure molecular dynamics for molecular Systems. *Mol Phys.* 1983;50:1055–1076.

[34] Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity Rescaling. *J Chem Phys.* 2007;126(1):014101–014107.

[35] Nosé S. A molecular dynamics method for simulations in the canonical Ensemble. *Mol Phys.* 1984;52:255–268.

[36] Hoover WG. Canonical dynamics: equilibrium phase-space distributions.

Phys Rev A. 1985;31:1695–1697.

[37] Klenk H-P, Clayton RA, Tomb J-F, White O, Nelson KE, Ketchum KA, Dodson RJ, Gwinn M, Hickey EK, Peterson JD, Richardson DL, Kerlavage AR, Graham DE, Kyripides NC, Fleischmann RD, Quackenbush J, Lee NH, Sutton GG, Gill S, Kirkness EF, Dougherty BA, McKenney K, Adams MD, Loftus B, Peterson S, Reich CI, McNeil LK, Badger JH, Glodek A, Zhou L, Overbeek R, Gocayne JD, Weidman JF, McDonald L, Utterback T, Cotton MD, Spriggs T, Artiach P, Kaine BP, Sykes SM, Sadow PW, D'Andrea KP, Bowman C, Fujii C, Garland SA, Mason TM, Olsen GJ, Fraser CM, Smith HO, Woese CR, Venter JC. The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature*. 1998;394:364-370.

[38] Humphrey W, Dalke A, Schulten K. VMD—visual molecular dynamics. *J Mol Graph*. 1996;14(1):33–38.

[39] The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

[40] Hese K, Otto C, Routier FH, Lehle L. The yeast oligosaccharyltransferase complex can be replaced by STT3 from *Leishmania major*. *Glycobiology*. 2009;19(2):160-171.

[41] Ng HW, Laughton CA, Doughty SW. Molecular Dynamics Simulations of the Adenosine A2a Receptor: Structural Stability, Sampling and Convergence. *J Chem Inf Model*. 2013;53(5):1168-78.

[42] Amadei A, Ceruso MA, Di Nola A. On the convergence of the conformational coordinates basis set obtained by the essential dynamics analysis of proteins' molecular

dynamics simulations. *Proteins*. 1999;36(4):419-424.

[43] Hess B. Convergence of sampling in protein simulations. *Phys Rev E Stat Nonlin Soft Matter Phys*. 2002;65(3):031910-031920.

[44] Pedebos C, Arantes PR, Giesel GM, Verli H. *In silico* Investigation of the PglB Active Site Reveals Transient Catalytic States and Octahedral Metal Ion Coordination. *Glycobiology*. 2015;25(11):1183–1195.

[45] Lee HS, Im W. Transmembrane Motions of PglB Induced by LLO are Coupled with EL5 Loop Conformational Changes Necessary for OST Activity. *Glycobiology*. 2017;27(8):734-742.

[46] Breton C, Šnajdrová L, Jeanneau C, Koča J, Imberty A. Structures and mechanisms of glycosyltransferases. *Glycobiology*. 2006;16(2):29R-37R.

[47] Li J, Rancour DM, Allende ML, Worth CA, Darling DS, Gilbert JB, Menon AK, Young Jr WW. The DXD motif is required for GM2 synthase activity but is not critical for nucleotide binding. *Glycobiology*. 2001;11(3):217-229.

Tables

Table 1. Distances from catalytic residues to Zn²⁺ ion during MD simulations (Å)

Residues and Atoms	MD Distance to Zn ²⁺ (Å)		
	AfAgIB-L-WT	AfAgIB-L-ΔP1	AfAgIB-L-ΔISP1
Asp47 Oδ1	2.2±0.2	2.2±0.2	2.3±0.4
Asp47 Oδ2	2.2±0.2	2.2±0.2	2.3±0.4
Asp161 Oδ1	2.2±0.2	2.3±0.3	2.4±0.5
Asp161 Oδ2	2.3±0.3	2.2±0.3	2.4±0.5
His163 Nε2	4.4±0.4	4.5±0.4	4.6±0.4
Glu360 Oε1	11.2±1.9	11.3±2.5	8.8±2.4
Glu360 Oε2	11.2±1.9	11.4±2.6	8.8±2.5

