Structural basis for arginine glycosylation of host substrates by bacterial effector proteins

Jun Bae Park, Young Hun Kim, Youngki Yoo, Juyeon Kim, Sung-Hoon Jun, Jin Won Cho, Samir El Qaidi, Samuel Walpole, Serena Monaco, Ana A. García-García, Miaomiao Wu, Michael P. Hays, Ramon Hurtado-Guerrero, Jesus Angulo, Philip R. Hardwidge, Jeon-Soo Shin & Hyun-Soo Cho

The bacterial effector proteins SseK and NleB glycosylate host proteins on arginine residues, leading to reduced NF-κB-dependent responses to infection. *Salmonella* SseK1 and SseK2 are *E. coli* NleB1 orthologs that behave as NleB1-like GTs, although they differ in protein substrate specificity. Here we report that these enzymes are retaining glycosyltransferases composed of a helix-loop-helix (HLH) domain, a lid domain, and a catalytic domain. A conserved HEN motif (His-Glu-Asn) in the active site is important for enzyme catalysis and bacterial virulence. We observe differences between SseK1 and SseK2 in interactions with substrates and identify substrate residues that are critical for enzyme recognition. Long Molecular Dynamics simulations suggest that the HLH domain determines substrate specificity and the lid-domain regulates the opening of the active site. Overall, our data suggest a front-face Sni mechanism, explain differences in activities among these effectors, and have implications for future drug development against enteric pathogens.
Protein glycosylation is a post-translational modification implicated in a wide range of cellular/biological processes, including cell development, signaling cascades, and tumorigenesis. Glycosyltransferases (GTs) catalyze the transfer of a sugar moiety to acceptor substrates and are classified according to their folding as GT-A, GT-B, GT-C or GT-D. Most GT-A fold GTs are single domain proteins that contain a Rossmann-like fold though exceptions to this rule exist. GT-A GTs also have a DxD (Asp-x-Asp) motif, which is required to coordinate the divalent cation (cofactor). The donor substrates include sugar-linked nucleotide diphosphates that also interact with the cofactor. Within proteins as acceptor substrates for GTs, the most prevalent glycosylated amino acids are serine and threonine (O-linked glycosylation), and asparagine (N-linked glycosylation).

Another type of glycosylation was recently reported from studies of bacterial virulence proteins. Enteropathogenic Escherichia coli (EPEC) and enterohemorrhagic Escherichia coli (EHEC) express numerous effector proteins which are injected into host cells via a type III secretion system (T3SS) to disrupt host cell functions. The T3SS effectors SseK1 and SseK2 from Salmonella enterica have GT-A fold and are single domain proteins that contain a Rossmann-like fold though exceptions to this rule exist. GT-A GTs also have a DxD (Asp-x-Asp) motif, which is required to coordinate the divalent cation (cofactor). The donor substrates include sugar-linked nucleotide diphosphates that also interact with the cofactor. Within proteins as acceptor substrates for GTs, the most prevalent glycosylated amino acids are serine and threonine (O-linked glycosylation), and asparagine (N-linked glycosylation).

Results
Anomeric configuration of glycosylated peptides. Recently it was proposed, though not experimentally demonstrated, that SseK3 is a retaining GT. However, and in contrast to this, another NleB study synthesized Arg-N-GlcNAc-containing glycopeptides in a β-configuration, implying that these enzymes are inverting GTs. To resolve these discrepancies, we investigated by NMR spectroscopy the glycosidic bond configuration of a GlcNAc-GAPDH analogue coupling at the anomeric position of the transferred GlcNAc to be 168 Hz, characteristic of an α-linkage (Fig. 1). These data suggest that the transfer of GlcNAc by SseK1 follows a retaining mechanism. Considering the conserved active site residues and the structural similarity between SseK1 and SseK2/SseK3, SseK2/SseK3 also might be retaining GTs (detailed information is described below).

Overall enzyme architecture. We solved the crystal structures of Salmonella enterica serovar Typhimurium SL1344 SseK1 in complex with UDP. SseK2 was solved both in its unliganded form and in complex with UDP and UDP-GlcNAc. NleB2 from E. coli O145:H28 was solved in its unliganded form. (Supplementary Table 1). For overexpression and crystallization, amino acids 1–20 and 1–33 at the N-terminus of SseK1 and SseK2 were truncated, respectively. This N-terminal region is predicted to be unstructured and presumably play a role in secretion and translocation into the host cell. Proteins containing these N-termini failed to crystallize. Point mutations (C39S, C210S) were introduced into SseK1 to prevent protein precipitation due to irregular intermolecular disulfide binding. For NleB2, amino acids 317–326 at the C-terminus were deleted for better crystal packing and Cys21 and Cys199 were also substituted to serine (Fig. 2a). Unambiguous electron density maps for uridine 5′-diphospho-N-acetylglucosamine (UDP-GlcNAc) or uridine 5′-diphosphate (UDP) were visualized in the active sites of the crystal structures (Fig. 2b).

The sequence identity among SseK1, SseK2, and NleB1 ranges from ~60–65%. The N-termini, whose function is presumably to facilitate protein translocation into host cells, differ the most among orthologs, while the rest of the sequence is highly conserved. Hereafter, we will focus primarily on the biochemical study for SseK1 because SseK1 is more active than SseK2. For structural analyses, we focus on SseK2 because the data-sets for this protein were obtained at a higher resolution, with three different snapshots of the active site. SseK2 possesses an overall protein fold composed of 15 α-helices and 9 β-strands that is highly similar to SseK1 and NleB2 (RMSD = 1.8 Å/1.8 Å, Z-score = 37.2/37.4, number of compared residues = 304/304 to SseK1/NleB2, respectively based on DALI pairwise comparison) (Fig. 2c, d). These structures belong to the GT-A class, which has two abutting β/α/β Rossmann-like domains (β3-α2-β4-α3-β5) and contains an Asp-x-Asp (DxD) motif in the active site (SseK2D239-x-D241, SseK1D223-x-D225, and NleB2D218-x-D220).

SseK2 can be divided into three types of sub-domains, namely the catalytic domain (40–147 and 185–336), which includes the

nucleotide and peptide substrates and find common features for the three peptides such as the recognition of the conserved Trp and Arg residues (WR-motif). Finally, molecular dynamics simulations reveal that the presence of GlcNAc in the donor site induces conformational changes on the side chains of the peptide substrate so that the final arginine acceptor becomes properly oriented for a front face attack to the anomeric C1 carbon of the sugar.
Rossmann-like domains, the protruded helix-loop-helix (HLH) domain (148–184), and the C-terminal lid domain (337–348) (Fig. 2e). The concave shape of the catalytic domain is composed of an α-helix and β-strand mixture, and similar to other GT-A structures, continuous central β-strands (β8, β9, β6, β3, β4, β5) form a mixture of parallel and anti-parallel strands. The C-terminal lid domain is highly flexible in the absence of a ligand. Therefore, the electron density map for this domain was not resolved. However, in the structures with UDP and UDP-GlcNAc, the substrate leads to an unambiguous electron density map for the C-terminal lid domain, implying that the domain is well ordered only in the presence of the nucleotide (detailed information is described below).

