Design of a high-affinity carbohydrate binding protein

Dissertation

zur Erlangung des akademischen Grades
"doctor rerum naturalium" (Dr. rer. nat.)
in der Wissenschaftsdisziplin "Biochemie"

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät der Universität Potsdam
Physikalische Biochemie

von
Ruth Sonja Kunstmann

Potsdam, den 28. April 2017

Published online at the
Institutional Repository of the University of Potsdam:
URN urn:nbn:de:kobv:517-opus4-403458
http://nbn-resolving.de/urn:nbn:de:kobv:517-opus4-403458

Declaration

I hereby declare that I am the sole author of the submitted thesis and no other sources or help apart from those specifically referred to have been used.
This thesis has not been submitted to an examination board before.

Potsdam, 28th of April 2017

Ruth Sonja Kunstmann

Publications and contributions

The following papers have been published in the scope of this work with contributions by Ruth Sonja Kunstmann listed as following:

Kang, Y.[#1, 5], Gohlke, U.[#2], Engström, O.[#6], Hamark, C.[6], Scheidt, T.[1, 7], Kunstmann, S.[1], Heinemann, U.[2, 3], Widmalm, G.[6], Santer, M.[4], and Barbirz, S.[1] **(2016)**. Bacteriophage tailspikes and bacterial O-antigens as a model system to study weak-affinity protein–polysaccharide interactions. *J. Am. Chem. Soc. 138,* 9109–9118. (Persons marked with # contributed equally to this work)

- ➔ Fluorescence titration and kinetics experiments included in this work in Chapter 3.1 in Fig. 11, Fig. 12 and Fig. 23.
- ➔ CD spectroscopy experiments for the thermostability control of Sf6 TSP not included in this work.
- ➔ Proof-reading of the manuscript

Broeker N.K.[1], Andres D.[1, 8], Kang Y.[5], Gohlke U.[2], Schmidt A.[1, 9], Kunstmann S.[1], Santer M.[4], and Barbirz S.[1] **(2017)**. Complex carbohydrate recognition by proteins: Fundamental insights from bacteriophage cell adhesion systems. *Perspectives in Science 11,* 45-52.

- ➔ Proof-reading of the manuscript

From data included in Chapter 3.4 a manuscript has been written with the following contributions (Persons marked with # contributed equally to this work):

Kunstmann S.[#1, 4], Broeker N.K.[#1], Gohlke U.[#2], Heinemann U.[2, 3], Santer M.[4] and Barbirz S.[1] Solvent networks tune thermodynamics of oligosaccharide complex formation in an extended protein binding site (*working title*). *In preparation.*

- ➔ Structural comparison of 27 X-ray crystal structures (Fig. 40,Fig. 41 and Fig. 43), MD simulations and water analysis (Fig. 44-53), free energy calculations (Fig. 54), writing and proof-reading of the manuscript by S. Kunstmann
- ➔ Protein mutation and purification; oligosaccharide purification, ITC experiments (Data included in Fig. 37 and Fig. 54A), writing and proof-reading of the manuscript by N.K. Broeker
- ➔ Protein crystallization, X-ray structure resolution, Structure file deposition at the PDB by U. Gohlke
- ➔ Scientific advice and proof-reading of the manuscript by U. Heinemann
- ➔ Project supervision, scientific advice for computational methods, writing and proof-reading of the manuscript by M. Santer
- ➔ Project planning, supervision and organization, scientific advice, writing and proof-reading of the manuscript by S. Barbirz

Data included in this work not specifically referred to on the previous page might be included in publications in the future. If contributions have not already been adviced differently, the data included in this work were produced by Ruth Sonja Kunstmann except for the following:

- ➔ Fluorescence binding data of Sf6 TSP_{DN} and enzymatical cleavage profiling of Sf2a PS by Sf6 TSP*wt* were conducted in the scope of the bachelor thesis and master student work of Tom Scheidt[1,7] before the start of this work.
- ➔ Computational starting structure of SfY dodecasaccharide, Computer scripts for the superimposition of ligands after parametrization in MD simulations by Yu Kang[4, 5]
- ➔ Computer script for the file convertion of GLYCAM parameters from AMBER to Gromacs by Marko Wehle[4, 10]
- ➔ Sf6 TSP SPR interaction experiments with LPS immobilization (Fig. 20A) were conducted by Lea Günther[1] in the scope of a six weeks practical Master module under the supervision and project planning of S. Kunstmann
- ➔ *Shigella flexneri* isolates as listed in Tab. 8 were characterized and provided by Angelika Fruth[11]
- ➔ ELITA experiments (Fig. 24) were performed in the scope of the Bachelor thesis of Saskia Buchwald[1, 12] under supervision and project planning of S. Kunstmann
- ➔ Sf2a O-antigen icosasaccharides were provided by Laurence Mulard[13]
- ➔ Protein fluorescence labeling, bacterial sensor fluorescence titration and kinetics experiments (Fig. 27) were performed in the scope of the Bachelor thesis of Alexandra Helm[1] under the supervision and project planning of S. Kunstmann
- ➔ Computer script for water selection and water position averaging in the hydrogen bond based water method was provided by Fabian Müller-Knapp[14]

Affiliations

1 Physikalische Biochemie, Universität Potsdam, Potsdam, Germany

2 Max-Delbrück-Centrum für Molekulare Medizin in der Helmholtz-Gemeinschaft, Berlin, Germany

3 Institut für Chemie und Biochemie, Freie Universität, Berlin, Germany

4 Max Planck Institute of Colloids and Interfaces, Potsdam, Germany

5 College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, PR China.

6 Department of Organic Chemistry, Arrhenius Laboratory, Stockholm University, Stockholm, Sweden

7 University of Cambridge, Department of Chemistry, Cambridge, U.K.

8 BayerPharma AG, Berlin, Germany

9 Sanofi-Aventis Deutschland GmbH, Industriepark Höchst, Frankfurt, Germany

10 Fachgebiet Regelungssysteme, Technische Universität Berlin, Berlin, Germany

11 Nationales Referenzzentrum für Salmonellen und andere bakterielle Enteritiserreger, Robert Koch Institute, Wernigerode, Germany.

12 Fachbereich Biochemie, Universität Greifswald, Greifswald, Germany

13 Institut Pasteur, Unité de Chimie des Biomolécules, Paris, France

14 Konzept-Energie GmbH, Berlin, Germany

Zusammenfassung

Kohlenhydrat-Protein Interaktionen sind in der Natur weitverbreitet. Sie stellen die Grundlage für viele biologische Prozesse dar, wie zum Beispiel Immunantworten, Wundheilung und Infektionsprozesse von pathogenen Viren oder Bakterien mit einem Wirt wie dem Menschen. Neben der Infektion von Menschen können aber auch Bakterien selbst durch so genannte Bakteriophagen infiziert werden, welche für den Menschen ungefährlich sind. Diese Infektion involviert die spezifische Erkennung der pathogenen Bakterien, die Vermehrung der Bakteriophagen und schließlich die Abtötung der Bakterien. Dabei können die Mechanismen der spezifischen Erkennung genutzt werden, pathogene Bakterien auf Lebensmitteln zu detektieren oder die Diagnose von Infektionen zu vereinfachen.

Die spezifische Erkennung von Enteritis-erzeugenden Bakterien wie *Escherichia coli*, *Salmonella spp.* oder *Shigella flexneri* durch Bakteriophagen der Familie der *Podoviridae* erfolgt über die Bindung eines sogenannten *tailspike proteins* des Bakteriophagen an das aus Kohlenhydraten-bestehende O-Antigen des Lipopolysaccharids von Gram-negativen Bakterien. Das *tailspike protein* spaltet das O-Antigen um den Bakteriophage an die Oberfläche des Bakteriums zu führen, damit eine Infektion stattfinden kann. Die Affinität des *tailspike proteins* zum O-Antigen ist dabei sehr niedrig, um nach Spaltung des O-Antigens das Spaltungsprodukt zu lösen und wiederum neues Substrat zu binden.

In dieser Arbeit wurde ein *tailspike protein* des Bakteriophagen Sf6 verwendet (Sf6 TSP), das spezifisch an das O-Antigen von *Shigella flexneri* Y bindet. Eine inaktive Variante des Sf6 TSP wurde verwendet um einen hoch-affin bindenden Sensor für pathogene *Shigella* zu entwickeln. Der *Shigella*-Sensor wurde durch Kopplung von unterschiedlichen Proteinmutanten mit einem fluoreszierendem Molekül erhalten. Dabei zeigte eine dieser Mutanten bei Bindung von *Shigella* O-Antigen ein Fluoreszenz-Signal im Bereich des sichtbaren Lichts. Molekulardynamische Simulationen wurde anhand der erzeugten Proteinmutanten als Methode zum rationalen Design von hoch-affin Kohlenhydrat-bindenden Proteinen getestet und die resultierenden Affinitätsvorhersagen wurden über Oberflächenplasmonresonanz-Experimente überprüft. Aus weiteren experimentellen und simulierten Daten konnten schließlich Schlussfolgerungen über die Ursprünge von Kohlenhydrat-Protein Interaktionen gezogen werden, die eine Einsicht über den Einfluss von Wasser in diesem Bindungsprozess lieferten.

Summary

Carbohydrate-protein interactions are ubiquitous in nature. They provide the initial molecular contacts in many cell-cell processes as for example immune responses, signal transduction, egg fertilization and infection processes of pathogenic viruses and bacteria. Furthermore, bacteria themselves are infected by bacteriophages, viruses which can cause the bacterial lysis, but do not affect other hosts. The infection process of a bacteriophage involves the specific detection and binding of the bacterium, which can be based on a carbohydrate-protein interaction. The mechanism of specific detection of pathogenic bacteria can thereby be useful for the development of bacteria sensors in the food industry or for tools in diagnostics.

Bacteriophages of the *Podoviridae* family use tailspike proteins for the specific detection of enteritis causing bacteria as *Escherichia coli*, *Salmonella spp.* or *Shigella flexneri*. The tailspike protein provides the first contact by binding to the carbohydrate containing O-antigen part of lipopolysaccharide in the Gram-negative cell wall. After binding to O-antigen repeating units, the enzymatic activity of tailspike proteins leads to cleavage of the carbohydrate chains, which enables the bacteriophage to approach the bacterial surface for DNA injection. Tailspike proteins thereby exhibit a relatively low affinity to the oligosaccharide structures of O-antigen due to the necessary binding, cleavage and release cycle, compared for example to antibodies.

In this work it was aimed to study the determinants that influence carbohydrate affinity in the extended TSP binding grooves. This is a prerequisite to design a high-affinity tailspike protein based bacteria sensor. For this purpose the tailspike protein of the bacteriophage Sf6 (Sf6 TSP) was used, which specifically binds *Shigella flexneri* Y O-antigen with tetrasaccharide repeating units consisting of [→2)-α-L-Rha*p*-(1→2)-α-L-Rha*p*-(1→3)-α-L-Rha*p*-(1→3)-β-D-Glc*p*NAc-(1→] at the intersubunits of the trimeric β-helix protein. The Sf6 TSP endorhamnosidase cleaves the O-antigen at the α(1→3)-rhamnose linkage, which leads to an octasaccharide as the main product. The binding affinity of inactive Sf6 TSP towards polysaccharide was characterized by fluorescence titration experiments and surface plasmon resonance (SPR). A dissociation constant in the nano- to micromolar range was obtained for the polysaccharide complex. Moreover, several techniques for the covalent immobilization of polysaccharide on the sensor surface were tested in order to evaluate whether these surfaces may serve as a model of the bacterial cell surface for quantitative binding experiments. Analysis of molecular contacts between oligosaccharide and protein in

X-ray crystal structures of the inactive Sf6 TSP and in models obtained with molecular dynamics simulations revealed that especially the rhamnoses at the nonreducing end of the octasaccharide receptor are bound by hydrogen bonds and water bridges in the binding site of Sf6 TSP.

Moreover, cysteine mutations were introduced into the Sf6 TSP binding site for the covalent thiol-coupling of an environment-sensitive fluorescent label to obtain a sensor for *Shigella flexneri* Y based on TSP-O-antigen recognition. This sensor showed a more than 100 % amplitude increase of a visible light fluorescence upon the binding of a polysaccharide test solution. From this, a detection limit for bacteria of ~400 cfu could be deduced. Further improvements are thus needed to increase the sensitivity of the TSP sensor to the necessary detection limit of infectious ten colony forming units of *Shigella flexneri* cells for example by including an enrichment step prior to detection.

The detection limit of the sensor could also be improved by increasing the tailspike affinity towards the O-antigen. Therefore molecular dynamics simulations evaluating ligand flexibility, hydrogen bond occupancies and water network distributions were used for affinity prediction on the available cysteine mutants of Sf6 TSP. The binding affinities were experimentally analyzed by SPR. This combined computational and experimental set-up for the design of a high-affinity carbohydrate binding protein could successfully distinguish strongly increased and decreased affinities of single amino acid mutants.

A thermodynamically and structurally well characterized set of another tailspike protein HK620 TSP with high-affinity mutants was used to evaluate the influence of water molecules on binding affinity. Water was analyzed both in hydrogen bond networks and as individual binding partner mediating carbohydrate-protein binding. The free enthalpy of HK620 TSP-oligosaccharide complex formation thereby either derived from the replacement of a conserved water molecule or by immobilization of two water molecules upon ligand binding. Furthermore, the enthalpic and entropic contributions of water molecules in a hydrophobic binding pocket could be assigned by free energy calculations. Here it was concluded that the separation of water molecules from the residual water network in a binding site leads to a high cost in enthalpy and entropy. The findings in this work can be helpful for the improvement of carbohydrate docking and carbohydrate binding protein engineering algorithms in the future.

Table of contents

List of figures

List of tables

1 Introduction

Design is mostly connected with something aesthetic or artistic. The word ‘design’ originates from the Latin word *designare* for ‘to draw’ or ‘to describe’. Products with exceptional design would be granted with the international Red Dot Award, where they are judged depending on their innovation, aesthetic quality, functionality, realization possibility and emotional content meaning how much the user enjoys the product (red-dot.org, 26.04.2017). For protein design or engineering, these aims also seem applicable, maybe except for the aesthetic quality as a high-resolution technique would be mandatory to see it. Herbert A. Simon, who has been awarded the Nobel Prize in Economics in 1978, wrote in “The Science of the Artificial” from 1969: “Everyone designs who devise courses of action aimed at changing existing situations into preferred ones.” This thesis is about modifying a Sf6 bacteriophage tailspike protein to be an improved binder and sensor and to advance the situation of *Shigella flexneri* pathogen diagnostics and the general knowledge of carbohydrate binding.

1.1 Carbohydrate-protein interactions

Carbohydrate-protein interactions are a prerequisite for most cellular processes (Varki et al., 2009). Carbohydrates depict the most diverse class of biological polymers due to their high number of functional groups, which can be cross linked in multiple combinations (Zeng et al., 2012). Therefore carbohydrate-protein interactions are involved in the recruitment of immune cells (Lasky, 1992), fertilization (Sinowatz et al., 1998), signal transduction pathways (Sacchettini et al., 2001) and infection process (Kline et al., 2009). Thus, carbohydrate protein interactions fulfil tasks in and around eukaryotic cells (Fig. 1). In the cell, glycan processing enzymes post-translationally modify proteins with N- and O-glycans. Therfore activated nucleotide sugars are formed with the help of nucleotide-sugar glycosyltransferases in the cytosol. The activated sugars are relocated via nucleotide-sugar transporters into the Golgi apparatus and then transfered onto the glycan modifications by glycosyltransferases (Coutinho et al., 2003). Outside the cell, glycosidases cleave mono- or oligosaccharides from glycans for modification purposes (Rye and Withers, 2000) or in the process of virus replication as for example the neuramidase of the Influenza virus cleaving sialic acids from glycans for virus release (Wagner et al., 2002).

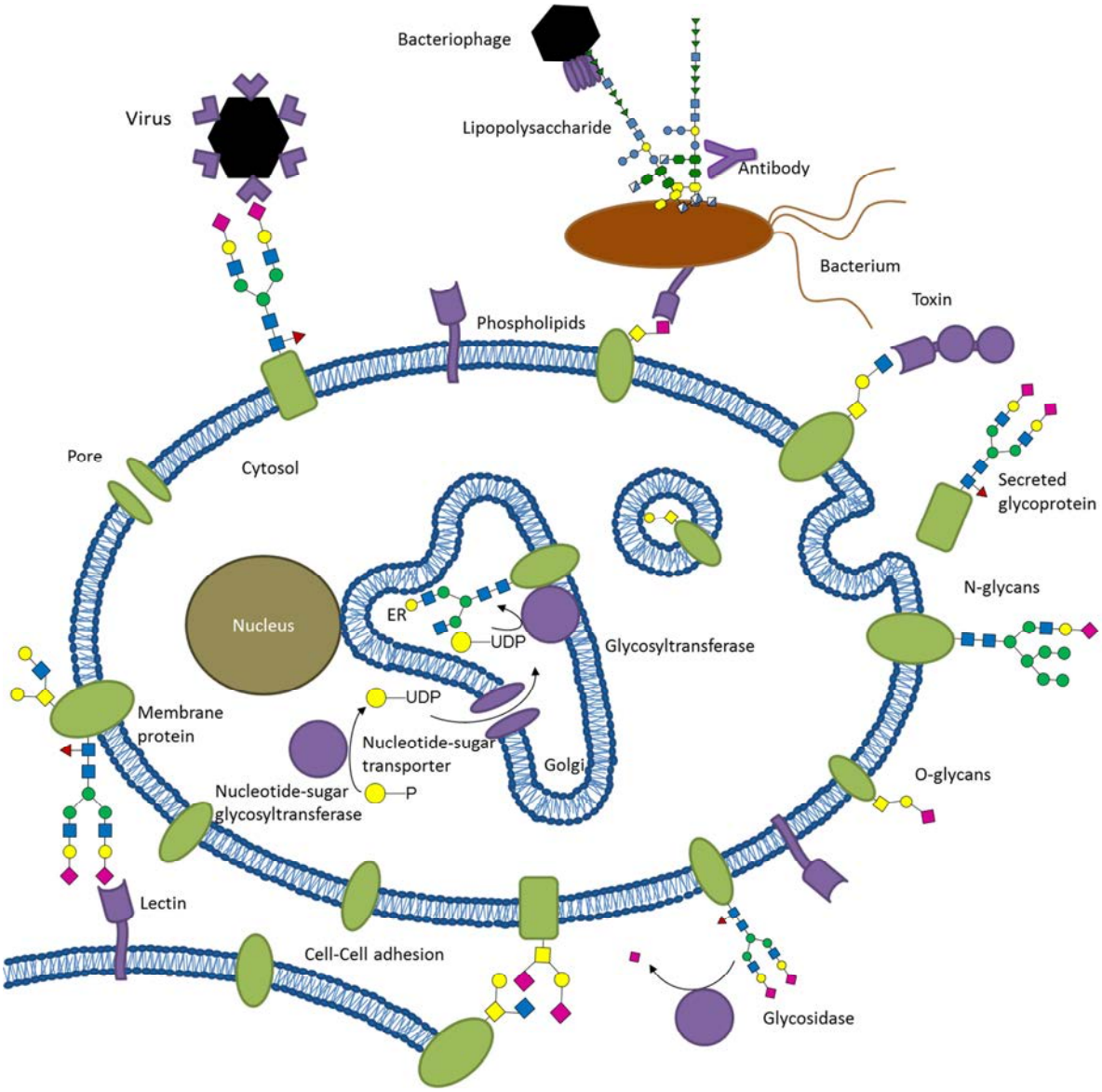

Fig. 1: Carbohydrate-protein interactions
A eukaryotic cell is presented by the lipid bilayer (blue) surrounding a nucleus (brown) and a cartoon of the membrane-based endoplasmic reticulum (ER), Golgi-apparatus and vesicles (blue). Proteins in the cellular membrane or in the ER/Golgi are depicted in green and carbohydrate-binding proteins and enzymes are shown in purple. Glycans are depicted according to the symbolic nomenclature for glycans as various forms of N- and O-glycans and bacterial lipopolysaccharides (Varki et al., 2016). A bacterium is depicted in brown and a virus and a bacteriophage in black.