Figure Legends

Figure 1. AfAgIB-L (*left* - PDB ID: 3WAK) and CIPgIB (*right* - PDB ID: 3RCE) crystallographic structures. The TM domain (orange) and the C-terminal domain structural units (CC -green, IS - blue, P1 -red) are depicted. The purple sphere corresponds to Zn²⁺ ion, while the green sphere corresponds to Mn²⁺ ion. The TH-Bundle is originally part of CC unit, although it occupies the same place of the IS unit from CIPgIB and interacts with the IS unit from AfAgIB.

Figure 2. RMSD and Secondary Structure content during simulations. (A) RMSD

relative to the crystal structure for each system at every 10ns with standard deviations.

(B) Secondary structure content for each system at every 10ns with standard deviations.

Results are shown as an average between replicates with standard deviation.

Figure 3. RMS fluctuation difference between backbone atoms as an average of each residue comparing the mutated systems to the wild-type. (A) AfAgIB-L-WT structure of the C-terminal domain. (B) *Top.* C-terminal domain of the AfAgIB- Δ P1 system: CC unit is colored in green, IS unit is colored in blue, TH-Bundle is colored in brown, and the high fluctuation areas are colored in yellow. *Bottom.* RMSF difference (red) between AfAgIB-L- Δ P1 and AfAgIB-L-WT with stars marking the highest fluctuations areas. (C) *Top.* C-terminal domain of the AfAgIB- Δ ISP1 system: CC unit is colored in green, and the high fluctuation areas are colored in yellow. *Bottom.* RMSF difference (green) between AfAgIB-L- Δ ISP1 and AfAgIB-L-WT with stars marking the highest fluctuations areas. TM domain is hidden on (B) and (C) for better visualization. WWDXGX and DK motifs are represented in sticks.

Figure 4. Top two PCs projections extracted with the PCA technique for the individual systems with concatenated trajectories. Individual replicates are represented by a color code of shades of black (AfAgIB-L-WT), red (AfAgIB-L- Δ P1), and green (AfAgIB-L- Δ ISP1).

Figure 5. Top two PCs projections for all systems with concatenated trajectories extracted with PCA technique. Individual replicates are represented by a color code of shades of black (AfAgIB-L-WT), red (AfAgIB-L- Δ P1), and green (AfAgIB-L- Δ ISP1). Each graph represents a different domain or structural unit of the enzyme, as indicated in

each legend.

Figure 6. Snapshots of representative coordination geometries for the (A) crystallographic structure and each system studied: (B) AfAgIB; (C) AfAgIB- Δ P1; and (C) AfAgIB- Δ ISP1. Coordination numbers varies between 6 and 7, depending on the number of waters nearby or the denticity of the amino acids in the coordination complex. Residues from the catalytic site are represented in yellow licorice, while the rest of the protein, the membrane, and other water molecules are hidden for better visualization. Zn²⁺ ion is represented as a pink sphere.

Figure 7. Structural comparison between AfAgIB-L, AfAgIB-S1 and S2, and PfAgIB (PDB ID 2ZAI)¹². (A) AfAgIB-L with all other C-terminal domains aligned to its corresponding domain. (B) PfAgIB with the deletion of IS unit folds in a way that resembles the other two paralogs from *A. fulgidus*. (C) Alignment of the C-terminal domain of PfAgIB-WT to AfAgIB- Δ ISP1 left the P1 and P2 units (orange) aligned near a sensible area (green spheres) of the mutated system (AfAgIB- Δ ISP1), possibly protecting that region in a hypothetical chimeric AfAgIB-L.

Figures

Figure 1.

Figure 2.

Figure 3.

Figure 4.

Figure 5.

Figure 6.

Figure 7.