Recently, the crystal structure of EarP, an arginine rhamnosyltransferase, was solved, revealing a GT-B fold and an inverting catalytic mechanism in which a glutamate residue acts as the catalytic base. Hence, SseK and EarP are likely to differ in their catalytic mechanisms (see below).

**Donor substrate binding mode.** Based on the complexes of SseK2 with UDP and UDP-GlcNAc, we identified the donor-substrate binding mode and substrate-mediated conformational changes. UDP-GlcNAc consists of three groups, namely the uridine, pyrophosphate, and GlcNAc, which will be discussed independently. The uridine group has an aromatic ring tethered by Phe203 and Trp65 through π-π stacking, water-mediated indirect hydrogen bonds (backbones of Arg68 and Ser346), and hydrogen bonding with the Phe66 backbone (Fig. 3a, upper panel). This sandwich-like π-π stacking is an unusual interaction in GTs because in most of them the sugar nucleotide uracil moiety is sandwiched between an aromatic and an apolar non-aromatic residue participating in π-π stacking and CH-π...
interactions, respectively27,28 (Supplementary Fig. 1a), implying that this unusual interaction is not a requirement for GTs that prefer uracil-containing sugar nucleotides. However, this sandwich-like π-π stacking interaction is unique for this family of enzymes and is determinant for recognition of the uracil moiety (see below).

Both SseK1 and SseK2 share sandwich-like aromatic π-π stacking interactions and both the tryptophan and phenylalanine residues are conserved in NleB2. However, in the sandwich-like π-π stacking, the interaction modes of SseK1 and SseK2 are slightly different. In contrast to SseK2, Trp331 from the C-terminal lid of SseK1 interacts with the uracil base instead of
**Fig. 3** UDP-GlcNAc binding mode in SseK2. 

**a** Uracil moiety of UDP-GlcNAc interacts with SseK2 through hydrogen bonds and π-π stacking (top panel), but SseK1 uses a slightly different mechanism (second and third panel). Uracil binding mode of SseK3 is similar to SseK2 instead of SseK1 (bottom panel). 

**b** GlcNAc moiety of UDP-GlcNAc interacts with Asp204, Arg207, Asp239, and Arg348 by hydrogen bonds. The carbonyl group of the acetyl of GlcNAc interacts with a water molecule to stabilize the divalent metal ion. 

**c** Manganese ion coordinates six oxygens from pyrophosphate, Ser340, Asn338, Asp241, and water. The DxD motif stabilizes both UDP-GlcNAc and manganese ion. *Amino acid numbering in brackets refers to conserved sequence of SseK1. Black dashed lines represent hydrogen bonds.*
Phe203 (Fig. 3a, second and third panels). To confirm this structural difference, we mutated the Phe residue of SseK1 and SseK2 and measured the $K_d$ for UDP-GlcNAc by using isothermal titration calorimetry (ITC) (Supplementary Fig. 1b, c). While the $K_d$ for UDP-GlcNAc to SseK1 F187A was similar to wild type SseK1, the $K_d$ for binding of UDP-GlcNAc to SseK2 F203A was increased about 13-fold as compared to wild type SseK2. These data suggest that although the sandwich-like $\pi$-$\pi$ stacking interactions are maintained, the uracil binding modes within SseK1 and SseK2 are slightly different. In the SseK3 structure, SseK2-like $\pi$-$\pi$ stacking interaction is conserved and Phe190 SseK3 and Trp52 SseK3 (corresponding to Phe203SseK2 and Trp65SseK2) participate in an interaction with the uridine group in the same orientation (Fig. 3a, bottom panel). Mutating Trp51SseK1 and Trp65SseK2 to alanine abrogated UDP-GlcNAc binding to SseK1 and SseK2 (Supplementary Fig. 1b, c). The activity of the W51A mutant was reduced more than the W331A mutant, as measured in NF-κB activation assays, which was consistent with ITC assay data (Supplementary Fig. 1b, c). Overall, Trp51SseK1 and Trp65SseK2 appear to be more critical than Trp331SseK1 and Phe203SseK2 for $\pi$-$\pi$ stacking interactions with UDP-GlcNAc. Note that for all sugar nucleotides the binding energy is dominated by a large negative enthalpic term and to a lesser extent by a non-favored entropic term (Supplementary Fig. 1b, c).

The GlcNAc moiety of UDP-GlcNAc establishes hydrogen bond interactions with Asp204, Arg207, Asp239, and Arg348. The acetyl group of GlcNAc stabilizes the manganese ion by water-mediated hydrogen bonds (Fig. 3b). The importance of the acetyl moiety was confirmed by ITC data that show increased $K_d$ of UDP-glucose (UDP-Glc) and UDP-galactose (UDP-Gal) as compared to the $K_d$ of UDP-GlcNAc for SseK1 (Supplementary Fig. 1b, c). In comparison to UDP-GlcNAc, UDP-Glc lacks the acetyl group, leading to a decrease in the enzyme-substrate binding affinity of about 16.5-fold. Moreover, in UDP-Gal, the absence of the acetyl group and the presence of an inverted C4 hydroxyl group may lead to steric hindrance with the enzyme, leading to the weakest $K_d$ (69.1-fold weaker than the $K_d$ of UDP-GlcNAc). The binding affinities of SseK1 and SseK3 for UDP-Glc and UDP-Gal are relatively different; however, their binding affinities for UDP-GlcNAc are similar (2.3, 1.2, and 1.9 μM for SseK1, SseK2, and SseK3, respectively). Overall, the SseK enzymes possess an architecture that is optimized for binding UDP-GlcNAc.

Most GT-A GTs have a DxD motif that is required for enzymatic activity. The DxD motif in SseK2 has two significant functions, the coordination of the manganese ion and the interaction with the GlcNAc group (Fig. 3c). The manganese ion acts as a bridge between SseK2 and the pyrophosphate of UDP-GlcNAc. In the absence of manganese, the DxD motif-mediated donor-substrate binding would not be expected to occur due to the negative charge repulsion between DxD and the pyrophosphate of UDP-GlcNAc. An octahedral molecular geometry was visualized for the manganese ion coordinated to six oxygens from the UDP pyrophosphate, Asp241, Ser340, Asn338, and a water molecule. Asp239 interacts with a water molecule and a GlcNAc moiety via hydrogen bonds. Asp241 interacts with both a manganese ion and with a water molecule that stabilizes the manganese ion. Most of the residues that interact with UDP-GlcNAc are highly conserved in the SseK and NleB families (Supplementary Fig. 2).