Carbohydrate processing enzymes are coupled to carbohydrate binding modules (CBM), providing the binding independently from the enzymatic domain (Boraston et al., 2004). Carbohydrate binding proteins without enzymatic activity as lectins provide the contacts between cells for example in the nucleophile recruitment (Lasky, 1992) or are produced by plants, bacteria or viruses for the binding on cell surface exposed glycans for infection processes or as toxins (Lis and Sharon, 1998). Furthermore, bacteriophages infect bacteria by binding to the bacterial surface carbohydrates by tailspike proteins (Broeker et al., 2017). In contrast, the immune system produces antibodies with carbohydrate based antigens for the recognition of virus and bacteria carbohydrates (Kosma and Müller-Loennies, 2011).

1.1.1 Lectins and antibodies

Lectins and carbohydrate binding antibodies constitute different strategies for the complex formation. While lectins bind mono- to tetrasaccharides on shallow protein binding sites, the carbohydrate binding sites on antibodies are buried in grooves with elongated oligosaccharide ligands (Fig. 2) (Bundle and Young, 1992).

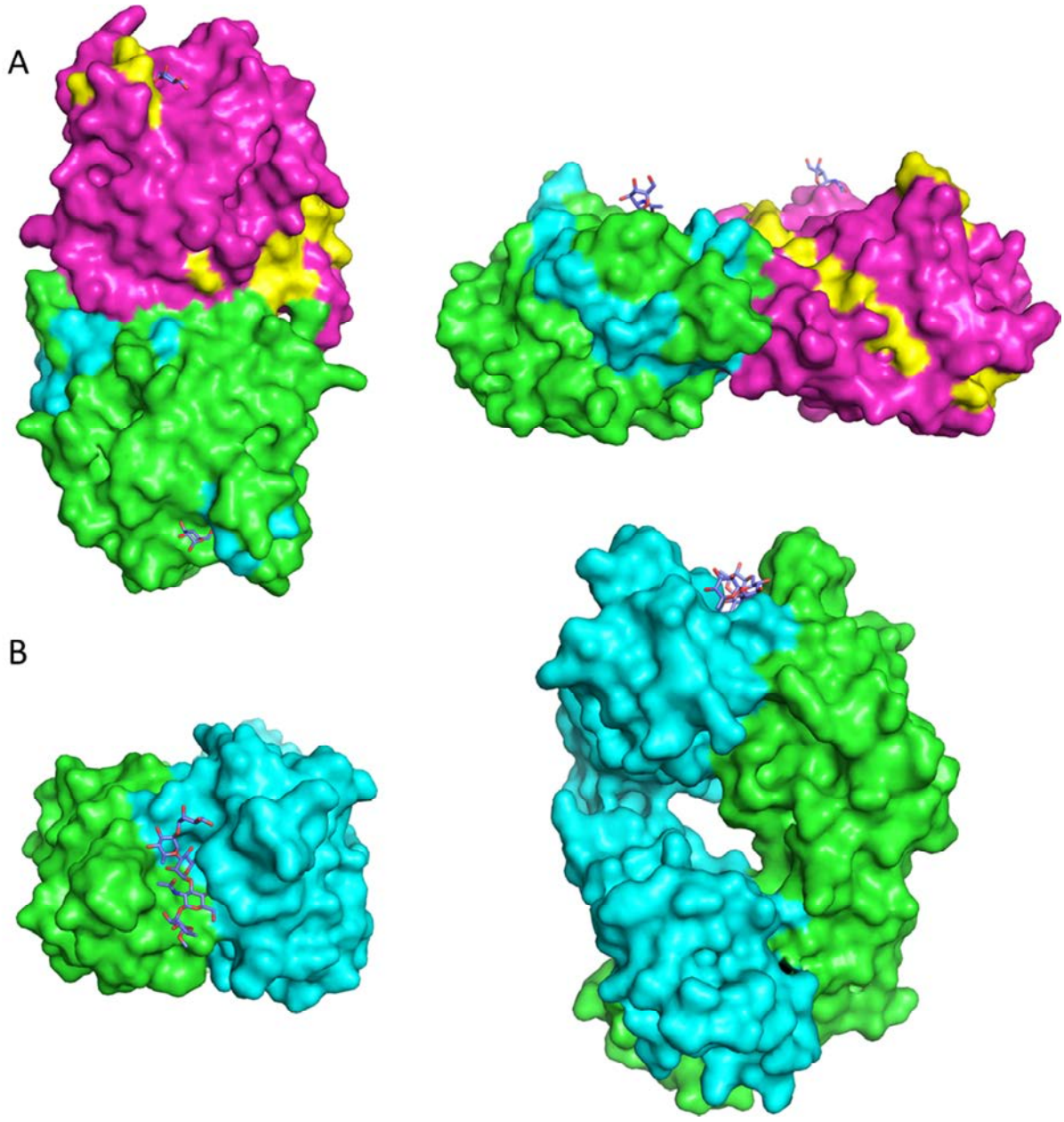

Fig. 2: Lectins and antibodies
(A) The Fava bean lectin is depicted as tetrameric protein (multicolored surface) with two Glc and two GlcNAc (blue sticks) in shallow binding sites on the top and the side of the protein (pdbID: 2B7Y) (Reeke and Becker, 1986). **(B)** The FAB fragment of the *Shigella flexneri* Y O-antigen antibody as dimer (green and cyan surface) with the pentasaccharide ligand (blue sticks) in a deep binding groove on the intersubunit site of the protein as top and side view (pdbID: 1M7I) (Vyas et al., 2002).

Lectins are mostly built up as multidomain proteins (Lee and Lee, 1995). The carbohydrate binding sites are thereby structurally related as some lectins exhibit the so-called lectin fold, which consists of two antiparallel β-sheets with six or seven strands superimposed on each other (Lis and Sharon, 1998). Ligand specificity might be restricted to one monosaccharide or be broadened by selective binding of only O3 and O4 of the monosaccharide, which would leave variabilities at O2 for example for glucose and mannose specific binders like Concanavalin A (ConA) (Lis and Sharon, 1998). Monosaccharides are bound to lectins with a low affinity in the millimolar range (Gabius, 2008). However multivalence is used to increase affinity either by binding oligosaccharide fragments or by clustering multiple protein domains (Lis and Sharon, 1998; Lundquist and Toone, 2002).

Carbohydrate binding antibodies can bind epitopes of oligosaccharides including one to seven monosaccharides with affinities in the micro- to nanomolar range (Vyas et al., 2002). Thereby small epitopes are bound in cavities and elongated epitopes in groove shaped sites (Padlan and Kabat, 1988). Antibody sites do not bind a segment on each residue, meaning that some carbohydrate moieties are solvent exposed as polysaccharides tend to form helical structures (Evans et al., 1995; Vulliez-Le Normand et al., 2008).

1.1.2 Binding contributions

Albeit the importance of carbohydrate-protein interactions in nature, their processes and driving forces are less well understood than for protein-protein or protein-carbohydrate interactions (Lee and Lee, 1995). Reasons for this are the high flexibility and diversity of carbohydrate structures, which makes the resolution in X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy more challenging (DeMarco and Woods, 2008). Moreover, the low binding affinities between proteins and carbohydrates in the milli- to micromolar range complicate interaction measurements by isothermal titration calorimetry (ITC), surface plasmon resonance (SPR) and fluorescence spectroscopy (Lee and Lee, 1995).

Carbohydrate-protein interactions are established by direct hydrogen bonds between polar amino acids and mainly hydroxyl groups of the carbohydrate, via indirect hydrogen bonds which involve a bridging water molecule and by CH,π-interactions of an aromatic ring side chain of the protein and the apolar side of a sugar ring perpendicular to each other with CH bonds pointing towards the aromatic ring (Fig. 3) (Lee and Lee, 1995). In plant legume lectins and animal C-lectins also metal coordination with calcium or manganese ions can occur (Lis and Sharon, 1998).

The intermolecular hydrogen bonds are primarily formed by Asp, Asn, Glu, Arg, His, Trp, Lys, Tyr, Gln, Ser and Thr (Lütteke and von der Lieth, 2009) which contribute with different strengths to the binding affinity.

Aromatic residues (Phe, Tyr and Trp) mostly coordinate neutral sugars like glucose or galactose or their N-acetylated forms (Muraki et al., 2002). These interactions are weaker than direct hydrogen bonds and depend on the aromatic residue. A CH,π-interaction accounts for 2-3 kcal/mol (Chávez et al., 2005).

Water molecules in the binding site provide stability for intra- and intermolecular connections (Kadirvelraj et al., 2008). In the protein binding site water molecules occupy the hydroxyl groups of the carbohydrate to pre-organize the binding site by orienting the residues involved in binding (Saraboji et al., 2012). Thus upon binding the entropic loss is minimized and the binding enthalpy is strongly negative due to the dynamic reorganization of water molecules (Chervenak and Toone, 1994). The energy gain for a release of a highly ordered water molecule accounts for ~2 kcal/mol (Dunitz, 1994). However, carbohydrate-protein interactions are governed by enthalpy-entropy compensation as the degrees of freedom of torsion angles of glycosidic bonds in the carbohydrate ligand are strongly reduced upon binding (Dunitz, 1995). This binding contributions and the involvement of water in binding thermodynamics are mainly studied by X-ray crystallography and NMR (Roldós et al., 2011), but also strongly rely on modeling techniques with explicit water molecules as in molecular dynamics simulations (MD) (Clarke et al., 2001).

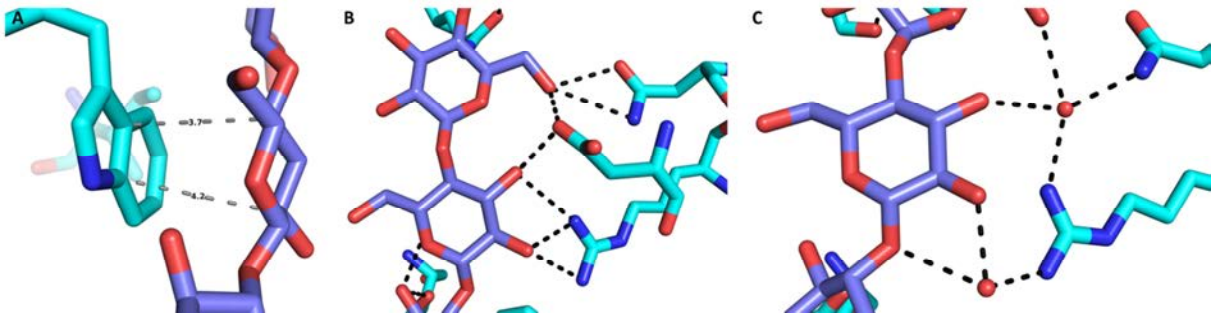

Fig. 3: Binding contacts between proteins and carbohydrates
Protein-carbohydrate interactions are depicted as CH,π-interactions **(A)**, direct hydrogen bonds **(B)** and water mediated hydrogen bonds **(C)** from the family 17 carbohydrate-binding module from *Clostridium cellulovorans* in complex with cellotetraose (pdbID: 1J84) (Notenboom et al., 2001). The carbohydrate is depicted as blue sticks, water as red spheres and protein residues are shown as cyan sticks. Hydrogen bonds (black dashes) have a maximal distance of 3.5 Å. Gray dashes indicate the distance between the aromatic tryptophan and the carbohydrate ring with 3.7 and 4.2 Å.

1.2 The bacterium *Shigella flexneri*

Bacillus dysentericus, later on named *Shigella dysenteriae,* was first described in 1897 by Kiyoshi Shiga in Japan (Yabuuchi, 2002). The taxonomic genus *Shigella* comprises Gram-negative bacteria which belong to the family of the *Enterobactericeae* (Skerman et al., 1980). They are divided into four subspecies with *Shigella* (*S.*) *dysenteriae*, *S. flexneri*, *S. boydii* and *S. sonnei* and are closely related to the enteroinvasive *Escherichia coli* (EIEC) (*E. coli*) with more than 90 % chromosomal DNA identity (Sasakawa, 1997). *Shigella* cause Bacillary dysentery or Shigellosis and are a major cause of infancy death in tropical regions (Kotloff et al., 2013). From these serogroups the main cause for Shigellosis infections in developing regions is *Shigella flexneri* (Anderson et al., 2016). The symptoms are watery and bloody diarrhea due to inflammation and ulceration of the colon. The infection can lead to severe sepsis which causes hypoglycemia or a hemolytic uremic syndrome with kidney failure (Sansonetti et al., 1999). *Shigella* cells are transmitted from person-to-person or by contaminated drinking water and several classical antibiotic resistances complicate the treatment of the infection (Anderson et al., 2016; Murray, 1986). Already a very small number of 10-500 cells can cause an infection in adults (DuPont et al., 1989). Thus, the bacteria developed an efficient mechanism for colonization and infection of the human host (Sansonetti et al., 1999; Sasakawa, 1997).

1.2.1 The infection mechanism of *Shigella*

Most of the genes involved in the infection process of *Shigella* and EIEC are coded on a virulence plasmid including the *ipa*, *mxi* and *spa* loci (Parsot, 2005). After ingestion of *Shigella* cells, the bacteria pass the acidic environment of the stomach by expression of the specific sigma factor σ^s resulting in improved resistance to acid (Small et al., 1994). The bacteria colonize the colon by adherence and passage of the mucosal layer, which is the least understood process of the infection. The two barriers in the colon are the natural microbiota and the mucus layer (Anderson et al., 2016; Day et al., 2015; Xu et al., 2016). The natural microbiota might be overcome by bacteriocin production (Calcuttawala et al., 2015). The mucus layer can be disrupted by the expression of mucinases and neuramidases of the bacteria and the passage to the epithelial cells is possible (Anderson et al., 2016; Henderson et al., 1999).

Shigella cells enter the epithelial cell barrier of the colon via M cells and internalization of macrophages. The bacteria escape the macrophage digestion by disruption of the vacuole and movement into the cytosol and the induction of macrophage apoptosis (Zychlinsky et al., 1992). Apoptosis leads to the release of IL-1 cytokines which attract nucleophiles and induce inflammation (Chen and Zychlinsky, 1994). Epithelial cells of the colon are invaded by *Shigella* cells from the basolateral site by interaction with components of the extracellular matrix like fibronectin, laminin and collagen type IV (Watarai et al., 1995). Mxi and Spa proteins are secreted by the type III secretion system of the bacteria. These induce polymerization of actin and the formation of protrusions in the epithelial cell membrane leading to macropinocytosis of the bacteria within 5-10 min (Francis et al., 1993; Parsot, 2005). *Shigella* cells escape into the cytoplasm by disruption of vacuoles by IpaB and IpaC (Agaisse, 2016). The intercellular spread between cells proceeds via the manipulation of the actin cytoskeleton of the host cell (Welch and Way, 2013). The bacterial factor IcsA interacts with N-WASP, the major actin nucleator, leading to actin nucleation at the bacterial pole. This actin-based mechanism enables mobility and the spread to the neighboring cell by repeated protrusion of the cell membrane (Agaisse, 2016). These processes provide the spread of bacterial cells and the disruption of the epithelium.

1.2.2 Diagnosis and vaccination

The four subspecies of *Shigella* are further divided into 15-20 serogroups each for *S. flexneri, S. dysenteriae* and *S. boydii* and one serotype for *S. sonnei*. The serotypes of *Shigella* are based on the O-antigen structure of the lipopolysaccharide (LPS) in the Gram-negative cell wall (see Chapter 1.3) (Allison and Verma, 2000). The O-antigen consists of a tetrasaccharide built up by three rhamnoses and one N-acetylglucosamine (GlcNAc, NAG), which is the backbone of the O-antigen and depicts the serogroup Y (Fig. 4) (Kenne et al., 1978). Modifications by glucosylation and acetylation are induced by serotype converting bacteriophages (see Chapter 1.4) leading to the variation of the serotypes (Allison and Verma, 2000; Gemski et al., 1975; Verma et al., 1991). This results in 15 serotypes for *Shigella flexneri* with the same backbone except for serogroup 6, where L-Rha*p* is replaced by an α(1-4) linked glucuronic acid (D-Glc*p*A). The *S. flexneri* serotypes are depicted in Fig. 4 based on Perepelov et al., 2012.

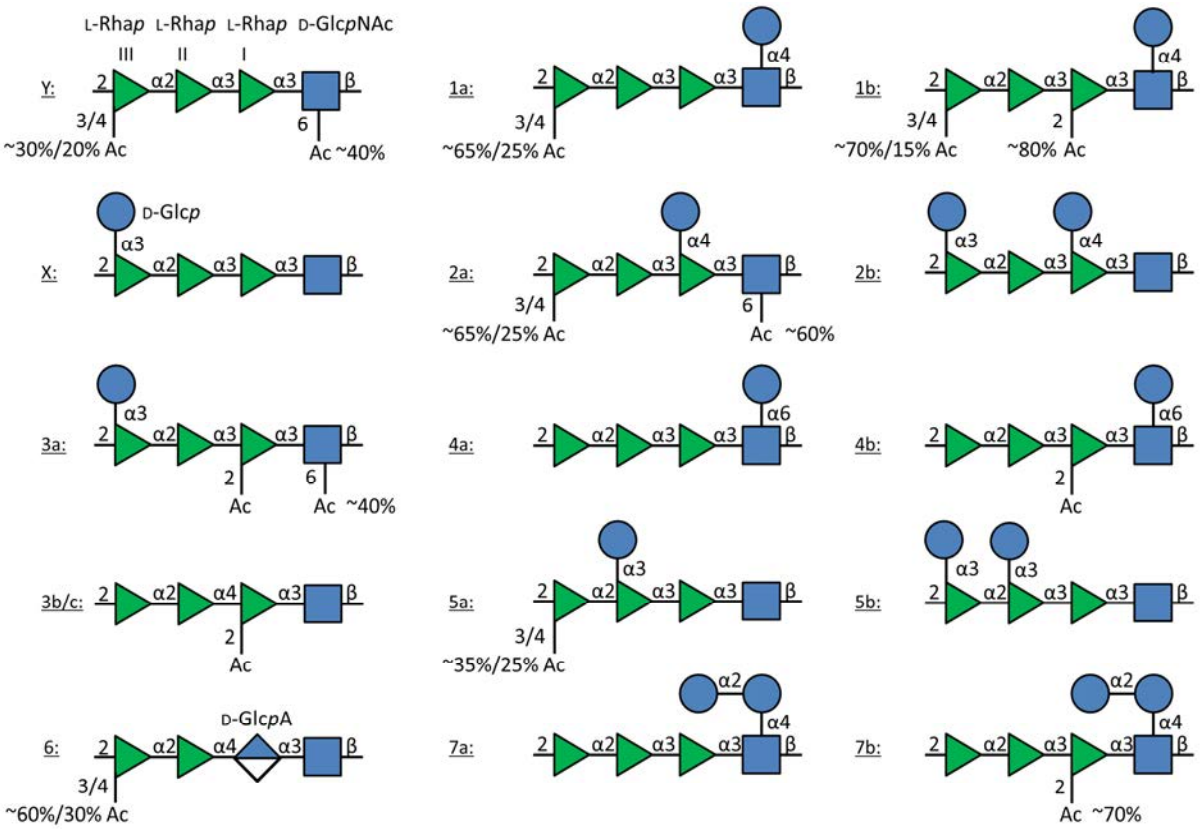

Fig. 4: *Shigella flexneri* O-antigen repeating units
The 15 serotypes of *Shigella flexneri* are depicted as glycan structures according to the symbolic nomenclature for glycans (Varki et al., 2016). The structures are taken from (Perepelov et al., 2012).

The serotypes are determined in diagnostics by antisera which are used in agglutination tests (Carlin and Lindberg, 1983; Carlin et al., 1986). Furthermore, genotyping is employed for diagnostics which target O-antigen modifying genes (Li et al., 2009; Sun et al., 2011) or genes which are known for *Shigella flexneri* virulence as *ipaH, cadA, set1B* and *ial*, coding for an invasion plasmid antigen, a lysine decarboxylase, an enterotoxin and an adhesion protein, respectively (Frankel et al., 1989; Li et al., 2009).

Shigella flexneri infections are treated by the antibiotic ciprofloxacin, which is problematic due to rising antibiotic resistances (Anderson et al., 2016). Infected patients develop serum LPS and Ipa specific antibodies (Lindberg and Pál, 1993). Therefore vaccine development is based on LPS-protein conjugates (Kämpf et al., 2015; Passwell et al., 2010).

1.3 The Gram-negative cell wall

The classification of the Gram-negative and positive cell walls of bacteria traces back to Hans Christian Gram in 1884, who used crystal violet for the staining of peptidoglycan in the bacterial cell wall. This resulted in a strong staining for Gram-positive bacteria due to the thick peptidoglycan layer and to a weak staining of Gram-negative bacteria with a thin peptidoglycan layer protected by an additional outer membrane (Popescu and Doyle, 1996).

The peptidoglycan is a polymer conjugate of N-acetylglucosamine and muramic acid peptides, which is located in the periplasmic space between the inner and outer membranes of the Gram-negative cell wall (Fig. 5) (Osborn et al., 1964). These membranes are constituted of high amounts of proteins and phospholipids and are anchored to the peptidoglycan via lipoproteins (Lindberg, 1973). Lipopolysaccharide (LPS) is anchored in the outer membrane with up to one million molecules per *E. coli* cell. LPS is known as endotoxin which can cause septic shock up to multiple organ failure (Raetz and Whitfield, 2002). The structure of LPS is explained exemplified for *Shigella flexneri* Y in the following section.

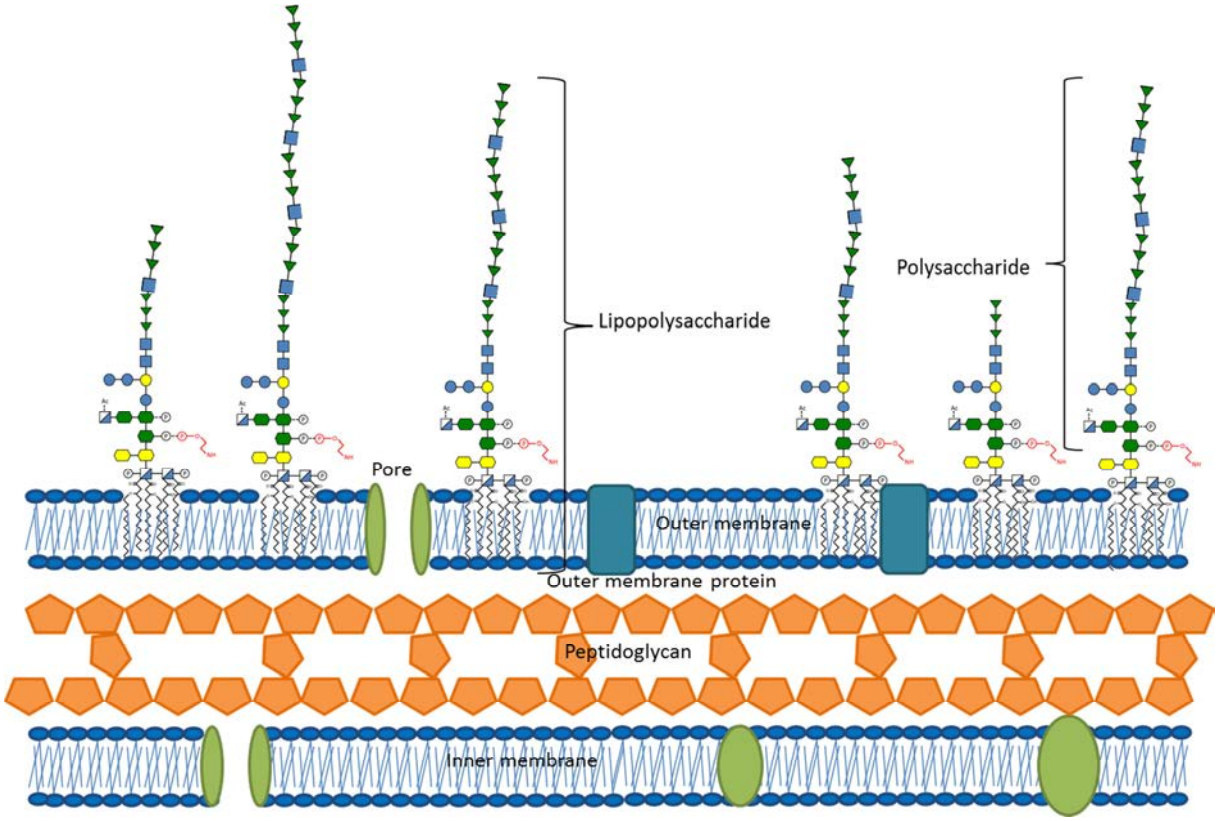

Fig. 5: The Gram-negative cell wall
The Gram-negative cell well is generally built up from the inner and outer membrane (blue) with a thin layer of peptidoglycan in the periplasmic space (orange). Transport of molecules is achieved by pores in both membrane layers (green). The outer membrane includes specific outer membrane proteins (Omp) with various functions (turquoise). In the outer membrane, lipopolysaccharide is anchored via the lipid A and the polysaccharide is located in the extracellular space (symbolic nomenclature for glycans, SNFG).

1.3.1 *Shigella flexneri* Y Lipopolysaccharide

Shigella flexneri Y LPS consists of the hydrophobic lipid A membrane anchor, the inner and outer core with a conserved carbohydrate sequence and the serotype specific O-antigen repeating units (Fig. 6). Bacteria expressing lipid A with the complete inner and outer core have a rough LPS phenotype (R-LPS). The addition of one O-antigen repetative unit (*RU*) constitutes a semi-rough phenotype (SR-LPS), while LPS with a large number of O-antigen *RU* results in the smooth LPS phenotype (S-LPS) (Simmons and Romanowska, 1987).

Lipid A is highly conserved among *Enterobactericeae* and consists of two phosphorylated glucosamines with six linked acyl chains of 12 and 14 carbon atoms (Galloway and Raetz, 1990; Raetz and Whitfield, 2002).

Inner and outer cores are structurally less conserved compared to lipid A. The inner core in most bacteria consists of two Kdo residues (3-deoxy-D-manno-oct-2-ulosonic acid), which are added during lipid A biosynthesis and are required for a minimal LPS structure for *in vitro* bacterial growth (Raetz, 1990). Additionally three L-glycero-α-D-heptopyranoses (L,D-Hep) are included in the inner core. The inner core structure can be modified non-stoichiometrically by phosphate, pyrophosphate, phosphoethanolamine and phosphocholine (Raetz and Whitfield, 2002).

The *E. coli* R3 core type, which is also the core type of *S. flexneri* LPS, contains about 30 % glucosamine on the branched L,D-Hep, but it is only present if the adjacent L,D-Hep is not phosphorylated (Knirel et al., 2011; Müller-Loennies et al., 2002; Simmons and Romanowska, 1987). The phosphoethanolamine (PEtN) modifications can occur at the inner core structure as single PEtN at the branched Kdo moiety (not shown in Fig. 6) or as a pyrophosphoethanolamine (PPEtN) modification at the first heptose structure (in red in Fig. 6) (Dröge et al., 1970; Lehmann et al., 1971). The amount of PEtN modification of the Kdo is increased with 5-50 mM Ca^{2+} ions in the growth medium (Kanipes et al., 2001). The PPEtN modification of L,D-HepI is catalyzed by the WaaP kinase and the EptC phosphoethanolamine transferase for phosphate and PEtN coupling, respectively (Klein et al., 2013; Yethon et al., 1998). PEtN seems to be required for outer membrane permeability and the modification is especially prevalent under phosphate limited growth conditions and addition of Zn^{2+} (Klein et al., 2013). Furthermore in *Shigella flexneri,* a plasmid encoded PEtN transferase Lpt-O was found to modify various *Shigella flexneri* O-antigen repeating units by PEtN modification at position 3 of the second rhamnose (RAMII) (Knirel et al., 2013; Sun et al., 2012).

The outer core consisting of a pentasaccharide with an α(1-3) linked sequence of GlcNAc, Gal and Glc with two α(1-2) branched glucoses at the galactose is conserved for *S. flexneri* strains (except serotype 6) and based on the *E. coli* R3 type (Fig. 6).

The chain length of the O-antigen is bimodally distributed for *S. flexneri* and regulated by a chromosomal chain length distributer gene *wzz* resulting in a length of 12-17 repetetive units and a plasmid coded *wzz* gene regulating very long O-antigen chains up to 90 *RU* (Morona et al., 1995; Stevenson et al., 1995). This distribution is hypothesized to be evolutionary optimized for virulence (Morona, 2003).

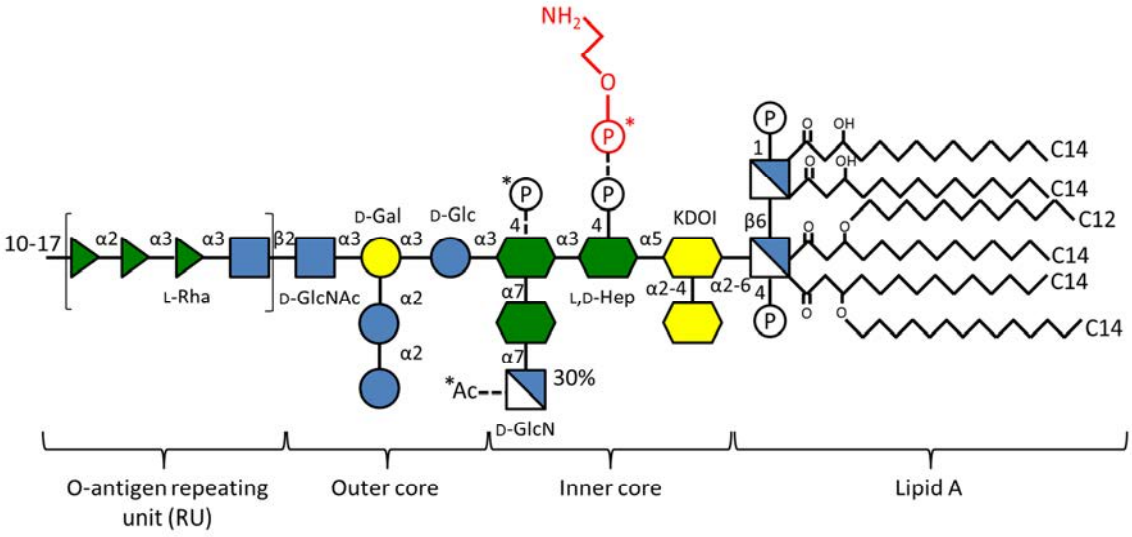

Fig. 6: *Shigella flexneri* Y lipopolysaccharide
Structure of *Shigella flexneri* Y LPS depicted according to the symbolic SNFG (Varki et al., 2016). Dashed lines and asterisks indicate non-stoichiometric modifications (P=Phosphate, Ac=Acetyl-).

1.4 Bacteriophages

Exactly 100 years ago the term bacteriophage was first used by Felix d'Herelle and was described as ultramicroscopic parasite of bacteria (Bradley, 1967). Nowadays bacteriophages are recognized as bacterial viruses as they resemble eukaryotic viruses. Nucleic acids are packed in a capsid which is connected with a tail region. Bacteriophages are distinguished into temperate and virulent. Temperate phages inject DNA in the bacterium and the phage genome is integrated into the bacterial chromosome. Due to an external trigger, formation of phage particles within the cell is induced and the bacterium is lysed. In contrast, virulent phages inject their DNA into a bacterium with direct lysis of the cell (Bradley, 1967). The morphology of bacteriophages was determined mainly by electron microscopy and bacteriophages were classified into six groups (Tab. 1).

Tab. 1: Classification of bacteriophages

Group	Morphology	Taxonomy of *Caudovirales*
A	contractile tail packed with dsDNA	*Myoviridae*
B	long noncontractile tail packed with dsDNA	*Siphoviridae*
C	short noncontractile tail packed with dsDNA	*Podoviridae*
D	large capsomer without tail packed with ssDNA	*Microviridae*
E	small capsomer without tail packed with ssRNA	*Leviviridae*
F	flexible filament packed with ssDNA	*Inoviridae*

dsDNA = double-stranded DNA, ssDNA = single-stranded DNA, ssRNA = single-stranded RNA; derived from Ackermann, 2011; Bradley, 1967 and Maniloff and Ackermann, 1998

Independently of their morphology, most bacteriophages are specific for bacteria of the *Enterobactericeae* family (Bradley, 1967). The life cycle of the bacteriophages includes the specific attachment on the cell, the injection of the nucleic acid, followed by replication of the nucleic acids in the cell and the subsequent release of new phages. The attachment is driven by specific polymeric molecules on the surface of Gram-negative and positive bacteria, which have been identified as LPS or teichoic acid, respectively (Lindberg, 1973) for Group A to C phages. Thereby the contact point was identified as the tail of the bacteriophage and group C has been connected with an endoglycosidase activity which cleaves the LPS O-antigen (Lindberg, 1973).

The most intensively studied and characterized candidate of the group C bacteriophages is P22 infecting *Salmonella.* It gives the name to a group of related phages known as “P22-like” (Maniloff and Ackermann, 1998; Parent et al., 2012; Zinder and Lederberg, 1952), for example Sf6 infecting *Shigella flexneri* Y and HK620 infecting *Escherichia coli* H (Dhillon et al., 1998; Gemski et al., 1975). The three named bacteriophages display a high genetic identity with up to 90 % for Sf6 and HK620 (Casjens et al., 2004). Using electron microscopy a complete P22 virion could be solved with a ~65-70 nm capsid head including the tail hub, needle and six tailspike proteins (Fig. 7A) (Chang et al., 2006; Israel et al., 1967) with similar observations also for Sf6 (Lindberg et al., 1978). For P22 and Sf6 bacteriophages an endorhamnosidase activity has been identified, leading to the cleavage of O-antigen repetative units used by the phage to approach the bacterial cell wall, which was speculated to derive from the tailspike proteins (Lindberg et al., 1978).

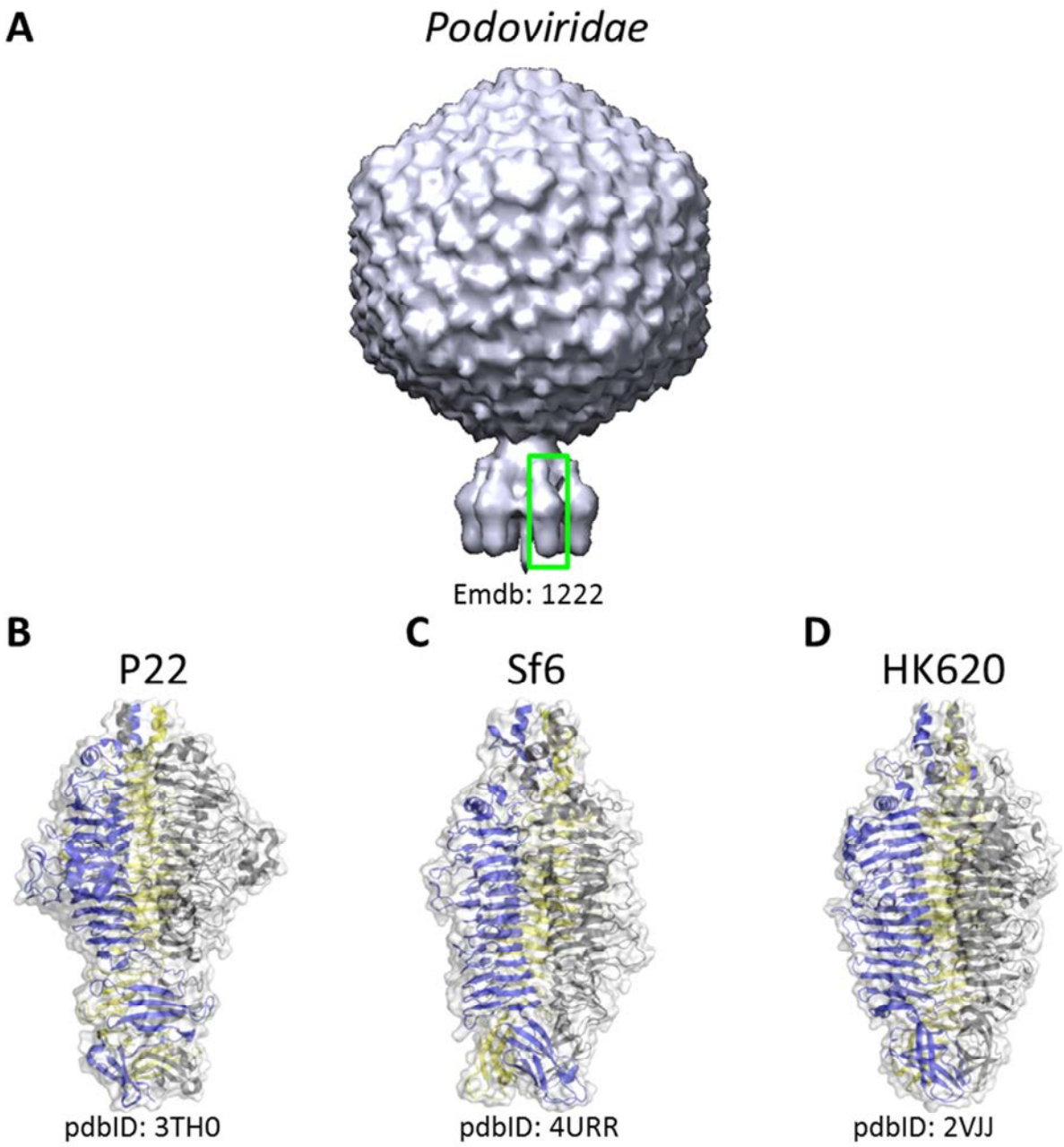

Fig. 7: *Podoviridae* bacteriophage with tailspike proteins
(A) Electron microscopy model of P22 bacteriophage with capsid, tail hub, needle and tailspike proteins (framed in green). Tailspike proteins of the bacteriophages P22 **(B)**, Sf6 **(C)** and HK620 **(D)** as trimeric proteins.