**Conformational change by the donor-substrate binding.** After donor-substrate binding, several GTs undergo large local conformational changes. For example, human glycogenin1 (hGYG1) has a lid, an acceptor arm, and a C-terminal loop. Their conformational rearrangement influences the accessibility of the substrate at the active site and in turn their catalytic activity. The SseK2 structure undergoes a dramatic conformational change induced by donor-substrate binding. In the ground state of SseK2, the C-terminal lid domain is highly flexible, impeding its visualization in the crystal structure. In this state, the donor-substrate binding site might be fully exposed to allow the access of UDP-GlcNAc (Fig. 4a, left panel). After UDP-GlcNAc binds to the active site, the α1 helix is tilted towards the UDP-GlcNAc by ~3.5° and the C-terminal lid domain covers up the active site to stabilize the bound substrate and to restrict the accessibility of water molecules (Fig. 4a, right panel and Fig. 4b). The closure of the C-terminal lid in the presence of UDP-GlcNAc determines a closed conformation for this family of enzymes. We truncated the lid domain (SseK1 1-321) and found that the $K_d$ of UDP-GlcNAc binding increased about 155.3-fold for the lid domain truncation, as compared to the wild type protein (Supplementary Fig. 1b, c), suggesting that the C-terminal lid domain plays a key role for donor-substrate binding.

Both Trp334 and Arg335 in the C-terminal lid domain of SseK3 (corresponding to Trp347SseK2 and Arg348SseK2) interact with UDP-GlcNAc. This interaction is highly similar to that of SseK2. It was reported that Trp334 and Arg335 in SseK3 are essential for enzyme activity. The amino acid sequence of the C-terminal lid domain of SseK1 is slightly different from SseK2 and SseK3. The conserved arginine residue in SseK2 and SseK3 (Arg348SseK2, Arg335SseK3), which is located in the lid domain, is substituted to an alanine residue in SseK1 (Ala332SseK1). In addition, an arginine residue is located next to an alanine residue (Arg333SseK1). In the crystal structure of UDP-bound SseK1, the backbone direction of Ala332 is located at the opposite side of UDP-GlcNAc (Fig. 4c). This implies that the lid domain of SseK1 is likely more flexible than SseK2 and SseK3. Furthermore, the sequence alignment shows that the NleB family lacks the arginine residue, though a conserved Trp is present at the C-terminus. We would expect that both the SseK and NleB families have a different conformational behavior of the lid-domain.

**Peptide substrate recognition by SseK1 and SseK2.** To obtain structural information on the molecular recognition of the substrates, we performed saturation transfer difference (STD) NMR experiments using short peptides from FADD, TRADD, and GAPDH. Standard homo- and heteronuclear 2D NMR techniques were used to obtain the chemical shift assignments of GAPDH195-203, FADD110-118, and TRADD229-237 (Supplementary Table 5–7). For each peptide, four different enzyme systems were prepared: apo-SseK1, apo-SseK2, holo-SseK1, and holo-SseK2, where apo and holo stand for the enzyme without and with Mn+2 and UDP, respectively. We observed that all three peptides bound to both SseK1 and SseK2, irrespectively of the forms used in the experiments (Supplementary Fig. 3, 4). These data imply that binding of the short peptide ligands occurs independently of enzymatic activity and can also take place in the absence of the sugar nucleotide. In STD NMR experiments, strong signal intensities from different hydrogen atoms of the ligand permit identification of the main contacts of the peptides with the enzyme in the bound state. After intensity normalization, binding epitope maps of the peptides were obtained (Fig. 5 and Supplementary Fig. 5). In all cases, high STD signals, indicating close contacts, were observed for the conserved Trp and Arg side chains. The results support the concept that a WR-motif (W112/R113 in FADD, W230/R231 in TRADD, and W196/R197 in GAPDH) appears to be central for recognition. For TRADD229-237, although rather similar binding modes to SseK1 and SseK2...
were observed, the binding epitope was spread across the entire molecule when bound to SseK1 but was more concentrated around the WR-motif for SseK2. The epitopes of FADD\textsubscript{110-118} are comparable when bound to either SseK1 or SseK2. For GAPDH\textsubscript{195-203}, differences in binding to both enzymes were observed. For the binding of GAPDH\textsubscript{195-203} to SseK1, the data support a significant conformational rearrangement of the peptide ligand upon addition of Mn\textsuperscript{2+} and UDP. This was evidenced by a significant change in the binding epitope mapping, particularly a large increase in STD intensities for the arginine side-
chains (Fig. 5c, d). However, for all other peptides, and for GAPDH195-203 binding to SseK2, no such rearrangement occurs (Supplementary Fig. 6).

To elucidate the specific role of the WR-motif, we measured the kinetics of SseK1 to four GAPDH187-203-derived synthetic peptides (designated as WT, W196A, R197A, and W196A/R197A). Each of the W196A, R197A, and W196A/R197A mutant forms decreased the catalytic efficiency of about 40.5%, 47.4%, and 17.3%, respectively, as compared to WT GAPDH187-203 peptide (Supplementary Fig. 16b). In particular, the double mutant form (W196A/R197A) synergistically decreased enzyme catalysis, supporting our STD-NMR experiments that suggested that the WR-motif of the peptides is of utmost importance for binding to these enzymes.

To investigate the relevance of the WR-motif for binding, we also carried out STD NMR experiments focused on analyzing the impact of single and double mutations on the affinity of the molecular recognition of the synthetic peptide TRADD229-237. As we were interested only in analyzing the impact on binding, we ran competition experiments for the interactions of the TRADD229-237-derived peptides with SseK1. Five synthetic peptides (designated as WT, W230A, R231A, W230A/R231A, and R235A) were analyzed. All of them bound to SseK1, as detected by STD NMR, yet their affinities were different, as reflected in their differences in average STD NMR intensities (e.g., the most intense alpha proton showed 14%, 9%, 6%, 4%, and 10% for WT, W230A, R231A, W230A/R231A, and R235A, respectively). This result indicates that the highest affinity for SseK1 is achieved when the full WR-motif is present. Again, the results were compatible with the double mutant showing the lowest binding affinity. We then confirmed the differences in affinity compared to the WT peptide by performing competition STD NMR experiments. In binary samples containing SseK1 and equimolar concentrations of the TRADD229-237 and one of the mutant TRADD peptides, none of the mutants was able to significantly displace the WT peptide, which demonstrates that modifications at the WR-motif impact negatively the affinity of the peptide for the enzyme (Supplementary Figs. 19, 20).

STD NMR data revealed that all the peptide ligands were recognized in solution. Hence, it is clear that differences in glycosylation of full-length FADD, TRADD, and GAPDH substrates by SseK1 and SseK2 are not due to differences in binding modes of their death domain sequences, but instead due to differences outside the binding site. In agreement with the similarity of binding modes of the short peptides detected by STD NMR, most of the sequence differences of SseK1 and SseK2 are likely in regions away from the binding site, including the HLH domain. Hence, differences in glycosylation specificity may be attributed to differences in the internal dynamics between the two enzymes at those distinct regions.