1.4.1 The Sf6 tailspike protein

In P22 and Sf6, the endorhamnosidase was attributed to the tailspike proteins (Fig. 7B-C) (Baxa et al., 1996; Chua, 1999). Additionally, HK620 tailspike exhibits endo-N-acetylglucosaminase activity (Fig. 7D) (Barbirz et al., 2008). Moreover, the tailspike proteins of the respective bacteriophages are structurally conserved suggesting an evolutionary relationship (Barbirz et al., 2008).

Tailspike proteins have been extensively studied as a model system for protein folding (Seckler, 1998). The carbohydrate binding protein constituted of a β-solenoid fold is stable to heat, detergent and proteases and folds slowly facilitating the study of folding in experimentally accessible time scales (Barbirz et al., 2009; Miller et al., 1998). The fold was first identified in the pectate lyase C (Yoder et al., 1993) and has been associated to the binding of elongated carbohydrates (Kajava and Steven, 2006).

A tailspike protein (TSP) consists of four domains comprising a capsid-head binding domain, an α-helical N-terminal domain, a right-handed β-helix domain and a C-terminal β-sandwich (Fig. 8A) (Müller et al., 2008). The three tailspike proteins P22, HK620 and Sf6 were crystallized as homo-trimeric proteins lacking the N-terminal capsid binding domain with their respective O-antigen binding partners as oligosaccharide fragments (Barbirz et al., 2008; Müller et al., 2008; Steinbacher et al., 1997). The conserved protein folds without the N-terminal capsid binding domain have a length of 12.7 nm and a diameter of 7 nm including 13 β-sheet rungs in a 7 nm long β-helix domain (Fig. 8B) (Barbirz et al., 2008). Each rung is thereby constituted of three β-sheets B1, B2 and B3, which are separated by the three turns T1, T2 und T3. T2 consists of two amino acids, while T1 and T3 range between five and ten amino acids in the case of Sf6 TSP (Fig. 8C) (Müller et al., 2008). The carbohydrate binding sites and endo-glycanase activities of P22 and HK620 TSP have been identified to lie within the intrasubunit cleft of one trimeric subunit, while for Sf6 TSP the binding site was identified in the intersubunit cleft (Fig. 8D) (Barbirz et al., 2008; Müller et al., 2008; Steinbacher et al., 1997). Even though HK620 and Sf6 TSP are more similar regarding their structure, Sf6 and P22 TSP both show endorhamnosidase activity and cleave the specific O-antigen receptor into oligosaccharides of two repeating units (Barbirz et al., 2008; Müller et al., 2008). The affinity towards this oligosaccharide are in the milli- to micromolar range (Baxa et al., 1996; Broeker et al., 2013; Kang et al., 2016). Thus, tailspike proteins depict highly stable O-antigen specific carbohydrate-binding proteins with endo-glycosidase activity and medium to low binding affinities towards oligosaccharide fragments.

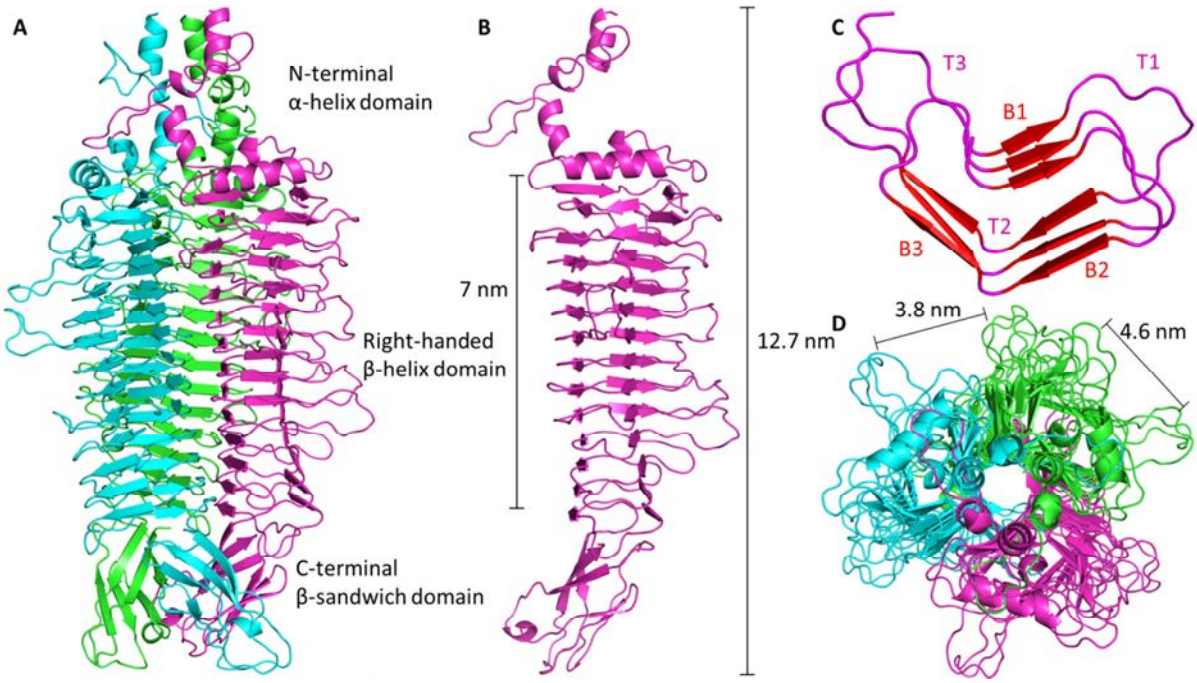

Fig. 8: Sf6 tailspike protein
(A) Sf6 tailspike protein as homotrimer with a molecular mass of 56 kDa per monomer crystallized as N-terminally truncated variant without capsid binding domain. The protein consists of an N-terminal α-helix domain and a C-terminal β-sandwich domain. In between the three right-handed β-helix domains form a left-handed superhelical twist. **(B)** One subunit of the homotrimer is shown. The protein has an overall length of 12.7 nm with the β-helix exhibiting the longest part of 7 nm. **(C)** Cross-section of the β-helix fold motif with three β-sheets B1, B2 and B3 connected by three turns T1, T2 and T3. **(D)** The trimeric protein shown from the N-terminus with six grooves with a wideness of 4.6 nm (intrasubunit) and 3.8 nm (intersubunit). The structural models were derived from the pdbID 4URR (Kang et al., 2016).

1.5 Protein design

Protein design or engineering is a technique for the improvement of protein properties like stability, enzyme activity or binding affinity (Pokala and Handel, 2001). Furthermore, *de novo* design of proteins with certain planned geometries (Höcker et al., 2001) or changes for binding specificity or enzymatic activity are achieved as so called scaffold design (Pokala and Handel, 2001). The design can be either random, by a natural change of coding DNA sequences by error-prone PCR or directed evolution (Arnold, 1998; Cirino et al., 2003) or it can be rational, which mostly includes computational approaches (Marshall et al., 2003). Random or rational derived sequence libraries coding for the protein of interest are tested in high-throughput methods involving phage- or yeast-display approaches (Rhiel et al., 2014) and the subsequent characterization of for example binding properties by SPR, NMR or mass spectrometry (Pande et al., 2010). Furthermore computational derived *de novo* or altered protein structures highly rely on experimental techniques allowing model deduction like X-ray crystallography, NMR and cryo-electron microscopy (Das and Baker, 2008).

To name some examples for carbohydrate binding protein design, phage-display is employed for the design of antibodies against charged carbohydrate structures associated with tumor cells (Mao et al., 1999; Schoonbroodt et al., 2008). In phage-display approaches, the peptide sequences derived from DNA libraries are fused to a phage surface protein, which is more precisely a filamentous M13 phage with the pIII coat protein. The displayed peptide chains are consequently screened for binding affinity or specificity by biopanning. In biopanning the ligand of interest is immobilized in a microtiter plate, phages are applied for binding, non-bound phages are removed by washing and the residual bound phages are amplified and used for further selection (Pande et al., 2010). As another example carbohydrate-binding proteins are used as protein scaffolds due to their high stability as high-affinity ligand binders in computational approaches (Morin et al., 2011). The most prominent example of computational protein design is nowadays ROSETTA, which is a modelling software based on Monte-Carlo simulations with protein side chain rotamer perturbations in implicit solvent with the aim of energy-landscape minimization (Rohl et al., 2004). More computationally expensive techniques, which in return allow a higher precision, are based on molecular dynamics simulations (Baron, 2012).

1.5.1 Rational protein design of carbohydrate-binding proteins

Rational design of high-affinity binding proteins for protein-ligand or protein-protein interactions is based on computational techniques. Thereby the design of protein-ligand high-affinity binders depicts a special challenge compared to enzyme specificity improvement or protein-protein interactions (Morin et al., 2011). Even though programs like ROSETTALIGAND (Meiler and Baker, 2006), which are commonly used for other high-affinity design purposes, are in the process to be optimized for protein-carbohydrate interactions as ROSETTADOCK (Kilambi et al., 2013), this is not a commonly used strategy for automated selection of mutations for design of high-affinity carbohydrate-binding proteins yet.

However, several tools have been developed for the prediction of carbohydrate binding sites on proteins and amino acids forming these binding sites (Kulharia et al., 2009; Malik et al., 2010; Shionyu-Mitsuyama et al., 2003; Taroni et al., 2000; Zhao et al., 2014). The prediction data set of these tools is thereby based on the protein-carbohydrate structures available in the protein data bank PDB (Berman et al., 2002).

Docking approaches are challenging for carbohydrates due to their high flexibility and the incorporation of water molecules in the binding site (DeMarco and Woods, 2008; Samsonov et al., 2011). Nevertheless, several approaches to incorporate scoring functions for carbohydrates in docking programs have been made (Kerzmann et al., 2008, 2008; Nivedha et al., 2014). Docking is mostly performed on small saccharide motifs from di- to tetrasaccharides. For example, docking of the SfY pentasaccharide to the binding antibody could not be performed successfully (Nivedha et al., 2014; Vyas et al., 2002).
An alternative to docking approaches is the calculation of alchemical binding free energies from molecular dynamics simulations. The method has been shown to be more accurate than Poisson-Boltzmann surface area methods (Gouda et al., 2003). Alchemical techniques have been successfully used for the calculation of binding free energies of carbohydrates with an error of ~4 kJ/mol (Mishra et al., 2015; Zacharias et al., 1993).

1.6 Aims of this study

Sf6 TSP as a specific receptor of the *Shigella flexneri* Y O-antigen was characterized as highly stable trimeric protein with an endorhamnosidase activity and an estimated dissociation constant to the octasaccharide O-antigen cleavage product in the millimolar range. However, the dissociation constant of the protein had not been exactly determined due to the low affinity. In contrast, simulations with Sf6 TSP were established, which successfully predicted the octasaccharide binding state (Kang et al., 2016). These suggested the protein as suitable for application of numerical tools for affinity prediction. Thus, on one hand Sf6 TSP is a potential candidate for the application and modification as *Shigella flexneri* Y specific bacterial sensor. On the other hand as a weak carbohydrate-binding protein, driving forces of the interactions and their affinity improvement could be studied. The work was therefore conducted with both experimental and computational based approaches.
For the characterization of the binding behavior of Sf6 TSP and the determination of the dissociation constant of Sf6 TSP-oligosaccharide complexes several methodologies as fluorescence spectroscopy or surface plasmon resonance were applied using *Shigella flexneri* Y polysaccharide ligands to increase the avidity of the binding event. The immobilization of the named polysaccharide was intended to achieve a mimic of the bacterial surface in a surface-based interaction assay and thereby to also obtain more insight about the natural binding architecture.

The application of Sf6 TSP as a bacterial sensor should be tested by the introduction of an environment sensitive fluorescent label in the binding site via thiol conjugation. Therefore the faster purification of Sf6 TSP was to be achieved by the establishment of an affinity chromatography based purification protocol to obtain cysteine mutants for fluorescent label conjugation. Furthermore, computational approaches by MD simulation and free energy calculations with the aim to obtain high affinity mutants of Sf6 TSP were to be tested.

The contribution of water molecules in carbohydrate-protein interactions in thermodynamically well characterized HK620 TSP mutants was investigated by MD simulation to gain more information about the enthalpic and entropic driving forces.

2 Material and Methods

2.1 Materials

2.1.1 Chemicals

All chemicals if not mentioned otherwise were purchased from Roth (Karlsruhe) or Merck (Darmstadt) including Sigma Aldrich (St. Louis, USA) with the purity of *pro analysis.* Solutions were prepared with pure water (ddH_2O) with conductivity lower than 0.055 µS/cm by a ELGA Purelab flex Viola water system.

2.1.2 Materials

Materials	Company
Centrifugal concentrator Amicon Ultra	Millipore (Schwalbach)
CMD200d Sensor chip	Xantec (Düsseldorf)
Dialysis tubes	Spectrum Labs (Houston, USA)
Microtiter plates, 96well, Nunc MAXI-SORP	Thermoscientific (Schwerte)
PD10 column	GE Healthcare (Freiburg)
Plastic cuvettes UV-Vis	Eppendorf (Hamburg)
QIAquick Gel Extraction Kit	Quiagen (Hilden)
QIAquick PCR purification Kit	Quiagen (Hilden)
Reaction tubes 1.5 ml/ 2.0 ml	Eppendorf (Hamburg) VWR (Darmstadt)
Roti®-Prep Plasmid MINI	Roth (Karlsuhe)
Sterile filter FP 30 (0.2/ 0.45 µm) CA-S	Whatman (Dassel)
Ultrafiltration membranes YM10, YM30	Millipore (Schwalbach)
Western Blot Filter Paper 1Chr	Whatman (Dassel)

2.1.3 Buffer and media

Buffer	Compounds
AP buffer	100 mM Tris/HCl pH 9.5; 100 mM NaCl; 5 mM $MgCl_2$
BCA solution A	54 mg/ml Na_2CO_3; 24 mg/ml $NaHCO_3$; 1.94 mg/ml Na^+_2-Bicinconinic acid salt (BCA)
BCA solution B	1.25 mg/ml $CuSO_4x5H_2O$; 1.25 mg/ml L-serine
Blotting buffer	48 mM Tris; 39 mM glycine
Buffer A	40 mM Tris/HCl, 200 mM NaCl; 50 mM imidazole
Buffer B	40 mM Tris/HCl, 200 mM NaCl; 400 mM imidazole
Coomassie stain	25 % (v/v) isopropanol; 10 % (v/v) acetic acid; 0.025-0.05 % (w/v) Coomassie Brillant blue (Serva, Heidelberg)
Coomassie destain	10 % (v/v) acetic acid
6x DNA loading buffer	0.25 % (w/v) bromophenol blue; 0.25 % (w/v) xylencyanol; 15 % (w/v) Ficoll Typ 400; 20 mM EDTA
HRP substrate solution	1x PBS with 2 mg/ml o-phenylene diamine and 0.03 % H_2O_2
2.5x Laemmli buffer (Laemmli, 1970)	163 mM Tris/HCl pH 6.8; 250 mg/ml glycerole; 5 % (w/v) SDS; 0.1 mg/ml DTT; 0.25 mg/ml bromophenol blue

Lysis buffer	25 mM Tris/HCl; 0.01 mM EDTA; 0.9 % (w/v) glucose; 0.04 mg/mL lysozyme; pH 8.0
PBS	150 mM NaCl; 10 mM sodium phosphate buffer pH 7.6
PBST	0.1 % Tween 20 in PBS
Rotiphoresegel 30	30 % acrylamide mix: 29.2 % acrylamide, 0.8 % bisacrylamide
SDS-PAGE running buffer	25 mM Tris/HCl pH 8.8; 192 mM glycine; 0.1 % (w/v) SDS
4x SDS-gel buffer 1	1.5 M Tris/HCl pH 8.8; 0.8 % (w/v) SDS
2x SDS-gel buffer 2	0.25 M Tris/HCl pH 6.8; 0.4 % (w/v) SDS
Silver stain developer	2.5 % (w/v) Na_2CO_3; 0.04 % (v/v) formaldehyde
Silver staining	0.5 ml 29 % (w/v) $AgNO_3$; 25 µl 37 % (v/v) formaldehyde in 100 ml dH2O
Silver stain fix solution 1	30 % (v/v) ethanol, 10 % (v/v) acetic acid
Silver stain fix solution 2	0.4 M sodium acetate; 0.5 % (v/v) acetic acid; 30 % (v/v) ethanol; 0.5 % (v/v) glutaraldehyde, 0.1 % (w/v) $Na_2S_2O_3$
Silver stain stop solution	50 mM EDTA; pH 8.0
1x TAE	40 mM Tris/Acetate pH 8,0; 1 mM EDTA
1x Tris/NaCl	40 mM Tris/HCl pH 8.0; 200 mM NaCl
1x TEC	40 mM Tris/HCl pH 8.0; 200 mM NaCl, 2 mM EDTA
TEV buffer	25 mM Tris/HCl pH 8.0, 125 mM NaCl, 1 mM EDTA, 1 mM DTT