To test that hypothesis, we ran long (800 ns) Gaussian accelerated molecular dynamics (GaMD) simulations of SseK1 and SseK2. Principal component analysis (PCA) showed that, in both cases, the motions of largest amplitude are indeed present around the HLH domain (Supplementary Fig. 7e, f), primarily due to rotation of the HLH towards the binding site (Supplementary Fig. 7a, b). Noticeably, the simulations showed that SseK1 is significantly more flexible than SseK2 in the loop region connecting the HLH (Supplementary Fig. 7e, f). Additionally, significant differences were observed at the tip of the HLH, as SseK2 exhibited a substantial tilting motion towards the binding site (Supplementary Fig. 7c, d). These data reveal that there exist significant differences in the dynamics of the HLH domain between SseK1 and SseK2 that affect the HLH domain approach towards the substrate binding pocket, which could explain the differences in substrate specificity of these two enzymes and in turn, glycosylation.

To understand further the molecular basis of substrate peptide recognition we generated a ternary SseK2:UDP-GlcNAc:FADD110-118 complex using an induced fit molecular docking protocol. Since this ternary complex was not accessible experimentally neither by X-ray crystallography nor by NMR spectroscopy, molecular modeling provides the only insight into the structure of the full complex in the presence of both the donor sugar nucleotide and the acceptor. Peptide structure prediction and NMR chemical shift indexing indicated that the peptide remained in its native helical conformation (Supplementary Fig. 8). Docking of FADD110-118 was only possible using the SseK2 structure with the C-terminal lid in the open conformation, since the closed lid precludes access to the binding site. The resulting model was in good agreement with STD NMR data, with Trp112FADD and Arg113FADD in close proximity to the protein surface (Supplementary Fig. 9a, b). In addition, the sidechain of Arg113FADD was also found in close proximity to His260, Glu271, and Asn338, as observed in the docking of the acceptor Arg to the SseK2-UDP-GlcNAc crystal structure (Supplementary Fig. 9c and see below).

Furthermore, it was possible to graft the published 3D structure of the full-length FADD protein (PDB 3EZQ) into our SseK2:UDP-GlcNAc:FADD110-118 model complex without any significant atomic overlap (Supplementary Fig. 10). We wanted to investigate whether the presence of the GlcNAc ring in the donor substrate might have an impact on the acceptor binding mode, and to analyze the dynamics of the full ternary SseK2:UDP-GlcNAc:FADD complex. We subjected this complex to a 500 ns Gaussian GaMD simulation. PCA showed that the most significant motions involved rotation and translation (Supplementary Fig. 11a, b) of both the SseK2 HLH and the FADD C-terminal a-helix towards one another. A distinct energy minimum was observed at short inter-helical distances (Supplementary Fig. 11c). At this minimum, a clear intermolecular complementarity was observed. In particular, electrostatic interactions were observed between Lys176SseK2 and Asp175FADD, and between Asp180SseK2 and Arg166FADD (Supplementary Fig. 12a). Leu172FADD interacts with a hydrophobic patch defined by Val169SseK2 and Leu176SseK2. The a3-helix of FADD is highly negatively charged and interacts closely with Lys264, Arg348, and the manganese ion of SseK2 (Supplementary Fig. 12b). Our simulation suggests that Asp123 of FADD directly coordinates the manganese ion. Finally, along the 500 ns of the
Fig. 5 Binding modes of short peptide substrates. Binding epitope mappings of a FADD_{110-118} b TRADD_{229-237} and c, d GAPDH_{195-203} peptides in the presence of 25 μM SseK1. Samples in a, b and d contained 25 μM Mn^{2+}, and 25 μM UDP. All STD intensities normalized against H\_2 of the tryptophan. Colored circles represent magnitude of normalized intensities (blue:< 40%, pink: 40–70%, red:> 70%). Comparison of GAPDH_{195-203} binding to SseK1, c in the absence, and, d in the presence of Mn^{2+} and UDP, reveals a significant change in the binding mode of the substrate peptide upon addition of the cofactor and the nucleotide diphosphate. For STD NMR study of binding to SseK2 see Supplementary Fig. 5.
GaMD simulations, significant conformational rearrangements of the arginine side chains in the FADD110–118 region were observed (Supplementary Fig. 13a–d). On average, Arg117 of FADD is the residue from FADD showing the shortest distance to the C1 anomeric carbon of UDP-GlcNAc, and is the only one simultaneously establishing close contacts with His260, Glu271, and the beta-phosphate of UDP-GlcNAc.

Together, these data provide evidence for specific recognition of FADD by SseK2 via interactions far from the active site, which may provide insight into differences in substrate glycosylation specificity of SseK1 and SseK2. These data also suggest that Arg117 is the best oriented residue from FADD for accepting the transferred glycosyl from the donor UDP-GlcNAc, in keeping with the role of this residue as the only acceptor site in FADD3,4.

**Catalytic importance of HEN motif.** In the SseK2-UDP-GlcNAc crystal structure (chain D), Asn272 (corresponding to Asn256seK1) is closer to the C1 anomeric carbon of UDP-GlcNAc (5.1 Å) as compared to Glu271 (5.6 Å). These data indicate that the Asn272 may play an important role in binding and catalysis. In addition, Glu271 is located at the entrance of the putative acceptor arginine-binding site. To obtain insight into the importance of these residues in substrate recognition, we conducted in-silico docking of the acceptor Arg to the SseK2-UDP crystal structure. It was observed that a closed conformation of SseK2 structure possesses a putative acceptor substrate binding site pocket that is connected to the anomeric carbon of GlcNAc. We used a closed conformation of SseK2 with a putative acceptor substrate binding site pocket connected to the anomeric carbon of GlcNAc for automated computational docking (Discovery Studio, Accelrys). As a result, the negatively charged pocket of the concave active site interacts with the positively charged guanidine group. In particular, Glu271, the beta-phosphate of UDP, and His260 were located in a position suitable for hydrogen bonding (Supplementary Fig. 14a). All of these residues are highly conserved in the SseK and NleB families (Supplementary Fig. 2).

In the docking structure of SseK2, a single negatively charged residue (Glu271SseK2) and an additional beta-phosphate from UDP are in close contact to the guanidinium group of the acceptor Arg. The Glu271SseK2 corresponds to the Glu253NleB1 whose mutation to Ala did not inhibit NF-kB signaling, in agreement with the importance of this residue in glycosylation of the acceptor Arg. Further studies are required to determine its precise role either in catalysis or binding (see below).

In addition, His260SseK2 is located near the guanidinium group of the acceptor Arg. A pH activity profile for both SseK1 and NleB1 revealed that NleB1 and SseK1 have an optimal pH between 6.0–8.0 and 6.0–8.5, respectively (Supplementary Fig. 14b). This highlights the potential role of histidine as a catalytic base because the pKₐ value of histidine is ~6.0. To test the role of these residues in substrate recognition and catalysis, we mutated His260, Glu271, and Asn272, which together form the ‘HEN’ motif. This motif is highly conserved at both the primary sequence and tertiary structure levels (Supplementary Fig. 15). Wild-type (WT) SseK1 and the HEN motif single mutants were overexpressed with TRADD in HEK293T cells. WT SseK1 inhibited PARP production (Fig. 6a). As expected, the DxD-AxA double mutant, resulted in an increase in cPARP level. The HEN motif single mutants (His260, Glu271, and Asn272 in SseK2) led to an increase in the cPARP levels, a similar outcome to the DxD-AxA double mutant. When we observed the oligomerization form in a non-reducing gel, TRADD oligomer was detected in the mutant forms, suggesting that each mutant loses its glycosylation activity and fails to inhibit TRADD oligomerization.