Media	**Compounds**
AIM medium (Studier, 2005)	1 % tryptone; 0.5 % yeast extract; 25 mM Na_2HPO_4; 25 mM KH_2PO_4; 50 mM NH_4Cl; 5 mM Na_2SO_4; 2 mM $MgSO_4$; 0.2x Trace metals; 0.5 % glycerol; 0.05 % glucose; 0.2 % lactose (ZYM-5052-medium)
AIMamp	AIM medium with 0.1 mg/ml ampicillin
LB-agar	LB medium; 2 % (w/v) agar-agar
LBamp-agar	LB agar with 0.2 mg/ml ampicillin
LB medium	1 % (w/v) tryptone; 0.5 % (w/v) yeast extract; 1 % (w/v) NaCl; pH 7.0
LBamp	LB medium with 0.1 mg/ml ampicillin
1000x Trace metals	50 mM $FeCl_2$; 20 mM $CaCl_2$; 10 mM $MnCl_2$; 10 mM $ZnSO_4$; 2 mM $CoCl_2$; 2 mM $CuCl_2$; 2 mM $NiCl_2$; 2 mM Na_2MoO_4; 2 mM Na_2SeO_3; 2 mM H_3BO_3 in 60 mM HCl

2.1.4 Enzymes and Molecular Mass Marker

Enzyme	**Company**
Benzoase	Novagen (Merck) (Darmstadt)
Pfu Plus! DNA Polymerase	Roboklon (Berlin)
Proteinase K	Sigma Aldrich (St. Louis, USA)
Q5 polymerase	New England Biolabs (Schwalbach)
Q5 polymerase buffer	New England Biolabs (Schwalbach)
Restriction enzymes	New England Biolabs (Schwalbach)
Restriction enzyme buffers	New England Biolabs (Schwalbach)
T4 Ligase	Stratagene (La Jolla USA)
Taq polymerase	New England Biolabs (Schwalbach)
ThermoPol buffer	New England Biolabs (Schwalbach)

Marker	**Company**
1 kb DNA ladder	10 µl per lane, New England Biolabs (Schwalbach)
Color Prestained Protein Standard	Broad Range (11–245 kDa); 5 µl per lane, New England Biolabs (Schwalbach)

2.1.5 Primer

Name[a]	**Sequence**	**Description**
T7prom	5‘-*TAATACGACTCACTATAGGG*-3‘	*Forward (fw) sequencing primer for pET11a vector plasmid*
T7term	5‘-*GCTAGTTATTGCTCAGCGG*-3‘	*Reverse (rv) sequencing primer for pET11a vector plasmid*

Sf6seq1	5‘-*GATAGCGTAGAGGCTTATTGA*-3‘	*fw sequencing primer in the Sf6 TSP sequence (bp 690-709) (Barbirz, 2005)*
Sf6seq2	5‘-*AACCAATGCACATGGTCTCA*-3‘	*fw sequencing primer in the Sf6 TSP sequence (bp 1095-1114) (Barbirz, 2005)*
Sf6seq3	5‘-*AACACAGATCCTGTAA*-3‘	*fw sequencing primer in the Sf6 TSP sequence (bp 1511-1530) (Barbirz, 2005)*
SK5	5‘-CGCCCAACGACCGAAAACCTGTATT TTCAGGGTGCTAGCGACCCTGATC-3‘	Insertion primer (fw) coding His_6-tag+TEV protease cleavage sequence for plasmid vector 54 (see plasmids) including the TEV protein cleavage sequence
SK6	5‘-GCCGCAATCGTGATGGTGATGGTGA TGGTGATGCATATGTATATCTCCTTCTTAAA GTTAAAC-3‘	Insertion primer (rv) coding His_6-tag+TEV protease cleavage sequence for plasmid vector 54 (see plasmids) including start codon and His_6-tag sequence
SK7	5‘-TCCCTGCGAGCAACCAAGCGTGTGG GTATATTGTTGGATCCA-3‘	Mutagenesis primer (fw) for Sf6 TSP dN V204C
SK8	5‘-TGGATCCAACAATATACCCACACGC TTGGTTGCTCGCAGGGA-3‘	Mutagenesis primer (rv) for Sf6 TSP dN V204C
SK9	5‘-GTTAAGGCTGTCTTATGATTGTGAC ACTATCGGCCGTTATCTGAG-3‘	Mutagenesis primer (fw) for Sf6 TSP dN S246C
SK10	5‘-CTCAGATAACGGCCGATAGTGTCAC AATCATAAGACAGCCTTAAC-3‘	Mutagenesis primer (rv) for Sf6 TSP dN S246C
SK11	5‘-GCATGTCAGGCTGGATGCTGTAGAG TCGGTCTTC-3‘	Mutagenesis primer (fw) for Sf6 TSP dN T315C
SK12	5‘-GAAGACCGACTCTACAGCATCCAGC CTGACATGC-3‘	Mutagenesis primer (rv) for Sf6 TSP dN T315C
SK13	5‘-CCTGCCATTTCAGCAGAGGATGTTAT TCTGCTGATGAAAGC-3‘	Mutagenesis primer (fw) for Sf6 TSP dN N340C
SK14	5‘-GCTTTCATCAGCAGAATAACATCCTC TGCTGAAATGGCAGG-3‘	Mutagenesis primer (rv) for Sf6 TSP dN N340C
SK15	5‘-TATGGAACAACTGGATTTAGCTTGTT GCGGCTCAACAGGCGTG-3‘	Mutagenesis primer (fw) for Sf6 TSP dN Y400C with D399A mutation
SK16	5‘-CACGCCTGTTGAGCCGCAACAAGCT AAATCCAGTTGTTCCATA-3‘	Mutagenesis primer (rv) for Sf6 TSP dN Y400C with D399A mutation
SK17	5‘-GAACGCCGACCTCATGCCAGTCACA TAAAATTGTCAGTGGTG-3‘	Mutagenesis primer (fw) for Sf6 TSP dN T443C
SK18	5‘-CACCACTGACAATTTTATGTGACTGG CATGAGGTCGGCGTTC-3‘	Mutagenesis primer (rv) for Sf6 TSP dN T443C
SK21	5‘-AAGTGGCGCCCCAACGACCGAAAAC CTG-3‘	Primer (fw) with SfoI cleavage site and TEV protease cleavage sequence for the N-term. of Sf6 TSP dN in 54 vector plasmid

a Primer in italic were provided from former works in the lab.

2.1.6 Plasmids

No	Name[a]	Description
54	*pET11a-Sf6dN E366A D399A*	*Sf6 bacteriophage tailspike protein* (“Tail spike protein - Shigella phage Sf6 (*Shigella flexneri* bacteriophage VI),” n.d.) *gene AF128887.2 (1970 bp; aa: 1-623)* (ENA European Nucleotide Archive, 2017) *truncated by the N-terminal domaine (aa: 1-109; bp: 1-327) with introduced mutations for E366A (gaa->gca) and D399A (gat->gct) in pET11a with Amp resistance*
109	pET11a-Sf6dN E366A D399A_His	Modification of plasmid 54: introduction of N-term. His_6-tag sequence and TEV protease cleavage site by SK5/6
117	pET11a-Sf6dN E366A D399A V204C_His	Modification of plasmid 109: introduction of mutated amino acid V204C (gtt->TGt) by SK7/8
118	pET11a-Sf6dN E366A D399A T315C_His	Modification of plasmid 109: introduction of mutated amino acid T315C (aac->aTG) by SK11/12
120	pET11a-Sf6dN E366A D399A Y400C_His	Modification of plasmid 109: introduction of mutated amino acid Y400C (tat->tGt) by SK15/16
121	pET11a-Sf6dN E366A D399A N340C_His	Modification of plasmid 109: introduction of mutated amino acid N340C (aat->TGt) by SK13/14
122	pET11a-Sf6dN E366A D399A T443C_His	Modification of plasmid 109: introduction of mutated amino acid T443C (acg->TGC) by SK17/18
123	pET11a-Sf6dN E366A D399A	Modification of plasmid 109: introduction of mutated amino acid S246C

	S246C_His	(tca->tGT) by SK9/10
135	pPR-IBA102-Sf6dN E366A D399A	Sf6 TSP gene from plasmid 109 reproduced by PCR with SK3/21, cleaved with SfoI/EcoRI into cleaved plasmid vector pPR-IBA102 provides N-terminal truncated Sf6 tailpike protein with N-term. *Strep*-tag®II and the mutations E366A D399A.

a The derived plasmids are based on the commercial available plasmids pET11a (Novagen, Madison, USA) and pPR-IBA102 (iba-lifescience, Göttingen). Plasmids in italic were provided from former works in the lab.

2.1.7 Bacteria Strains

Bacteria strain	**Origin/Company**
E. coli BL21	Novagen (über Merck, Darmstadt): F- *omp*T *hsdSB* (rB-mB-) *gal dcm*(DE3) (Dubendorf and Studier, 1991)
E. coli DH5α	Invitrogen (Karlsruhe): F- φ80d*lac*ZΔM15Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17(rk-,mk+) *pho*A *sup*E44 λ- *thi*-1 *gyr*A96 *rel*A1.
E. coli XL1 blue	Stratagene (La Jolla, CA, USA): recA1 endA1 gyrA96 thi-1 hsdR17supE44 relA1 lac[F' *proAB lacIq ZΔM15* Tn10 (Tetr)]c (Bullock et al., 1987)
E. coli H TD 2158	Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures (Braunschweig, Germany) with DSM serial number 10809
Salmonella typhimurium 06-01900	National Reference Centre for *Salmonella* and other Bacterial Enterics, Robert Koch Institute, Wernigerode, Germany
Shigella flexneri Y 99-2001	
Shigella flexneri Y 03-650	
Shigella flexneri 2a 03-6557	
Shigella flexneri 2a 08-7230	

2.1.8 Instruments

Instrument	**Name**	**Company**
Absorption spectrometer	Chemstation (Typ 89090A)	Agilent (Böblingen)
	Cary 50 UV-VIS	Varian (Darmstadt)
	Pharmacia Biotech Ultraspec 1000	GE Healthcare (Freiburg)
	NovaspecII photometer	Pharmacia (New Jersey, USA)
Agarose gel electrophoresis	Agagel Mini	Biometra (Göttingen)
Balance	Mirco Scale B 120 S	Satorius (Göttingen)
	Lab scale PB	Kern (Ballingen-Frommen)
CD spectrometer	J-815	JASCO (Groß-Umstadt)
Centrifuge	5403; 5415C	Eppendorf (Hamburg)
	Megafuge 16R; Biofuge pico; Pico 17	Heraeus (Hanau)
	Sorvall Lynx 4000; Sorvall Discovery M150	ThermoFisher Scientific (Waltham, USA)
Centrifuge rotor	F12-6 x 500 LEX, GSA, A27 8x50, F28/50, GS3, TX-400, S55-A2	Sorvall (Buckinghamshire, UK)
Clean bench	HERAsafe	Heraeus (Hanau)
Fluorescence spectrometer	Fluoromax 3	Horiba Jobin Yvon (Unterhaching)
FPLC	Äkta FPLC	GE Healthcare (Freiburg)
	Äkta Prime	
	Äkta Purifier	
	UPC-900 monitor	
	P-920 pump	
French-Press	SLM-Aminco	Polytec (Waldbronn)
Gel documentation	Gel iX Imager	Intas Science Imaging (Göttingen)
Gelscanner	Imagescanner III	GE Healthcare (Freiburg)
HPLC	SCL10A system	Shimadzu (Kyoto, Japan)
	VP/LC10AD pump	
	SPD M10A absorption detector	
	RF-10A fluorescence detector	
Incubator	Innova 2300 platform shaker	New Brunswick Scientific (Enfield, USA)
	Innova 42	
Lyophylle	VaCo 5/2	Zirbus technology (Bad Grund)
Mass spectrometry	REFLEX™ 2	BRUKER (Bremen)

PCR Thermocycler	T Gradient, Tprofessional Trio	Biometra (Göttingen)
pH-Meter	597	WTW (Weilheim)
Pipettes	Finnpipette®	ThermoFisher Scientific (Waltham, USA)
Plate reader	Epoch G5 spectrometer	Epock, BioTEK, (Winooski, USA)
Pure water supply	PURELAB flex 3	ELGA Labwater (Celle)
Refractometer	OptiLab DSP	Wyatt Technology, Dernbach
SDS-PAGE electrophoresis	LKB 2050 MIDGET	Biorad (München)
Surface plasmon resonance	Biacore™ *J*	BIAcore AB (Uppsala, Sweden)
	SR7500DC	Reichert (Düsseldorf)
Thermoshaker		Eppendorf (Hamburg)
Ultrafiltration	Amicon	Millipore (Schwalbach)
Vacuum evaporator	Speedvac RVC 2-18	Christ (Osterode)
Vacuum pump	Trivac	Oerlikon Leybold Vacuum (Köln)
Voltage source	Bio 105, PP4000	Biometra (Göttingen)
Water thermostat	CC-K12/PilotONE	Huber, Offenburg

2.1.9 Software

Program	**Source/Company**
Amber12/ AmberTools13	Alma Rosa Agorilla; AMBER Software Administrator; MC2280; (San Francisco, USA) (Case et al., 2013, 2012)
BIAevaluation3.1	BIAcore AB (Uppsala, Sweden)
GROMACS 4.5.5/4.6.3/5.0.2	Gromacs (Uppsala, Sweden) (Abraham et al., 2015, 2015; Hess et al., 2008; Pronk et al., 2013; Van Der Spoel et al., 2005)
LabScan v6.0	GE Healthcare (Freiburg)
Mmass v5.5	Strohalm (Strohalm et al., 2008)
MobyWat	Hetényi (Jeszenői et al., 2016, 2015)
PyMol 1.8.2.3	Schroedinger LLC (Schrodinger, 2015)
SerialCloner2.6	Serial Basics
SigmaPlot v11.0/v13.0	Systat Software (Chicago, USA)
SpectraManager 1.54.03/2.06.00	Jasco (Groß-Umstadt)
Swiss PDBViewer4.0	www.expasy.org/spdbv/ (Guex and Peitsch, 1997)
TraceDrawer 1.7	Reichert, Inc. (Düsseldorf)
Unicorn 6	GE Healthcare (Freiburg)
VMD 1.9.3	Theoretical and Computational Biophysics Group, NIH Center for macromolecular modeling & bioinformatics (Illinois, USA) (Humphrey et al., 1996)

2.2 Molecular biological methods

2.2.1 Polymerase chain reaction (PCR)

Amplification PCR

PCRs were performed in a 50 µl reaction volume for the amplification of the Sf6 TSP ΔN E366A D399A gene with N-term. cleavage sites for TEV proteases and His_6-tag (SK5/6) and SfoI/EcoRI restriction enzymes (SK3/21). The PCR reaction mix included the following components:

5x Q5 buffer (NEB)	10 µl
10 µM dNTPs	1 µl
10 µM Primer sense/antisense	2.5 µl
Template DNA (5 ng end concentration)	1 µl
2 U/µl Q5 HF DNA Polymerase (NEB)	0.5 µl
ddH_2O (autoclaved)	32.5 µl
	50 µl

For amplification the lid was heated to 99 °C and DNA was denatured before amplification for 30 s at 98 °C. After the last cycle amplified PCR products were elongated for additional 5 min at 72 °C and the DNA was stored at 4 °C in the instrument. The following program was run with 25/35 cycles on a thermocycler for amplification.

Denaturation (98 °C)	10 s
Annealing (60/65 °C)	30 s
Elongation (72 °C)	3.5 min/45 s

Colony PCR

Colony PCRs were performed for verification of successful ligation of PCR products. Therefore transformed bacteria were picked by an autoclaved toothpick from a LBamp-agar plate and resuspended in 10 µl 0.96 % NaCl solution. The same toothpick was transferred into a colony numbered master LBamp-agar plate. The colony number master plate was incubated at 37 °C for ~3 h. For the colony PCR a reaction mixture of 10 µl was prepared with the following components:

10x ThermoPol reaction buffer (NEB)	1 µl
DMSO (Roth)	0.3 µl
10 mM dNTPs	0.2 µl
5 U/ml Taq DNA polymerase	0.1 µl
10 µM Primer (T7prom/term)	0.2 µl
Bacteria suspension	1 µl
ddH_2O (autoclaved)	7 µl
	10 µl

For amplification the thermocycler lid was heated on 99 °C and sample DNA was denatured for 30 s at 95 °C. After the end of the cycling program amplified DNA was elongated for 10 min at 68 °C and the samples were stored at 4 °C in the instrument. The cycling program was repeated for 26 cycles with the following program:

Denaturation (95 °C)	15 s
Annealing (55 °C)	20 s
Elongation (68 °C)	90 s

Mutagenesis-PCR

Mutagenesis PCR was performed for the introduction of single point mutation into the amplified plasmid. Controls were run with plasmid and without primer or *vice versa* for each combination. The reaction mixture included the following compounds:

10x Amplibuffer A (Roboklon)	2.5 µl
25 mM $MgSO_4$	2.5 µl
10 mM dNTPs	1 µl
10 mM Primer sense/antisense	1 µl
50 ng/µl Template DNA	0.3 µl
2 U/µl Q5 HF DNA Polymerase (NEB)	0.25 µl
ddH_2O (autoclaved)	16 µl
	25 µl

For amplification of the plasmid DNA, the thermocycler lid was heated to 99 °C and sample DNA was denatured for 5 min at 95 °C. After amplification cycles the product DNA was elongated for 10 min at 72 °C and stored at 4 °C in the instrument. The amplification cycle was repeated 17 times with the following protocol:

Denaturation (95 °C)	30 s
Amplification (55 °C)	1 min
Elongation (72 °C)	7 min

2.2.2 Agarose gel electrophoresis

Agarose gel electrophoresis for the separation of DNA was performed with 1 % agarose gels in 50 ml volume for a small chamber with Agarose Premium (Serva, Heidelberg). DNA stain was added to the gel with either 1:10000 1 % ethidium bromide (Roth) or 1:20000 HD green (Intas, Göttingen). DNA size was determined by the 1 kb-DNA ladder from NEB. Electrophoresis was run with 100 mA current for 30-60 min. Separated DNA was visualized by UV light (312 nm) in the Intas Gel iX Imager.

2.2.3 DNA digestion and ligation

Restriction enzyme digestion for ligation and sequence verification was performed with restriction enzymes and the corresponding buffers from NEB in 20 µl volume for 1 h at 37 °C. For the digestions NEB buffer was used in a 1:10 ratio, 1 mg/mL BSA, ~500 ng DNA and an enzyme concentration of about 5 U.

Amplified DNA from mutagenesis PCR was used in restriction digestion with *DpnI* to remove methylated template DNA with 1:10 NEB Cut smart buffer and 1:50 DpnI (20 U/µl) enzyme for 1 h at 37 °C.

DNA ligation in a 30 µl volume with 1:20 T4 ligase or Quick ligase and 1:2 of the corresponding buffer (T4 ligase buffer/ Quick ligase buffer) was incubated for 30 min or 10 min at room temperature (RT), respectively. DNA was provided with 1:8 ratio of template and insert in the ligation.

2.2.4 DNA purification, concentration measurement and sequencing

PCR product DNA was purified with QIAquick PCR purification Kit. Plasmid DNA was purified with the Roti®-Prep Plasmid MINI. Elution from the anion exchange column was performed with 50 µl ddH_2O (autoclaved).

DNA concentration and purity were measured by absorbance at 260 and 280 nm and DNA sequencing was performed on purified plasmid DNA with sequencing primers by GATC (Konstanz).

2.2.5 Transformation with *E. coli* BL21 and DH5α

Transformation was performed with chemical competent cell batches of *E. coli* BL21 for protein expression and *E. coli* DH5α/XL1-blue for plasmid purification. Batches of 100 µl bacteria suspension were mixed with ligation reaction samples or 1 µl of plasmid DNA and incubated on ice for 20 min. Bacteria batches were heated on 42 °C for 45 s and immediately cooled on ice for 2 min. The whole bacteria suspension was inoculated on a LBamp-agar plate in two dilutions and incubated over night at 37 °C.

2.2.6 Storage and cultivation of bacteria

For liquid culture 5 ml LB medium (±antibiotics) were inoculated with a colony from a LB-agar plate (±antibiotics). Bacteria strains were cultivated in liquid culture for 6 h with 120 rpm at 37 °C until a visible turbidity. The bacteria suspension was mixed to a 17.4 % glycerol end concentration and bacteria glycerol stocks are stored at -80 °C.

2.3 Protein biochemical methods

2.3.1 Overexpression of recombinant proteins in *E. coli* BL21

Pre-cultures of 10 ml LBamp were inoculated from a cryo glycerol stock or a single colony from a LBamp-agar plate from transformed *E. coli* BL21 and cultivated over night at 37 °C with 120 rpm. 2 L AIMamp (His_6-tagged proteins) or 5 L LBamp (*Strep*-tagged®II proteins) were inoculated 1:100 with pre-culture and cultivated at 37 °C and 120 rpm. Bacteria were grown to an OD_{600} of 0.7-1.0, subsequently the cultivation temperature was lowered to 30 °C and protein expression in LBamp was induced by 1 mM IPTG. Bacterial expression cultures were incubated over night at 30 °C with 120 rpm.

2.3.2 Harvest and lysis of bacteria

Bacterial liquid cultures were harvested by centrifugation with 5000 rpm (Rotor: F12 6x 500 Lex) at 4 °C for 15 min. Cell pellets were resuspended in 10 ml 1x Tris/NaCl with 0.25 mg/ml AEBSF protease inhibitor. Cells were lysed under pressure with 1000 psi three times in a French press cell. The supernatant of the cell lysate was separated from cell debris by centrifugation at 18000 rpm (Rotor: F28/50) at 4 °C for 1 h.

2.3.3 Protein purification and storage

Affinity chromatography based protein purification for His_6-tagged proteins

Protein purification was performed on an Äkta FPLC system (GE Healthcare). Affinity chromatography columns (Profinity™ IMAC resin, 30 ml column volume (CV), maximal pressure 0.7 MPa) were loaded with metal ions by equilibration with 50 mM sodium acetate buffer pH 4 with 0.3 M NaCl. Metal ions were loaded as 0.1 M $NiCl_2$ and residual ions were removed by washing with sodium acetate buffer, ddH_2O and buffer A in 5 CV each. The soluble protein fraction was diluted 1:5 with buffer A and loaded on the column with a flow of 1 ml/min and subsequently washed with 8 CV buffer A at 2 ml/min. 10 ml fractions were collected throughout the whole washing process of unbound proteins in chromatography tubes containing 100 µl 0.5 M EDTA pH 8 each. Proteins with affinity tag were eluted via a 10 CV linear gradient of 50-400 mM imidazole (Buffer A to B) with 1 ml/min flow. The protein content of collected fractions was monitored using absorption at 280 nm. Fractions containing the protein of interest were pooled and the His_6-tag was proteolytically removed by TEV protease cleavage with a molar protein-protease ratio of 1:100 overnight in TEV buffer at room temperature under agitation. The chromatography step described above was repeated with the cleaved protein solution.

Classical protein purification for *Strep*®II-tagged proteins

The protein lysate was precipitated by 35 % ammonium sulfate for 1.5 h at 4 °C under agitation. The protein purification was performed on an Äkta FPLC system via three chromatography steps. The protein precipitate was diluted in 1x TEC with 200 mM ammonium sulfate and loaded on a hydrophobic interaction chromatography column (HiPrep Phenyl FF (High Sub) 16/10, CV=20 ml) using a flow rate of 1.5 ml/min. Proteins were eluted using a linear gradient of 15 CV from 200-0 mM ammonium sulfate and fractions

containing the protein of interest were collected and pooled. The obtained protein solution was loaded on an anion exchange column (Q-sepharose activated with 2 M NaCl, CV=23 ml). Protein fractions were eluted in a linear gradient from 0-200 mM NaCl with 4 ml/min in 8 CV. Collected protein fractions were separated on a size exclusion chromatography column (HiLoad 16/60 Superdex 200, CV=120 ml) with 1 ml/min and an isocratic flow of 1x TEC with 200 mM NaCl and 200 mM Urea.
Protein concentrations were determined using absorption at 280 nm. Purified protein solutions were concentrated in 1x TEC (Amicon Filter; MWCO 30k) and stored as 35 % ammonium sulfate precipitates at 4 °C.

2.3.4 SDS-Polyacrylamide gel electrophoresis (SDS-PAGE)

For separation of denatured proteins a 10 % running and a 3 % stacking gel were prepared:

Compounds:	Running gel (10 %):	Stacking gel (3 %):
ddH_2O	4.15 ml	1 ml
30 % Acrylamide mix	3.35 ml	0.5 ml
4x SDS-gel buffer 1	2.5 ml	-
2x SDS-gel buffer 2	-	1.5 ml
10 % ammonium persulfate (APS)	100 µl	60 µl
Tetramethylethylenediamine (TEMED)	6 µl	3 µl

Samples were prepared by dilution of 2:3 with 2.5x Laemmli buffer and proteins were denatured at 100 °C for 10 min. 10-20 µl protein samples were loaded per lane and protein sizes were determined by comparing with 5 µl Color Protein Standard Marker per lane for Coomassie staining and 2.5 µl for silver staining per lane. Electrophoretic separation was performed for 20 min at 80 V and subsequently 60 min at 120 V.

2.3.5 SDS-polyacrylamide-gel staining

Coomassie staining

Protein bands with concentrations >1 µg on SDS-polyacrylamide-gels were incubated overnight on a shaker in Coomassie staining solution and destained in 10 % acetic acid for two times 1.5 h. Developed gels were documented via the gelscanner.

Silver staining

Protein bands with concentrations >10 ng were detected using the silver staining method for SDS-polyacrylamide-gels (Heukeshoven and Dernick, 1988). Gels were incubated with 135 rpm for 15 min in silver stain fix solution 1, followed by incubation in silver stain fix solution 2 for 30 min. After three washing steps in pure water for 5 min each, gels were stained in the silver stain solution for 30 min with 135 rpm. Residual silver stain solution was removed by one washing step with pure water and SDS-gels were developed in silver stain developer until protein bands were visible. The reaction was stopped by stop solution and developed gels were documented in the gelscanner.

2.3.6 Western-Blot analysis

Blotting of proteins on a PVDF membrane (Immobilon^{TM}-P, Millipore) was realized with the semidry blotting system from Biometra. All components and the anode and cathode platform of the blotter were wetted with blotting buffer. Five thin filter papers (Whatman), the nitrocellulose membrane, the gel and a second layer of five filter papers were stacked (9x6 cm). Proteins were blotted by electrical force for 1 h with 2 mA/cm^2 membrane.

Blotted protein size standard was marked on the membrane by a pen. The membrane was washed for 5 min under agitation with 1x PBS and blocked for 1 h at RT and 135 rpm or overnight at 4 °C with 1 % (w/v) BSA in 1x PBST. After three washing steps of 5 min with 1x PBST the first antibody (α-His_6 (mouse) antibody (mIgG) from Novagen (1:1000)) was added to the membrane and incubated for 1 h at RT or overnight at 4 °C in PBST. The membranes were washed three times for 5 min with PBST and subsequently incubated with the secondary antibody (α-mouse (goat) antibody (mIgG) conjugated with alkaline phosphatase from Sigma Aldrich (1:2000)) for 1 h at RT with 135 rpm. After two washing steps of 5 min each with 1x PBST and 1x PBS the membrane was incubated for 10 min in AP buffer.

The membrane was developed with 33 µl BCIP and 17.5 µl NBT (both 50 mg/mL in 70 % DMF) in AP buffer until protein bands were visible and development was stopped with ddH_2O. Developed membranes were documented by the gelscanner in reflective mode.

2.3.7 Fluorescece labeling of proteins

Cystein mutants of Sf6 TSP were labeled with an environment-sensitive fluorescent dye (Name of the dye not named as the project is in the patenting process). Protein samples (1 mg/ml) were reduced by 20 mM Tris(2-carboxyethyl)phosphine (TCEP) in 400 mM sodium phosphate pH 7 at 4 °C overnight with 150 rpm after the samples were degassed for 10 min at room temperature. The reducing agent was removed by ultrafiltration of samples in Amicon Ultra 4ml 30 K tubes for four times and 4000 xg. Reduced protein samples (1 mg/ml in 50 mM sodium phosphate pH 7) were fluorescently labeled in the dark with 20 % (v/v) 10 mM fluorescence dye in DMSO for 1 h at 56 °C with 150 rpm. Free dye was removed by size exclusion chromatography on a PD10 column in the dark.
The purity of labeled protein solutions was determined by size exclusion chromatography with a Superdex 200 10/300 column (CV=24 ml) on a Shimadzu HPLC system with absorption detector (PDA) between 260-300 nm and fluorescence detector (Detector A) excitation at 492 nm and emission at 545 nm. Labeled protein samples (1 mg/ml) were centrifuged for 20 min at 40000 rpm (SS-55-A2) and 50 µl sample was injected on the column and eluted for 60 min and maximal pressure of 1.5 MPa in an isocratic flow of 50 mM sodium phosphate pH 7.0.

2.4 Carbohydrate biochemical methods

2.4.1 Purification of polysaccharides

Polysaccharide was purified from 10 L bacteria culture (LB medium over night at 37 °C) of *Salmonella typhimurium* 06-01900 or *Shigella flexneri* Y 99-2001. Bacteria cultivation and harvest was performed as described in section 2.2.6 and 2.3.2. LPS of *Shigella flexneri* Y was provided by Nils Carlin (Scandinavia Biopharma). Due to impurities the LPS was purified and cleaved to polysaccharide with the same protocol as described in the following for bacteria.

Bacteria pellets were washed twice with 10 mM Tris/HCl, 2 mM $MgCl_2$ pH 7.6 and twice with water before cells or LPS were resuspended in 100 ml 10 % acetic acid. Bacteria or LPS were hydrolyzed for 1.5 h at 99 °C and samples were cooled on ice. LPS hydrolysis samples were frozen at -40 °C overnight. Hydrolysis batches were centrifuged at 4 °C for 20 min and 8000 rpm (A27-8x50). The hydrolysis process was repeated with the bacterial pellet. The obtained pellet was washed twice with 1 % acetic acid and centrifuged supernatants were collected together. Supernatants were dialyzed against running tap water for at least three days or adjusted to pH 7.0 with 1 M NaOH. Precipitates in the supernatant were removed by centrifugation with 8000 rpm at 4 °C for 20 min. Polysaccharide in the supernatant was precipitated in 80 % ethanol at -40 °C for 2 h. Precipitates were centrifuged with 18000 rpm at 4 °C for 30 min (F12-6x 500 LEX). Pellets were resolubilized in 10 mM Tris/HCl, 4 mM $MgCl_2$ pH 7.8 and the remaining precipitates were removed by centrifugation. DNA impurities were cleaved by 2 U/µl benzonase for 3 h at 37 °C and subsequent protein impurities were cleaved with 15 µg/ml Proteinase K for 3 h at 65 °C or overnight at 37 °C. Ethanol precipitation was repeated until no pellet was formed after dilution of the precipitate in water. Obtained polysaccharide samples were separated by anion exchange chromatography from residual impurities either on a DEAE-sepharose column (CV=40 ml) or on a Q-sepharose fast flow column (CV=35 ml).

Prior to loading on the DEAE column (activated with 2 M NaCl), polysaccharide was diluted in 1x TEC after the last precipitation step. Elution was achieved with 4 CV 1x TEC.

Prior to loading, the Q-sepharose column was equilibrated with 10 mM ammonium sulfate. Polysaccharide was diluted in 10 mM ammonium sulfate and loaded on the column. The column was washed with 2 CV 10 mM ammonium sulfate and polysaccharide was eluted with a linear gradient from 0-1 M NaCl in 2 CV. Incident polysaccharide was monitored by changes in the refractive index using the OptiLab DSP refractometer.

Fractions from both chromatographic methods were precipitated with ethanol and the precipitate was diluted in 1 ml ddH_2O and freeze-dried. Polysaccharide was quantified by mass and purity was determined for 1 mg/ml samples by absorption measurements at 260 and 280 nm. Samples were considered pure if absorption was smaller than 0.1 for both wavelengths.

2.4.2 Oxidation of polysaccharide for surface activation

Polysaccharide was activated for surface immobilization by oxidation using sodium periodate with two different strategies. *Shigella flexneri* Y polysaccharide was prepared as 100 µl of 10 mg/ml solutions in ddH_2O and subsequently diluted in 900 µl 10 mM sodium phosphate pH 6.2 (sample 1) or 20 mM sodium acetate, 150 mM NaCl pH 5.0 (sample 2) and 100 µl 100 mM sodium periodate in ddH_2O for oxidation (total volume 1 ml) (Gedig, 2008; Wolfe and Hage, 1995). Samples 1 and 2 were incubated at 130 rpm for 30 min at 25 °C and 30 min at 4 °C, respectively. The oxidation process was stopped with 250 µl ethylene glycol and oxidized polysaccharide was directly separated from the oxidation products by size exclusion chromatography on a PD10 column with elution using ddH_2O. Activated polysaccharide samples were concentrated using an Amicon 4 Ultrafiltration 4k tube with 4000 g to a volume of 100 µl and stored at -20 °C.

2.4.3 Determination of reducing ends by the BCA method

The number of activated aldehyde or ketone groups on oxidized polysaccharides was determined by the BCA method. As standard a concentration series of glucose with the concentrations 0, 2.5, 10, 20, 30, 40 and 50 µM was used. BCA solution 1 and 2 were mixed in a 1:2 ratio to obtain the BCA working solution which was stored in the dark. Carbohydrate samples or glucose standard and BCA working solution were mixed in a 1:2 ratio and incubated for 15 min at 100 °C. Absorption was measured at 560 nm.

2.4.4 Production of oligosaccharides

Shigella flexneri Y polysaccharide was diluted in 1x TEC and incubated with 50 µg/ml Sf6 TSP*wt* overnight at RT at 120 rpm. The polysaccharide digest was freeze dried. Dried samples were diluted in 50 µl ddH_2O and loaded on a HiLoad 26/60 Superdex 30 column (CV=300 ml) equilibrated with ddH_2O. Oligosaccharides were eluted using an isocratic run with ddH_2O and were monitored via their refractive indeces. Fractions from different populations were pooled and freeze-dried separately. Each fraction was chromatographed a second time and freeze dried oligosaccharides were quantified by mass spectrometry.

2.4.5 MALDI-TOF mass spectrometry for oligosaccharide characterization

Freeze dried oligosaccharide samples were dissolved in 6 µl 0.1 % trifluoro actic acid (TFA) and 4 µl acetonitrile. A DHB/DMA matrix was used including 30 mg 2,5-dihydroxy acetophenone (DHB) in 1 ml 50 % acetonitrile and 0.1 % TFA (Ropartz et al., 2011). A mixture of 4 µl oligosaccharide sample and 4 µl matrix were spotted on the MALDI sample tray and crystallized. Ionization was performed in the positive mode and mass spectra were analyzed by Mmass.

2.5 Tailspike adsorption assay ELITA

The ELISA-like tailspike adsorption assay was developed by Schmidt et al., 2016. Bacteria were grown in 10 ml LB medium until an OD_{600} of 0.7-1.0. Bacteria were harvested by centrifugation (TX-400: 4000 g, 10 °C, 10 min) and washed twice with 1x PBS. Bacteria were inactivated with 2 % glutaraldehyde for 1 h and washed twice with 1x PBS and finally resuspended in 1x PBS with an OD_{600} of 0.1. 200 µl bacteria suspensions were added to the wells of a 96-well microtiter plate (Nunc F Maxisorp) and were incubated for adsorption overnight at 4 °C. Supernatants were removed and adsorbed bacteria were washed twice with 210 µl 1x PBS. Each well was blocked with 220 µl 1 % BSA in 1x PBS for 2 h at RT and washed once with 220 µl 1x PBS. Tailspike adsorption was achieved by incubation of each well with 190 µl 30-80 nM *Strep*-tag®II TSP (inactive) for 20 min with 130 rpm. Unbound tailspike proteins were removed by three washing steps with 200 µl 1x PBST for 4 min each. For detection 190 µl HRP-*Strep*-Tactin® conjugate (IBA, Göttingen) (1:10000 in 1x PBST) was incubated in each well for 15 min and 130 rpm. Unbound conjugate was removed by three washings steps with 200 µl 1x PBST and one additional washing step with 210 µl 1x PBS. ELITA assays were developed with 200 µl 1 mg/ml o-phenylenediamine with 0.03 % H_2O_2 in 50 mM citrate buffer pH 6.0 per well for 5 min. The reaction was stopped with 50 µl 2 M H_2SO_4 per well and the adsorption was measured within 5 min at 492 nm in the microtiter plate reader (Epoch, BioTek). Measured absorptions were measured in triplicates and mean values with standard deviations were depicted.

2.6 Biophysical methods

2.6.1 Fluorescence spectroscopy

Fluorescence spectroscopy steady state and kinetic measurements for monitoring changes in intrinsic protein fluorescence were performed as described before (Kang et al., 2016). The measurements were performed on a Fluoromax 3 spectrofluorimeter (HORIBA Jobin Yvon, Bensheim, Germany). Full volume cuvettes (poly(methyl methacrylate) (PMMA)) were filled with 3 ml 184 nM TSP in 50 mM sodium phosphate pH 7.0 (filtered by 0.45 pore size and degassed). Measurements were performed at 10 °C with stirring at level 4. Protein samples (1 mg/ml) were cleaned from particles by centrifugation with 40000 rpm for 20 min at 15 °C (SS55-A2) prior to measurements.