NF-kB activity data in A549 cells correlated highly with the results from the PARP cleavage assay (Fig. 6b). Surprisingly, NF-kB levels increased more in H244A, E255A, and N256A than in the AxA mutant. Furthermore, we also investigated enzyme kinetics using the L-arginine substrate and the purified recombinant SseK1 and SseK2 (Fig. 6c, Supplementary Fig. 16a). The catalytic activities of the mutants (H244A/H260A, E255A/E271A, and N256A/N272A in SseK1/SseK2) decreased significantly compared to WT, though they were not essential for activity. The H244A/H260A and N256A/N272A mutants showed lower catalytic activity than the DxD-AxA double mutant.

We also studied the glycosylation activity of WT and HEN mutant enzymes in vitro, in cell culture, and in mouse infection experiments using C. rodentium. We first incubated NleB1, NleB2, SseK1, or SseK2 with FADD, TRADD, or GAPDH. We observed that, consistent with previous studies,7,18, WT NleB1, SseK1, and SseK2 glycosylated FADD (Fig. 6d). None of the point mutations in the HEN motif of any of the enzymes retained the ability to glycosylate FADD (Fig. 6d). Similar data were observed in studies of TRADD and GAPDH glycosylation (Fig. 6e, f). We also noticed that NleB1 and SseK1 are self-glycosylated, though the functional importance of this modification is unknown.

To extend these data, we conducted a series of mouse challenge experiments with Citrobacter rodentium. C. rodentium has only 1 copy of NleB, which functions similarly to EHEC NleB118. To evaluate whether the HEN motif of NleB is important to C. rodentium virulence, we deleted the C. rodentium nleB gene and then complemented this mutant with different EHEC nleB1 and nleB2 expression plasmids. We infected mice with C. rodentium ΔnleB strains expressing either WT NleB1 (ΔnleB/pnleB1) or the mutant enzymes H242A, E253A, and N254A. Mice infected with ΔnleB C. rodentium showed an approximately 100-fold reduction in colonization magnitude after 14 days, as compared with WT C. rodentium, in support of previous findings7 (Fig. 6i). While this mutant was fully complemented by expressing WT EHEC NleB1, none of the HEN mutants complemented the colonization defect (Fig. 6i). Additionally, neither WT EHEC NleB2 nor an NleB2 mutant in which all HEN amino acids were mutated to alanines complemented the colonization defect. These data demonstrate that the HEN motif is highly important for enzymatic activity and virulence.

**Proposed mechanism.** Three reaction mechanisms have been proposed for retaining GTs, namely SN2, SNi, and orthogonal mechanisms. While the SN2 mechanism involves a double-displacement reaction requiring a nucleophilic residue to form a covalent glycosyl intermediate34, both the SNi and the orthogonal mechanisms involve a single displacement reaction in which the beta-phosphate of the nucleotide acts as the catalytic base35–37. However, the SNi and the orthogonal mechanisms differ in their reaction profiles and the timing of bond formation and bond breakage37. These differences lead to a dissociative and associative transition state for the SNi and the orthogonal mechanism, respectively37. Unlike the SNi mechanism that has been probed extensively, the SN2 and the orthogonal mechanism have never been demonstrated experimentally34–37. In our crystal structure, Asn256SseK1 was located in a possible position for the back-side attack of the C1 anomeric carbon of UDP-GlcNAc through an SN2 reaction. However, our kinetic assay results showed that this
residue is not essential for catalysis (its mutation to alanine in both SseK1 and SseK2 reduced its $K_{cat}$ by 80 % as compared to the WT enzymes). This result is unexpected because other GTs completely lose their catalytic activity (650-fold to 23,000-fold decrease) when a potential nucleophile residue was mutated to a non-nucleophile residue. It was proposed that Glu258 of SseK3 acts as the nucleophile residue for an $S_N2$ reaction (corresponding to Glu255 in SseK1 and Glu271 in SseK2). However, the E255A/SseK1 and E271A/SseK2 mutants still possess an activity of 70 and 24%, respectively, ruling out its potential role as a nucleophilic residue. Additionally, our GaMD simulations of the grafted complex of SseK2 with FADD supported the relevance of Glu271.
Fig. 6 HEN motif plays a key role in NleB/SseK enzyme activity. a SseK1 mutants were generated and the cellular function in HEK293T cells was investigated. A non-reducing gel (right panel) was used to confirm the presence of the TRADD oligomer. Mutants in red represent mutations of residues proposed to be catalytically important. Data represent at least three repetitions. b The NF-κB level in A549-NF-κB luc cells was measured to check the enzymatic functions. Data represent the mean and standard deviation in triplicate. Multiple comparisons performed by one-way ANOVA followed by Turkey’s Multiple Comparison Test (***P < 0.01, ****P < 0.001 compared to WT). c Enzyme kinetic assays of SseK1 and SseK2, respectively. d In vitro glycosylation of TRADD by NleB1, NleB2 (top panel) and SseK1 and SseK2 HEN mutants (bottom panel). e In vitro glycosylation of TRADD by NleB1, NleB2 (top panel) and SseK1 and SseK2 HEN mutants (bottom panel). f In vitro glycosylation of GAPDH by NleB1, NleB2 (top panel) and SseK1 and SseK2 HEN mutants (bottom panel). g Glycosylation of TRADD after co-transfection with either NleB1 or SseK1 (WT and HEN mutants) in HEK293T cells. FLAG-TRADD was immunoprecipitated and then immunoblotted using an anti-Arg-GlcNAc antibody. h Glycosylation of GAPDH after co-transfection with either NleB1 or SseK1 (WT and HEN mutants) in HEK293T cells. FLAG-TRADD was immunoprecipitated and then immunoblotted using an anti-Arg-GlcNAc antibody. i Colonization (log10 CFUs/g colon) of indicated C. rodentium strains (14 days post-gavage) in C57BL/6J mice (n = 6). Asterisks indicate significantly different colonization magnitude as compared to WT; Kruskal-Wallis test. Uncropped blots are shown in Supplementary Figs. 21 and 22.