Steady state measurements

Intrinsic protein fluorescence was excited at λ_{Ex}=280 nm or 295 nm and emission was monitored between λ=300-450 nm. Extrinsic fluorescence of the fluorescent label was excited at λ_{Ex}=492 nm and emission spectra were monitored between λ=500-650 nm. In both cases excitation and emission slit widths of 2.4 and 6.4 nm were used, respectively.

Polysaccharide solutions (10 mg/ml in ddH_2O) were added in 1,3 or 10 µl volumes and spectra were recorded after 3 min incubation time. Spectra were corrected for dilution and fluorescence of polysaccharide impurities using control measurements with addition of polysaccharide to buffer samples.

The same procedure was used for synthetic Sf2a oligosaccharide samples (provided by Laurence Mulard, Institute Pasteur, France) without correction for oligosaccharide impurities during data processing (limited sample amount). Emission maxima for oligosaccharide titration were monitored at λ_{Em}=340 nm and binding constants were calculated using the following formula (Baxa et al., 1996):

$$\Delta F = 1 + \left(\frac{\Delta F_{max}}{2[B]_{ges}} \cdot \left[\left(K_D + [B]_{ges} + [L]_0\right) - \sqrt{\left(K_D + [B]_{ges} + [L]_0\right)^2 - 4 \cdot [L]_0[B]_{ges}}\right]\right)$$

ΔF is the quotient of the measured fluorescence and fluorescence of the unbound protein, ΔF_{max} is the relative maximal fluorescence change in complex saturation, $[B]_{ges}$ is the total concentration of tailspike protein binding sites and $[L]_0$ the total ligand concentration. The dissociation constant K_D is derived by this equation.

Kinetic measurements

Fluorescence changes of protein and label were used to monitor polysaccharide binding kinetics at λ_{Em}=336 nm and λ_{Em}=540 nm, respectively. The slit width was 2.5 nm for excitation and 5 nm for emission. Traces were monitored with a death time of 4 s. Kinetics were fitted by a biexponential equation for the apparent rate constants $k_{obs1/2}$ and the maximal responses $a_{max1/2}$:

$$f(x) = a_{max,1} \cdot (1 - e^{-x \cdot k_{obs,1}}) + a_{max,2} \cdot (1 - e^{-x \cdot k_{obs,2}}) + 1$$

Due to the heterogeneity of polysaccharide the molecular mass and dissociation constant of the complex cannot be calculated. Therefore only the dissociation rate constant k_{diss} could be obtained from apparent rate constants $k_{obs,2}$ plotted against the polysaccharide concentration. The apparent rate $k_{obs,1}$ is measured due to manual mixing. The described equation was used for the linear fit with [P] and [L] for the protein and ligand concentration, respectively (Baxa et al., 2001):

$$k_{obs} = k_{diss} + k_{diss}([P] + [L])$$

2.6.2 Circular dichroism spectroscopy

Near-UV circular dichroism (CD) spectra were recorded on a J815 Jasco CD spectropolarimeter in a 5 mm pathlength quartz glass cuvette. Protein solutions (>2 mg/ml) were separated from particles by centrifugation with 40000 rpm for 20 min at 15 °C (SS55-A2). 50 mM sodium phosphate buffer (pH 7.0) was filtered (pore size 0.45 µm) and degassed prior to measurements. CD-spectra were recorded using protein concentrations of 2 mg/ml at 20 °C. Spectra were monitored from 350-250 nm with three accumulations with an increment and band width of 1 nm each and a response time of 8 s. Spectra were corrected for buffer signals.

The molar ellipticity [Θ_{MRW}] was derived by the described formula (Creighton, 1997) with Θ being the measured ellipticity, M_w being the protein molecular mass, c being the protein concentration in mg/ml, l being the cuvette thickness in cm and N_A being the number of amino acids:

$$[\Theta_{MRW}] = \frac{\Theta \cdot 100 \cdot M_w}{c \cdot l \cdot N_A}$$

2.6.3 Surface Plasmon Resonance

Lipopolysaccharide-Chip preparation via hydrophobic interaction

Immobilization was performed on the BIAcore *J* instrument with filtered (0.45 µm pore size) and degassed 50 mM sodium phosphate 4 mM $MgCl_2$ buffer pH 7.0. LPS was diluted in the running buffer and extruded to homogeneously sized micelles for 21 turns at 45 °C (MacDonald et al., 1991). Flow rates in the BIAcore *J* are not further quantified than low, medium and fast by the instrument instructions. The used L1-Chip (BIAcore) was cleaned by 12 min injection of 40 mM CHAPS with a low flow rate at 25 °C. LPS was immobilized by 12 min. injection of extruded 5 mg/ml LPS in running buffer at low flow rate and subsequent 3 min. injection of 10 mM NaOH and 1 mM NaCl, followed by 2 min injection of 1 % BSA to block free binding sites. SfY LPS and *Salmonella typhimurium* LPS were immobilized on measurement and reference channel, respectively.

Polysaccharide-Chip preparation via phosphoethanolamine modifications

Immobilization was performed on the BIAcore *J* instrument with filtered (0.45 µm pore size) and degassed 10 mM sodium phosphate buffer pH 7.0 following the described protocol (Gedig, 2008). For immobilization, the carboxymethyldextran surface chip CMD200d (Xantec, Düsseldorf) was inserted into the instrument and the flow was put on low level at 25 °C. Freshly prepared activation solution of 0.05 M NHS (N-hydroxysuccinimde) and 0.2 M EDC (1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide) in 10 mM sodium phosphate pH 7.0 was injected passing the sensor chip for 7 min. After a short buffer wash a 5 mg/ml polysaccharide solution in ddH_2O was injected for 5 min. Residual reactive groups on the chip surface were inactivated with a 2 min injection of 1 M ethanolamine.

Polysaccharide-Chip preparation via activated carbohydrates

Immobilization with activated polysaccharide was performed on a Xantec SPR instrument following the described protocol (Gedig, 2008). The running buffer of 10 mM sodium phosphate pH 6.0 and all used solutions for the immobilization were filtered (0.45 µm pore size) and degassed. The carboxymethyldextran surface chip CMD200d from Xantec was inserted into the instrument and the instrument was equilibrated at 25 °C. The surface was cleaned for 10 min with a flow rate of 7 µl/min with 2 M NaCl/10 mM NaOH and washed with buffer for 3 min. After each injection a pump wash was performed. For activation of carboxyl groups on the chip a freshly prepared mixture of 0.05 M NHS/0.2 M EDC in 0.5 M

MES (2-(N-morpholino)ethane sulfonic acid) buffer pH 5.5 was injected for 3 min with a flow rate of 23 µl/min. The surface was washed with buffer for 1 min. For hydrazide activation, 15 min injection of 0.1 M adipic acid dihydrazide hydrochloride pH 8.0 was performed with 4 µl/min flow rate. After surface equilibration for 5 min with buffer, residual NHS esters were quenched for 30 min with 1 M ethanolamine with 2 µl/min flow rate. The surface was equilibrated with buffer for 10 min. Activated polysaccharide (10 mg/ml) was injected for 20 min with a flow rate of 4 µl/min followed by 7 min washing and the reduction of the formed bond for 20 min with 50 mM sodium cyanoborhydride in 0.1 M acetate buffer pH 4.0 with the same flow rate. The surface was equilibrated with buffer until no signal drift was visible.

Interaction measurements and analysis

Interaction experiments at the BIAcore J 2000 instrument were performed with a medium flow rate at 25 °C with running buffer (50 mM sodium phosphate, 200 mM NaCl, 0.005 % Tween 20 pH 7.0). Protein samples were dialyzed in running buffer and centrifuged for 20 min with 40000 rpm (SS55-A2) for removal of air bubbles and precipitates prior to measurements. Several protein concentrations were injected for 3 min over the surface. After equilibration of the dissociation signal of 2 min, the chip surface was regenerated with 100 mM sodium acetate pH 4.0 for 3 min and running buffer for 1 min. For LPS-covered surfaces, an injection of 2 min with 1 % BSA solution was performed after each interaction measurement. Sensograms were processed with the program BIAevaluation3.1 and equilibrium data were fitted by a single ligand binding site model for the dissociation constant K_D with the maximal response units RU_{max} of the sensograms, the ligand concentration [L] and the maximal binding concentration B_{max}:

$$RU_{max} = \frac{B_{max}[L]}{K_D + [L]}$$

Interaction experiments at the Xantec instrument were performed with 20 µl/min flow rate and with 50 mM sodium phosphate pH 7.0 as running buffer. The instrument was equilibrated at 25 °C. Protein samples prepared as stated above were injected for 3 min over the surface and dissocation was equilibrated for 5 min. The surface was regenerated by a 4 min injection with 100 mM sodium acetate pH 4.0 and 1 min of running buffer. Data were processed with the program TraceDrawer1.7 or BIAevaluation3.1 and association and dissociation rate constants were fitted based on models for bivalent and heterogeneous binding models provided by the program (in detail with Fig. S 3 in the Appendix).

2.7 Computational methods

2.7.1 Alignment of protein structures

The published crystal structures of HK620 TSP mutants (pdbID: 2VJI, 2VJJ (Barbirz et al., 2008), 2X6W, 2X6X, 2X6Y, 2X85, 4AVZ (Broeker et al., 2013), 4PAT, 4XKW, 4XKV, 4XLA, 4XL9, 4XLF, 4XLE, 4XLC, 4XMY, 4YEL, 4XLH, 4XM3, 4XN3, 4XN0, 4YEJ, 4XR6, 4XNF, 4XOR, 4XOP, 4XON, 4XOT, 4XQ7, 4XQH, 4XQI (Kunstmann et al., *in preparation*) or derived structures from MD simulations were aligned with *alignto* to the WT-apo structure as reference by Pymol (Schrodinger, 2015). Distances, angles and torsions between amino acid side chains were determined with the measurement tool of Pymol. Hydrogen bonds were defined by the *distance* tool with 3.5 Å maximal distance and mode=2 selected for only polar groups as donor or acceptor.

2.7.2 Structure preparation for Molecular dynamic simulations (MD)

Structures of HK620TSP E372A were taken from crystal structures with one protein subunit with penta- and hexasaccharide ligands (pdbID: 4XM3, 4XN3). Water molecules and ligands were removed. Structures were parameterized with the AMBER03 force field for proteins (Duan et al., 2003) and the GLYCAM06 force field for glycans (Kirschner et al., 2008). Glycam parameters were processed into a Gromacs compatible format by *glycam2gmx.pl* (provided by M. Wehle). The penta- or hexasaccharide ligand was fitted in the binding site by superposition with the original ligand of the crystal structure by the script *rotate_octamer.f90* (provided by Y. Kang). The complex was placed in an orthogonal box with side lengths of approximately 70x80x150 Å filled with ~25900-30000 water molecules of type TIP3P, TIP4P or TIP5P (Tab. 2) (Jorgensen et al., 1983; Mahoney and Jorgensen, 2000). The system was neutralized with 10 Na^+-ions assuming a total negative charge of -10 using the *genion* routine.

Tab. 2: Details of HK620 TSP MD simulations

	W3	*W4*	*W5*	*P3*	*P4*	*P5*	*H3*	*H4*	*H5*
Protein (pdbID)	4XM3	4XM3	4XM3	4XM3	4XM3	4XM3	4XN3	4XN3	4XN3
Ligand[a]	-	-	-	Pent	Pent	Pent	Hex	Hex	Hex
No. atoms (system)	86828	112227	156007	86797	128158	155896	98320	127331	155652
No. atoms (protein)	8798	8797	8797	8797	8797	8797	8797	8797	8797
No. atoms (ligand)	-	-	-	119	119	119	140	140	140
No. (water)	26003	25851	29440	25957	29808	29394	29791	29596	29341
No. (Na^+)	10	10	10	10	10	10	10	10	10

Pent: -2)-α-L-Rha-(1-6)-α-D-Glc-(1-4)-[β-D-GlcNAc-(1-3)]-α-D-Gal-(1-3)-α-D-GlcNAc(1-
Hex: -2)-α-L-Rha-(1-6)-α-D-Glc-(1-4)-[β-D-GlcNAc-(1-3)]-α-D-Gal-(1-3)-[α-D-Glc-(1-6)]-α-D-GlcNAc(1-

Simulations with Sf6 TSP were based on two subunits with one binding site cleft of the inactive mutant E366A D399A (EADA) (pdbID: 4URR) (Kang et al., 2016). Cysteine mutations were introduced by atom name changes in the pdb-file before parametrization (Tab. 3). Dodecasaccharide geometry and parameters were provided by Y. Kang (Kang et al., 2016). The simulation complex was placed in an orthogonal box with the dimensions ~140x100x120 $Å^3$. Protein structure and ligand preparation were carried out as described for HK620 TSP with the TIP3P water model (Jorgensen et al., 1983).

Tab. 3: Details to Sf6 TSP MD simulations

	EADA	*EADAdod*	*V204C*	*S246C*	*T315C*	*N340C*	*Y400C*	*T443C*	*T443Cdod*
Protein (pdbID)	4URR	4URR	4URR	4URR	4URR	4URR	4URR	4URR	4URR
Ligand	Oct	Dodeca	Oct	Oct	Oct	Oct	Oct	Oct	Dodeca
No. atoms (system)	176007	176001	176015	176004	176001	176004	175999	176001	176013
No. atoms (protein)	15102	15102	15092	15102	15096	15096	15082	15096	15096
No. atoms (ligand)	177	264	177	177	177	177	177	177	264
No. (water)	53570	53539	53576	53569	53570	53571	53574	53570	53545
No. (Na^+)	18	18	18	18	18	18	18	18	18

Oct: -[3)-α-L-Rha-(1-3)-α-D-GlcNAc-(1-2)-α-L-Rha-(1-2)-α-L-Rha(1]^2-
Dodeca: -[3)-α-L-Rha-(1-3)-α-D-GlcNAc-(1-2)-α-L-Rha-(1-2)-α-L-Rha(1]^3-

2.7.3 Energy minimization, equilibration and MD simulation

MD simulations were carried out using the GROMACS4.5.5 program package (Hess et al., 2008; Pronk et al., 2013).

For HK620 TSP, after energy minimization to a maximum force smaller than 1000 kJ/mol/nm (steep descent) the system was equilibrated in two simulations with 100 ps each. All simulations were run under isothermal-isobaric (NPT) ensemble conditions with Parrinello-Rahman barostat coupling (reference pressure of 1.0 bar and coupling time constant of 0.5 ps) (Parrinello, 1981) and a Langevin thermostat (reference temperature 298 K and coupling time constant 1.0 ps). Simulation time step was 2 fs. Hydrogen bonds and protein backbone were constrained using the LINCS algorithm in GROMACS with 1000 kJ mol^{-1} $Å^{-2}$ (Hess et al., 1997).The systems were simulated without ligand for 50 ns and with ligand for 100 ns.

For Sf6 TSP, the protein backbone was constrained with 1000 kJ mol^{-1} $Å^{-2}$ except for turns and coils, which were kept flexible. The energy minimization, equilibration and MD simulation were performed as described for HK620 TSP with equilibrations for 1 ns and MD simulations for 100 ns each.

2.7.4 Free energy calculations (FE)

FE calculations of distinct water molecules

Free energy calculations were based on the double-decoupling method from Hamelberg and McCammon, 2004 and Samsonov et al., 2008. Nine crystal structures of HK620 TSP were used for 10 ns simulations (Tab. 4). Ligands were removed from the crystal structure and the protein was limited to the area of 12 Å distance to the hexasaccharide ligand (pdbID: 2VJJ), resulting in a fragment of amino acids 206-521 with constrained backbone, except for the loop region from amino acid 469-472. Crystal water molecules were removed from the crystal structure except for water molecules in superposition of Glc2 of the hexasaccharide ligand. The protein was parametrized (Amber03), the parametrized ligand (GLYCAM06) was superimposed on the crystallographic ligand position. The complex was placed in a cubic box with side length of 80 Å filled with around 15600 TIP3P water molecules (Jorgensen et al., 1983). The system was neutralized with 5-6 Na^+-ions using the *genion* routine.

Tab. 4: Details of 10 ns MD simulations of HK620 TSP

	WT-apo	***WT-pent***	***WT-hex***	***EQ-apo***	***EQ-pent***	***EQ-hex***	***EA-apo***	***EA-pent***	***EA-hex***
Protein (pdbID)	2VJI	4PAT	2VJJ	4AVZ	4XOT	2X6W	4XN0	4XM3	4XN3
Ligand	-	Pent	Hex	-	Pent	Hex	-	Pent	Hex
No. atoms (system)	51520	51528	51537	51500	51547	51514	51514	51534	51561
No. atoms (protein)	4616	4615	4616	4617	4617	4617	4610	4610	4610
No. atoms (ligand)	-	119	140	-	119	140	-	119	140
No. (water)	15633	15596	15592	15626	15602	15584	15633	15600	15602
No. (Na^+)	5	6	5	5	5	5	5	5	5

Pent: -2)-α-L-Rha-(1-6)-α-D-Glc-(1-4)-[β-D-GlcNAc-(1-3)]-α-D-Gal-(1-3)-α-D-GlcNAc(1-
Hex: -2)-α-L-Rha-(1-6)-α-D-Glc-(1-4)-[β-D-GlcNAc-(1-3)]-α-D-Gal-(1-3)-[α-D-Glc-(1-6)]-α-D-GlcNAc(1-

Energy minimization until 1000 kJ/mol/nm (steepest descent) with 500 kcal/mol $Å^2$ harmonic restraints on protein and ligand atoms was performed, followed by a NVT ensemble (Langevin-thermostat) with a harmonic restraint of 100 kcal/mol $Å^2$ on the protein-ligand complex and another energy minimization (steepest descent) with 100 kcal/mol $Å^2$ harmonic constraints on the complex. Subsequently, the whole system with only backbone restraints was equilibrated for 10 ps in the NVT ensemble (Langevin-thermostat) and 30 ps for the NPT ensemble (Parrinello-Rahman barostat). MD simulations were carried out for 10 ns.