Discussion

The overall crystal structures of SseK1 and SseK2 from S. typhi-murium as well as NleB2 from enteropathogenic E. coli are similar, revealing high identities between the amino acids at the active binding site level. These enzymes glycosylate the guanidinium moiety of Arg residues, which are residues with poor nucleophilic character because electrons on this moiety are partially delocalized. In addition, we identified several characteristics in the active site that are compatible with an S_{N}i mechanism and with the role of HEN motif residues in catalysis: (a) we experimentally demonstrate by NMR that these enzymes are retaining GTs; (b) donor-substrate-mediated fixation of the C-terminal lid could shield the active site from the hydrophilic environment, avoiding unwanted hydrolysis of UDP-GlcNAc, a behavior reported in other retaining GT-A fold GTs such as Legionella pneumophila glucosyltransferase^{27} and the toxin B;^{10} (c) mutations of the HEN motif lead to a significant reduction in glycosylation both in vitro and in vivo levels, implying that these residues affect enzyme catalysis. Based upon our kinetic, structural, and computational studies, we suggest that His and Glu might improve the poor nucleophile character of the acceptor Arg guanidinium moiety to facilitate catalysis; and (d) based on our MD simulations, the acceptor Arg is facing the anomeric carbon, which is compatible with a front face mechanism for glycosyl transfer with retention of anomeric configuration.

Each HEN residue of SseK3 is on the β-strand, but in the case of SseK1, SseK2, and NleB2, the HEN residues are located on the loop structure. This has implications for the differences in the activity of the enzymes due to the HEN motif because the loop structure is more flexible than the β-strand. In addition, there are differences in the regulation of the C-terminal lid domain between SseK1, SseK2, and SseK3. The arginine on the C-terminus participates in an interaction with UDP-GlcNAc in SseK2 and SseK3, but not in SseK1. Based on this result, the departure of the leaving group is likely to be easier, due to a weaker enzyme-substrate interaction network in SseK1, than in SseK2 and SseK3. We suggest that this may also be a reason for the differences in enzyme activity in the SseK family.

Although the tertiary structure of NleB1 is likely similar to NleB2, SseK1, and SseK2, a previous study has reported that NleB2 has a lower activity than NleB1 on the same target (TRADD)^{6}. In this study, we also demonstrate that the enzymatic activity of SseK1 is ~62 times higher than that of SseK2 based on enzyme kinetic assays. The substrate specificities of the NleB/SseK family of C. rodentium, Escherichia coli and Salmonella enterica are different^{18}. Based on these differences, we can infer from the crystal structures that these discrepancies between orthologs of this family might be attributed partially to the HLH domain, which is a structural feature not present in other GTs (Supplementary Fig. 18a). The amino acids of this domain are not conserved and are structurally flexible (Supplementary Fig. 18b), as confirmed by the long GaMD simulations (Supplementary Fig. 7). The flexible HLH domain is close to the active site, which indicates this domain may be involved in the recognition of the acceptor protein substrates containing death domains. In fact, GaMD simulations of the grafted SseK2:UDP-GlcNAc:FADD complex show that the HLH domain interacts directly with FADD C-terminal α-helix through side chain complementarity (Supplementary Fig. 12). These data support the role of the HLH domain in the recognition of the acceptor protein substrates containing death domains.

STD NMR spectroscopy revealed that, in contrast to their enzymatic activity profile, both SseK1 and SseK2 interact with all short peptides from FADD, TRADD, and GAPDH (Fig. 5, Supplementary Fig. 5). Each of these peptides contains a conserved WR-motif, which forms a key structural requirement for binding to the enzymes, as revealed by STD NMR spectroscopy and molecular modeling. Therefore it appears that recognition of death domains may be due to complementarity with the active site, whilst distinction between different death domain-containing proteins is mediated through interactions far from the active site. Overall our data provide compelling evidence of the molecular basis for Arg glycosylation, the differences in substrate specificity among orthologs, and will provide a framework for the design of pan-NleB/SseK inhibitors targeting enteric pathogens.

Methods

Protein purification. SseK1(121-336) gene was generated and amplified by PCR from synthesized DNA and cloned into a modified pET28a (Novagen) in which the thrombin cleavage site was replaced with a tobacco tech virus (TEV) protease cleavage site. SseK2 (34-348) and NleB2 (1-316) gene were generated and amplified by PCR from Salmonella typhimurium (strain SL1344), Escherichia coli O157:H7 (strain RIM12581) respectively and cloned into the pVFmTS vector (Korean patent 10-0690230), which has 6xHis-thioredoxin (Trx) and TEV protease cleavage site (Supplementary Table 2, Supplementary Table 3). PCR-based site-directed
mutagenesis was employed to generate various point mutations. Complete amino acid sequences are shown in Supplementary Table 4. Each sub-cloned plasmid was transformed into E. coli BL21(DE3) (Novagen) and grown in high salt Luria-Broth medium. When the O.D600 reached 0.6 ~ 0.8, the temperature was decreased to 17 °C and the culture was induced with 0.3 mM IPTG (isopropyl-1-thio-D-galactopyranoside). After 16 h incubation, each protein was purified using nickel-affinity chromatography. Cell was lysed using lysis buffer (20 mM Tris-HCl pH 7.5, 300 mM NaCl, 30 mM imidazole, 10% glycerol) and the proteins were eluted using an elution buffer of 300 mM imidazole in lysis buffer. Thereafter, the TEV recognition site was cleaved using TEV protease. After desalting to 20 mM Tris-HCl (pH 7.5), 50 mM NaCl, each protein was loaded into an anion-exchange chromatography column (Hitrap-Q, GE healthcare) and then gel-filtration chromatography (Superdex-200, GE healthcare) and then gel-filtration chromatography. MOLREP, REFMAC5, and COOT were used, using the particle mesh Ewald to calculate electrostatics.

### Molecular docking calculations for FADD-SseK2
Crystal structures of SseK1, SseK2, and FADD (PDB 3EZQ) were imported into Schrödinger Maestro 43 and further modeled, respectively. Native NleB2 and SseK2 were diffraction data sets were processed and scaled with the programs immolim and Aimless from the CCP4 program suite. The phasing information was solved by SAD method from SseK2 derivative NleB2 crystal using AutoSol program and the other proteins were solved by molecular replacement using NleB2 structure. MOLREP, REFMAC5, and COOT were used for molecular replacement, structure refinement, and further modeling, respectively. All figures were prepared using PYMOL.

### Molecular dynamics
UDP charges for use with UDP-GlcNAc were derived using the RESP fitting method implemented on the RED server. The UDP fragment was generated by replacing the GlcNAc with a methyl group. In accordance with GLYCAM 43, the HF/3-31G* level of theory was used with a weight factor of 0.01 and all aliphatic protons were constrained to a charge of 0. The total charge of the UDP fragment was set to −2. The charge of the methyl group was set to 0.194 before removing to give a final fragment with net charge −2.194, in keeping with the polarity of GLYCAM.

Molecular dynamics simulations of SseK1, SseK2, and SseK2:FADD complex were performed using the Amber PMEMD software 42. Protein atoms were parameterized using the Amber ff14SB forcefield and the Min 24 ion model was used using 12-6 L-J type potentials (Ambertools 1.6, 1264_tdp3p). UDP-GlcNAc was parameterized with GLYCAM 06j and GAFF. Each system was solvated in a truncated octahedral box of TIP3P water, with at least 10 Å between the solute and the edge of the box, before neutralizing with Na+ ions. The system was minimized using the conjugate gradient algorithm, converging on a threshold of 10−4 kcal mol−1 Å−1, first with 20 kcal mol−1 Å−2 restraints on solute atoms, before repeating with no restraints. The system was slowly heated to 310 K over 500 ps (NVT), before equilibrating the pressure to 1 atm (NPT) over a further 500 ps. In both cases with 20 kcal mol−1 Å−2 restraints were used on solute atoms. These restraints were then slowly released over 800 ps before performing Gaussian accelerated molecular dynamics (GAMD) simulations (800 ns SseK1/SseK2, 500 ns SseK2:FADD complex), as implemented in AMBER, using a boost potential on both the dihedral and total potential energies. Here, the simulation was split into 4 distinct stages. First, conventional dynamics were run for 2 ns to automatically calculate an initial boost potential. The calculated boost potential was then applied and equilibrated for 400 ps before adapting to the boost potential. Resulting boost potential was then fixed before performing production dynamics for 800 ns (SseK1/SseK2) or 500 ns (SseK2:FADD complex), saving coordinates every 100 ps. In all cases, the SHAKE algorithm was used to restrain all bonds involving hydrogen, allowing, for a time step of 2 fs. A Langevin thermostat was used with a collision frequency of 5 ps−1 and the barostat used an isotropic Berendsen method.

### Cell culture
Human embryonic kidney (HEK) 293T (ATCC, ATCC® CRL-3126®) and A549 NF-kB luciferase cells (Panomics, RC0002) were cultured in DMEM
supplemented with 10% fetal bovine serum (FBS, Cellgro), 100 U/ml penicillin, 100 μg/ml streptomycin, and 2 mM L-glutamine at 37 °C in 5% CO2.

**Western blot analysis and immunoprecipitation.** HEK293T cells were transfected with various combinations of plasmids using Lipofectamine 2000 (Invitrogen) as specified by the manufacturer. Cells were washed with phosphate-buffered saline (PBS) and lysed in 1 × RIPA buffer (GenDEPOT, Barker, TX, USA) containing 150 mM NaCl, 1% Triton X-100, 1% deoxycholic acid sodium salt, 0.1% sodium dodecyl sulfate (SDS), 50 mM Tris-HCl (pH 7.5), 2 mM EDTA, and a protease inhibitor cocktail. Whole cell lysates (WCLs) were centrifuged at 13,000 rpm for 10 min at 4 °C. To detect TRADD oligomerization, WCLs were separated using a non-reducing sample buffer. About 30 μg of proteins were separated by SDS-PAGE and transferred to nitrocellulose membranes (GE Healthcare, Little Chalfont, UK). Non-specific binding was blocked, and anti-GFP (1:3000, Santa Cruz, CA, USA; SC-8334:9960), anti-actin (1:5000, Cell Signaling Technology, Danvers, MA, USA, #4967), and anti-PARP (1:1000, Cell Signaling Technology, Danvers, MA, USA, #9542) anti-c-Myc (1:3000, Invitrogen, Camarillo, CA, USA, 12-5000) antibodies were used as primary antibodies. After washing, membranes were probed with the HRP-conjugated secondary antibody for 1 h. Enhanced chemiluminescent substrate (GenDEPOT, Barker, TX, USA) was used for visualization.

Immuno precipitation was performed with 15 μl of dynabeads protein G (Invitrogen, Camarillo, CA, USA). The beads were washed and incubated with 1 μg of the antibody for 1 h at RT. The beads were incubated with 300 μg WCL over night at 4 °C after washing. Samples were separated using SDS-PAGE for immunoblottting. UDP-GlcNAc (1 mM) and MnCl2 (5 mM) were added to recombinant SseK1 and incubated at 37 °C for 1 h. The same amount of wild type and auto-glycosylated SseK2 were loaded into 15% SDS-PAGE gel and anti-GlcNAc antibody (1:5000, CTD110.6, Santa Cruz, CA, USA, #sc-59623 used to detect GlcNAcylated arginine).

**NF-kB luciferase assay.** A549 cells (Panomics, RC0002) stably expressing NF-kB were transfected with a mixture of plNLL.1T[PK/Nlc/TK], GFP-TRADD and various Myc-SseK1 plasmids. plNLL.1T[PK/Nlc/TK] was used for transfection control. After 24 h, the cells were treated with 20 ng/ml TNF-α for 6 h. Luciferase assay was performed using Dual-Luciferase Reporter Assay system (Promega, E1910). Briefly, cells were lysed with Cell Culture Lysis Reagent and Luciferase Assay Reagent II was added to measure the luciferase activity. Stop & Glo Reagent was added into tube to quench firefly luciferase activity, and nano.luc luciferase activity was measured using a luminometer.

**Glycosyltransferase kinetics assay.** Recombinant SseK1, SseK2 and point mutant proteins were prepared as described and GT kinetics were measured using UDP-GlcNAc as glycosyl acceptor. Enzyme reaction buffer (ERB) was prepared as 25 mM Tris-HCl [pH 7.5], 50 mM NaCl, 1 mM MgCl2, and 1 mM MnCl2, sample fractions corresponding to the single UV280 peak were collected and concentrated to 0.2 mM by using 10 kDa cut-off Amicon tubes. Ligands (4 mM) were dissolved in the same SEC buffer and were titrated to variant SseK proteins at 25 °C. Binding stoichiometry, entropy variation, dissociation constant and Chi-square values were calculated using MicroCal Origin software.

**Data availability** Coordinates and structure factors have been deposited in the Protein Data Bank under the accession codes 5HSY (NleB X-ray structure), 5H60 (UDP bound SseK1 X-ray structure), 5H61 (SseK2 X-ray structure), 5H62 (UDP bound SseK2 X-ray structure) and 5H63 (UDP-GlcNAc bound SseK2 X-ray structure). Other data are available from the corresponding authors upon reasonable request.

Received: 13 April 2018 Accepted: 19 September 2018

Published online: 16 October 2018

**References**

1. Haltiwanger, R. S. & Lowe, J. B. Role of glycosylation in development. Annu. Rev. Biochem. 73, 491–537 (2004).
2. Glotter, T. M. Advances in understanding glycosyltransferases from a structural perspective. Curr. Opin. Struct. Biol. 28, 131–141 (2014).
3. Zhang, H. et al. The highly conserved domain of unknown function 1792 has a distinct glycosyltransferase fold. Nat. Commun. 5, 4339 (2014).
4. Lairson, L. L., Henriassat, B., Davies, G. J. & Withers, S. G. Glycosyltransferases: structures, functions, and mechanisms. Annu. Rev. Biochem. 77, 521–555 (2008).
5. Pearson, J. S. et al. A type III effector antagonizes death receptor signalling during bacterial gut infection. Nature 501, 247–251 (2013).
6. Li, S. et al. Pathogen blocks host death receptor signalling by arginine GlcNAcylation of death domains. Nature 501, 242–246 (2013).
7. Gao, X. et al. NleC, a bacterial effector with glycosyltransferase activity, targets GDPAPF function to inhibit NF-κappaB activation. Cell Host. Microbe 13, 87–99 (2013).
8. Kaper, J. B., Nataro, J. P. & Mobley, H. L. Pathogenic Escherichia coli. Nat. Rev. Microbiol. 2, 123–140 (2004).
9. Hauser, A. R. The type III secretion system of Pseudomonas aeruginosa: infection by injection. Nat. Rev. Microbiol. 7, 654–665 (2009).
10. Park, H. H. et al. The death domain superfamily in intracellular signaling of apoptosis and inflammation. Annu. Rev. Immunol. 25, 561–586 (2007).
11. Pearson, J. S., Zhang, Y., Newton, H. J. & Hartland, E. L. Post-modern pathogens: surprising activities of translocated effectors from E. coli and Legionella. Curr. Opin. Microbiol. 23, 73–79 (2015).
12. Grishin, A. M. et al. NleH degrades type III effector substrates. J. Biol. Chem. 288, 21143–21149 (2013).
13. Elliott, S. J. et al. The complete sequence of the locus of enteroctye effacement (LEE) from enteropathogenic Escherichia coli E2348/69. Mol. Microbiol. 28, 1–4 (1998).
14. Garimenda, J., Frankel, G. & Crepin, V. F. Enteropathogenic and enterohemorrhagic Escherichia coli infections: translocation, translocation, translocation. Infect. Immun. 73, 2573–2585 (2005).
15. Gao, X. et al. Citrobacter rodentium Nle8 inhibits tumor necrosis factor (TNF) receptor-associated factor 3 (TRAF3) ubiquitination to reduce host type I interferon production. J. Biol. Chem. 291, 18232–18238 (2016).
16. El Qaidi, S. et al. NleH/SseK effectors from Citrobacter rodentium, Escherichia coli, and Salmonella enterica display distinct differences in host substrate specificity. J. Biol. Chem. 292, 11423–11430 (2017).
17. Esposito, D. et al. Structural basis for the glycosyltransferase activity of the Salmonella effector SseK3. J. Biol. Chem. 293, 5064–5078 (2018).
18. Pan, M. et al. Synthesis of and Specific Antibody Generation for Glycopeptides with Arginine N-GlcNAcylation. Angew. Chem. Int. Ed. 53, 14517–14521 (2014).
19. Ghosh, P. Process of protein transport by the type III secretion system. Microbiol. Mol. Biol. Rev. 68, 771–795 (2004).
20. Choy, S. L. K. et al. SseK1 and SseK2 are novel translocated proteins of Salmonella enterica serovar typhimurium. Infect. Immun. 72, 5115–5125 (2004).
21. Holm, L. & Lasko, M. Dali server update. Nucleic Acids Res. 44, W351–W355 (2016).
22. Sengoku, T. et al. Structural basis of protein arginine rhamnosylation by glycosyltransferase EarP. Nat. Chem. Biol. 14, 368–374 (2018).
25. Zhang, X., Zhou, L. & Cheng, X. D. Crystal structure of the conserved core of protein arginine methyltransferase PRMT3. EMBO J. 19, 3509–3519 (2000).

26. Rust, H. L., Zurita-Lopez, C. I., Clarke, S. & Thompson, P. R. Mechanistic Studies on Transcriptional Coactivator Protein Arginine Methyltransferase 1. Biochemistry 50, 3332–3345 (2011).

27. Hurtado-Guerrero, R. et al. Molecular mechanism of elongation factor 1A inhibition by a Legionella pneumophila glycosyltransferase. Biochem. J. 426, 281–292 (2010).

28. Lira-Navarrete, E. et al. Dynamic interplay between catalytic and lectin domains of GaNAC-transferases modulates protein O-glycosylation. Nat. Commun. 6, 6937 (2015).

29. Wiggins, C. A. R. & Munro, S. Activity of the yeast MNN1 alpha-1,3-mannosyltransferase requires a motif conserved in many other families of glycosyltransferases. Proc. Natl Acad. Sci. USA 108, 21028–21033 (2011).

30. Mayer, M. & Meyer, B. Characterisation of ligand binding by saturation transfer difference NMR spectroscopy. Angew. Chem. Int. Ed. 38, 1784–1788 (1999).

31. Mayer, M. & Meyer, B. Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. J. Am. Chem. Soc. 123, 6108–6117 (2001).

32. Miao, Y. L., Feher, V. A. & McCammon, J. A. Gaussian accelerated molecular dynamics: unbiased enhanced sampling and free energy calculation. J. Chem. Theory Comput. 11, 3584–3595 (2015).

33. Albesa-Jove, D., Sainz-Polo, M. A., Marina, A. & Guerin, M. E. Structural Snapshots of alpha-1,3-galactosyltransferase with native substrates: insight into the catalytic mechanism of retaining glycosyltransferases. Angew. Chem. Int. Ed. 56, 14853–14857 (2017).

34. Lira-Navarrete, E. et al. Substrate-guided front-face reaction revealed by combined structural snapshots and metadynamics for the polypeptide N-acetylgalactosaminyltransferase 2. Angew. Chem. Int. Ed. 53, 8206–8210 (2014).

35. Ardevol, A. & Rovira, C. The molecular mechanism of enzymatic glycol transfer with retention of configuration: evidence for a short-lived oxocarbenium-like species. Angew. Chem. Int. Ed. 50, 10897–10901 (2011).

36. Sevastyanov, R., Evans, S. V. & Fyles, T. M. Geometric attributes of retaining glycosyltransferase enzymes favor an orthogonal mechanism. PLoS One 8, e71077 (2013).

37. Blacker, R. J. et al. Glycosyltransfer in mutants of putative catalytic residue Glu303 of the human ABO(H) A and B blood group glycosyltransferases GATA and GTB proceeds through a labile active site. Glycobiology 27, 370–380 (2017).

38. Zhang, Y. et al. Roles of individual enzyme-substrate interactions by alpha-1,3-galactosyltransferase in catalysis and specificity. Biochemistry 42, 13512–13521 (2003).

39. Reintert, D. J., Jank, T., Aktories, K. & Schulz, G. E. Structural basis for the function of Clostridium difficile toxin B. J. Mol. Biol. 351, 973–981 (2005).

40. Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. iMOSFLM: a new graphical interface for diffraction-image processing with automatic beam centre and spot picking. J. Appl. Crystallogr. 42, 747–753 (2009).

41. Halgren, T. A. et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J. Med. Chem. 47, 1739–1749 (2004).

42. Jacobson, M. P. et al. A hierarchical approach to all-atom protein loop prediction. Proteins 55, 351–367 (2004).

43. Olsson, M. H. M., Sondergaard, C. R., Rostkowski, M. & Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pK(a) Predictions. J. Chem. Theory Comput. 7, 525–537 (2011).

44. Vanquem, E. et al. In silico server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. Nucleic Acids Res. 39, W511–W517 (2011).

45. The PyMOL Molecular Graphics System, Version 1.8. (Schrodinger, LLC, 2015).

46. NATURE COMMUNICATIONS | DOI: 10.1038/s41467-018-06680-6 ARTICLE

© The Author(s) 2018

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.