Root mean square fluctuations of crystal water molecules were calculated from 1 ns simulation by *g_rms* to determine the atomic fluctuation <RMSF> of the respective water oxygen. Atomic fluctuations were converted into the harmonic potential k to constrain the respective water coordinate using the following formula with R=8.134 10^{-3} kJ mol^{-1} K^{-1} and T=298 K (Hamelberg and McCammon, 2004):

$$k = \frac{3RT}{< RMSF >^2}$$

The decoupling for single water molecules each was performed with 21 *lambda* (λ) points each for Coulomb, followed by van der Waals (vdW) interactions in forward and reverse directions. Steps were separated between λ=0 and λ=1 in Δλ=0.05. To avoid end-point singularities, vdW interactions were treated via soft-core interactions with a soft-core parameter α=0.5, an exponent of 1 and σ=0.3 nm. Integration was performed with the *g_bar* (Bennett acceptance ratio) implementation of Gromacs/4.5.5 (Bennett, 1976; Pronk et al., 2013). The free energy differences were calculated using: $\Delta\Delta G=\Delta G_{2(\text{Water in Water})}-\Delta G_{1(\text{Water in Protein})}$. Hereby ΔG_2 was derived from a separately conducted simulation with decoupling one water molecule in a box filled with TIP3P water with the same size as for protein including simulations. FE calculation of ΔG_2 for water from water resulted in 6.3 kcal/mol, confirming results reported in literature (Ben-Naim and Marcus, 1984). ΔG_1 was calculated by the given formula (Hamelberg and McCammon, 2004):

$$\Delta G_1 = \int \langle \frac{\partial U(\lambda, r_A, r_B, \zeta_B, r_S)}{\partial \lambda} \rangle_\lambda \, d\lambda - RTln\left(\frac{\sigma_{AB}}{\sigma_A \sigma_B}\right) + RTln(C^0 V_i) + RTln(\xi_i/8\pi^2) + P^0(\bar{V}_A - \bar{V}_{AB}) \quad (1)$$

The formula (1) is separated into five product terms. The first product term of the equation is derived from the result of the FE calculations ΔG_{FE}. The second term describes the symmetry of water, which is assumed symmetrical with a symmetry number of $\sigma_B=2$ and the whole term results in 1.68 kJ/mol. The third term is influenced by the value of k. The standard concentration C^0 for 1 mol/l resulted in 6.023 10^{-4} $Å^{-3}$. The standard volume V_i was calculated with the given constants using the following formula:

$$V_i = (\frac{2\pi RT}{k})^{3/2}$$

The fourth term results in zero as it describes the rotational space of the water molecule, which is allowed to rotate freely. The fifth and last term describes the change in volume when the water is decoupled from the system which is negligible in constant normal pressure simulations (Gilson et al., 1997). In conclusion, ΔG_1 was calculated using the following values in formula (1):

$$\Delta G_1 = \Delta G_{FE} + 1.68 \frac{kJ}{mol} + RT \ln[6.023\ 10^{-4} \text{Å}^{-3} \cdot (\frac{2}{3}\pi < RMSF >^2)^{3/2}]$$

FE calculation of side chain residues and ligand

The simulations were set up identical to the one for water molecule FE calculations.

The FE MD simulations were carried out with the same conditions and force fields as described for the elongated MD simulation for 10 ns with constrained protein backbone. The FE was performed with 21 *lambda* (λ) points each for decoupling Coulomb followed by van der Waals (vdW) interactions. These steps were separated between λ=0 and λ=1 in 0.05 steps. To avoid end-point singularities, vdW interactions were treated via soft-core interactions with a soft-core parameter α=0.5, an exponent of 1 and σ=0.3 nm.

Parameters for mutated residues GLU372→GLN, GLU372→ALA or GLN372→ALA were adjusted within the FE regarding atom charge (Coulomb), bond, angle and dihedral (vdW). Masses were kept stable. Atoms without counter atom in the target structure were changed to uncharged dummy atoms with mass of the start atom. To avoid charge differences between start and end structure for the mutation from GLU→GLN or GLU→ALA the nearby HIS374 was changed in charge simultaneously. The ligand was changed by replacing the Glc2→-OH with the same strategy. Free energies were integrated by *g_bar* (Bennett acceptance ratio) of Gromacs/4.5.5 (Bennett, 1976; Pronk et al., 2013). The ΔΔG was derived by mutation of side chains with and without ligand and by transformation of the ligand in the binding site and in a water box.

2.7.5 Flexibility analysis

The RMSD of protein backbone atoms of residues 469-472 for the glucose binding groove lid in HK620 TSP and binding site residues were calculated with the *gmx rms* tool. The binding site residues were defined as all amino acids in a 5 Å distance of hexa- or octasaccharide ligand in HK620 TSP wild type (pdbID: 2VJJ) or Sf6 TSP EADA (pdbID: 4URR), respectively.

The tools *g_rmsf* and *g_cluster* were used to calculate the fluctuations of the carbohydrate ligands, the binding site residues and the glucose binding pocket for both proteins HK620 and Sf6 TSP (manual.gromacs.org/archive/4.6.6/online; 22.2.2017).

2.7.6 Water network assignment

Conserved water assignment by the End point method

The End point method was based on (Kang et al., 2016). The trajectory of 50 ns and 100 ns simulations were treated with the *trjconv –pbc nojump* implementation of Gromacs/4.5.5 (Pronk et al., 2013).

All water molecules crossing the simulation box barrier were treated without periodic boundary conditions. The positions of water molecules in 5 Å distance to the protein in the last frame of the simulation were averaged with the protein with *g_covar* (Hess et al., 2008). The resulting pdb-structure was again reduced by all water in 5 Å distance to the protein.

Conserved water assignment by the Hydrogen bond based method

The hydrogen bond analysis was performed using the Visual Molecular Dynamics (VMD) tool (Eargle et al., 2006; Humphrey et al., 1996). Hydrogen bonds were defined by a unique donor-acceptor pair of only polar atoms O, N, S and F as donor with 3.5 Å distance and an angle of 40°. The analysis was performed for the protein-water interaction with defining the binding site amino acids of HK620 TSP as donor or acceptor (G211, H212, Q242, F247, P252, L282, T307, W308, T311, G313, W314, N315, D339, H340, S341, Y344, N346, C369, E372, H374, Y393, V395, H397, A399, E400, I426, L427,G 428, S429, D430, D466, G468, P469, D470, N471, Q478 and R505) to all water molecules in the box. Water-water interactions were determined as water within 5 Å distance to the listed amino acids as donor or acceptor to all water molecules in the box. Similar conditions were used to analyze protein-glycan interactions and glycan-water interactions.

Hydrogen bonds with occupancies higher than 2 % were taken into account. Occupied water positions on a donor/acceptor atom position were averaged with a maximal distance of 3.7 Å by *compress.pl* (provided by F. Müller-Knapp). The positions were averaged with *cpptraj average* from AMBER14 (Roe and Cheatham, 2013) and placed on the averaged protein/protein-ligand structure derived by *g_covar* (Hess et al., 2008). Water positions were quantified by counting the different water molecule indices within a radius of 1.3 Å divided by the overall number of in total occupied frames leading to a quantification factor (Q-factor).

Mobile water network determination by *MobyWat*

Water position analysis was performed using the implementation provided from *MobyWat* (Jeszenői et al., 2016, 2015). Simulation trajectories of simulations were treated with *gmx trjconv* for preparation of the trajectories to handle periodic boundary condition effects. The system thereby is centered in the box and frames are fitted subsequently on top of the first frame (as suggested by *MobyWat* manual 1.0; mobywat.com; 16.01.2017). Simulations were analyzed in prediction mode with protein as target and all water molecules of the simulation as water. Clustering and prediction tolerance were set to 1.0 and 2.5 Å, respectively. The maximal distance was set by default at 5 Å.

3 Results

3.1 Affinity studies of Sf6 TSP as a carbohydrate binding protein

The structure of Sf6 TSP was first solved for the wild type protein (Sf6 TSP*wt*) as a complex with one tetramer repetative unit of *Shigella flexneri* Y O-antigen with the sequence [→2)-α-L-Rha*p*-(1→2)-α-L-Rha*p*-(1→3)-α-L-Rha*p*-(1→3)-β-D-Glc*p*NAc-(1→] (Müller et al., 2008). Nevertheless, octasaccharides (2 *RU*) were found as the main product of dodecasaccharide hydrolysis. The two acidic residues Glu366 and Asp399 were identified as part of the active site and thus in spatial proximity to the reducing end of the octasaccharide. Mutagenesis studies showed that related non-charged amino acids (Gln and Asn) at these positions led to a reduction of the hydrolytic activity k_{cat} of Sf6 TSP by 1000-100000 fold (Müller et al., 2008). As the two described mutations (E366Q and D399N) still exhibited residual activity, double mutants of both residues to the non-charged analogues were purified: E366Q D399N (Sf6 TSP_{EQDN}) as well as to alanine E366A D399A (Sf6 TSP_{EADA}). The Sf6 TSP_{EADA} mutant was successfully crystallized with octasaccharide (Kang et al., 2016) (Fig. 9).

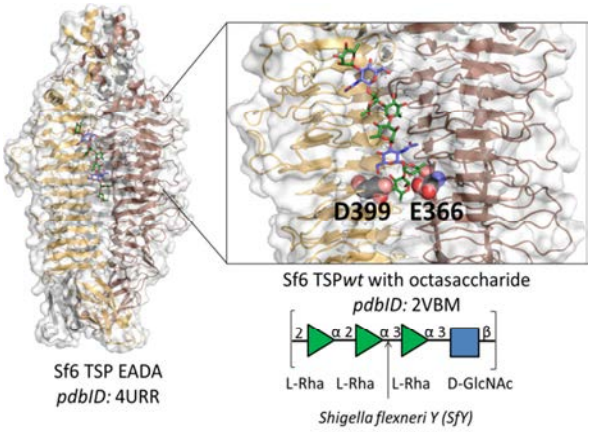

Fig. 9: Binding site of Sf6 TSP_{EADA} with octasaccharide ligand from SfY PS.
The trimeric β-helix structure of the Sf6 TSP was crystallized as inactive variant Sf6 TSP_{EADA} with two repetative units of SfY polysaccharide resulting in an octasaccharide ligand (pdbID: 4URR). The natural repetative unit of SfY O-antigen consists of three rhamnoses and one GlcNAc, which is cleaved at the Rha-(α1-3)-Rha glycosidic bond by the endorhamnosidase activity of Sf6 TSP*wt*. The enzymatic active residues E366 and D366 (spheres) are situated adjacent to the reducing end of an octasaccharide in the wild type structure (pdbID: 2VBM).

The dissociation constant of a Sf6 TSPwt-octasaccharide complex was estimated to lie in the milli- to micromolar range (Kang et al., 2016). Hence, measuring binding affinities with several methods in solution was challenging, due to the large amount of oligosaccharide ligand needed in these experiments. As a consequence, *Shigella flexneri* Y polysaccharide (SfY PS) was used in the present work for affinity experiments. Polysaccharide preparations were easier to obtain and promising to show an increased affinity based on their more elongated structure. Prerequisite for measurements with SfY PS was the use either of Sf6 TSP_{EADA} or Sf6 TSP_{EQDN} as inactive variants to minimize cleavage of the ligand during the interaction experiments.

3.1.1 Experimental characterization of the enzymatically inactive Sf6 TSP_{EADA}

3.1.1.1 Fast Purification of Sf6 TSP

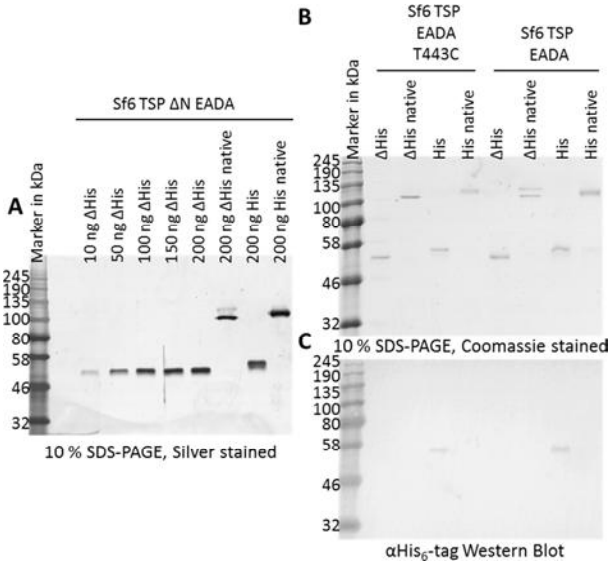

Fig. 10: Purification of Sf6 TSP with a His_6-tag
(A) The purity of Sf6 TSP_{EADA} was shown by a 10 % SDS-PAGE with silver staining. Protein bands at 56 kDa were assigned to the denatured monomer with a theoretical molecular mass of 58 kDa after incubation at 99 °C. Protein without heat treatment are native trimeric proteins with a theoretical molecular mass of 174 kDa detected as bands at 120 kDa. Samples with cleaved His_6-tag are labeled with ΔHis. **(B)** Protein samples of purified Sf6 TSP_{EADA} mutants were separated via 10 % SDS-PAGE with and without N-terminal His_6-tag. **(C)** Western blot analysis: TSP monomers with His_6-tag were detected at 56 kDa.

As all proteins in this work are without N-terminus, the ΔN will not be included further in the protein name. The purified Sf6 TSP is shortened by the N-terminal capsid domain resulting in Sf6 TSP ΔN. The protein was heterologously expressed in *E. coli* with an additional sequence of Met-Ala-Ser on the N-terminus followed by the amino acid sequence Lys^{108}-Ile^{623} as described previously (Freiberg et al., 2003). This included classical protein purification involving an ammonium sulfate precipitation and three chromatography steps, namely hydrophobic interaction, anion exchange and size exclusion chromatography. In order to simplify this purification protocol, an additional nucleotide sequence was cloned to the *sf6 tsp* gene, encoding six histidine residues (His_6-tag) and a linker region including a TEV-protease cleavage site sequence at the N-terminus. Expression of the modified *sf6 tsp* gene was possible after insertion into an expression vector. The respective protein was purified by an ion metal affinity chromatography. Subsequently, the N-terminal His_6-tag was removed by a TEV-protease digestion and a second affinity chromatography was done to separate the Sf6 TSP from the cleaved His_6-tags and the His_6-tagged protease. Due to the recognition site of the TEV protease, one additional glycine was added to the N-terminus, resulting in a changed amino acid sequence of Met-Gly-Ala-Ser-Lys^{108}-Ile^{623}. The new purification protocol yielded 25 mg Sf6 TSP from one liter of bacterial culture with a purity higher than 99 % as determined by SDS-PAGE with silver staining (Fig. 10A). His_6-tag removal was successful as no His_6-tag was detected by Western Blot analysis after the TEV-protease digestion (Fig. 10B-C). Thus the established protocol supplied pure protein of interest in a high yield and opened a route for the fast and easy purification of a variety of Sf6 TSP mutants.

3.1.1.2 Intrinsic protein fluorescence properties of Sf6 TSP mutants

The two inactive variants Sf6 TSP_{EADA} and Sf6 TSP_{EQDN} were tested in fluorescence titration binding experiments with polysaccharide ligands (Fig. 11A-B). For both proteins the maximum of intrinsic tryptophan and tyrosine fluorescence with excitation at 280 nm was detected at 339 nm. Upon addition of 33 µg/ml SfY PS an amplitude increase of 40 % for Sf6 TSP_{EADA} and of 6 % for Sf6 TSP_{EQDN} was observed. A hypsochromic shift of 3 nm occurred for both proteins, whereas the signal of Sf6 TSP_{EADA} was saturated by the addition of 33 µg/ml SfY PS. However, the signal did not reach saturation for Sf6 TSP_{EQDN}. The fluorescence increase and the blue shift indicate a more hydrophobic environment around the fluorescence emitting tryptophan residues in the binding site upon ligand binding. This results possibly from the removal of water molecules from the binding site that were replaced by the ligand.

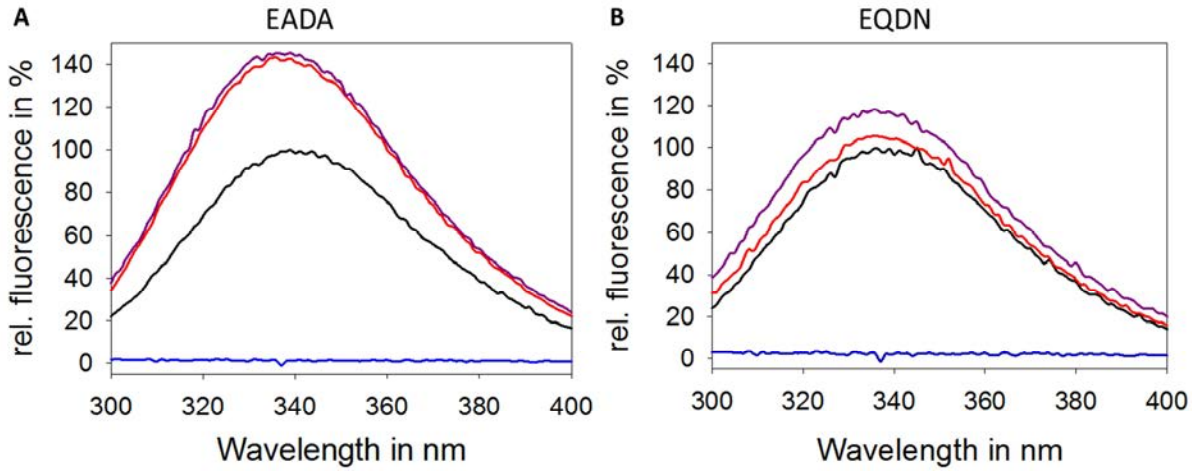

Fig. 11: Fluorescence binding experiments with the inactive variants of Sf6 TSP and SfY polysaccharide.
Intrinsic tryptophan fluorescence of Sf6 TSP_{EADA} **(A)** or Sf6 TSP_{EQDN} **(B)** was excited at 280 nm and the fluorescence emission was monitored between 300-400 nm. SfY PS was added sequentially which led to an increase of the intrinsic fluorescence of the protein without ligand (black) and with 33 µg/ml (red) and 66 µg/ml SfY PS (purple). The addition of polysaccharide without protein (blue) showed no signal.

The same intensity increase was observed upon addition of organic solvent in different concentrations to Sf6 TSP_{EADA} (Fig. S 1A-C). The protein solution was incubated for 3 h with methanol, acetonitrile and DMSO at 25 °C. A fluorescence increase up to 20 % was observed together with a 4 nm blue shift in these organic solvents. At organic solvent concentrations higher than 20 % fluorescence amplitudes decreased and the respective spectra red-shifted, suggesting a denaturation of Sf6 TSP under these conditions (Fig. S 1A-C). The stability of the protein tertiary structure of Sf6 TSP_{EADA} in 20 % organic solvent was confirmed by near-UV CD-spectroscopy (Fig. S 1D).

The differences in the intensity increase observed for both Sf6 TSP_{EADA} and Sf6 TSP_{EQDN} suggest that latter has a lower binding affinity than Sf6 TSP_{EADA}. However, the surrounding environment of the aromatic residues might be influenced differently in Sf6 TSP_{EQDN} compared to Sf6 TSP_{EADA} causing a smaller increase of the fluorescence signal during the binding event. The fluorescence signal in proteins derives from the three aromatic residues Phe, Tyr and Trp. However, the contribution of Phe can be seen as negligible due to the low quantum yield of the amino acid. The excitation maxima of Tyr and Trp are at 274 and 280 nm, respectively. Tyr fluorescence has an emission maximum at 305 nm, which is mostly quenched by an energy transfer to Trp or to an adjacent amide bond. Given this, Trp is considered as the main fluorescence source. The emission maximum of Trp can range from 308-350 nm depending on the polarity of the surrounding environment (Ghisaidoobe and Chung, 2014).

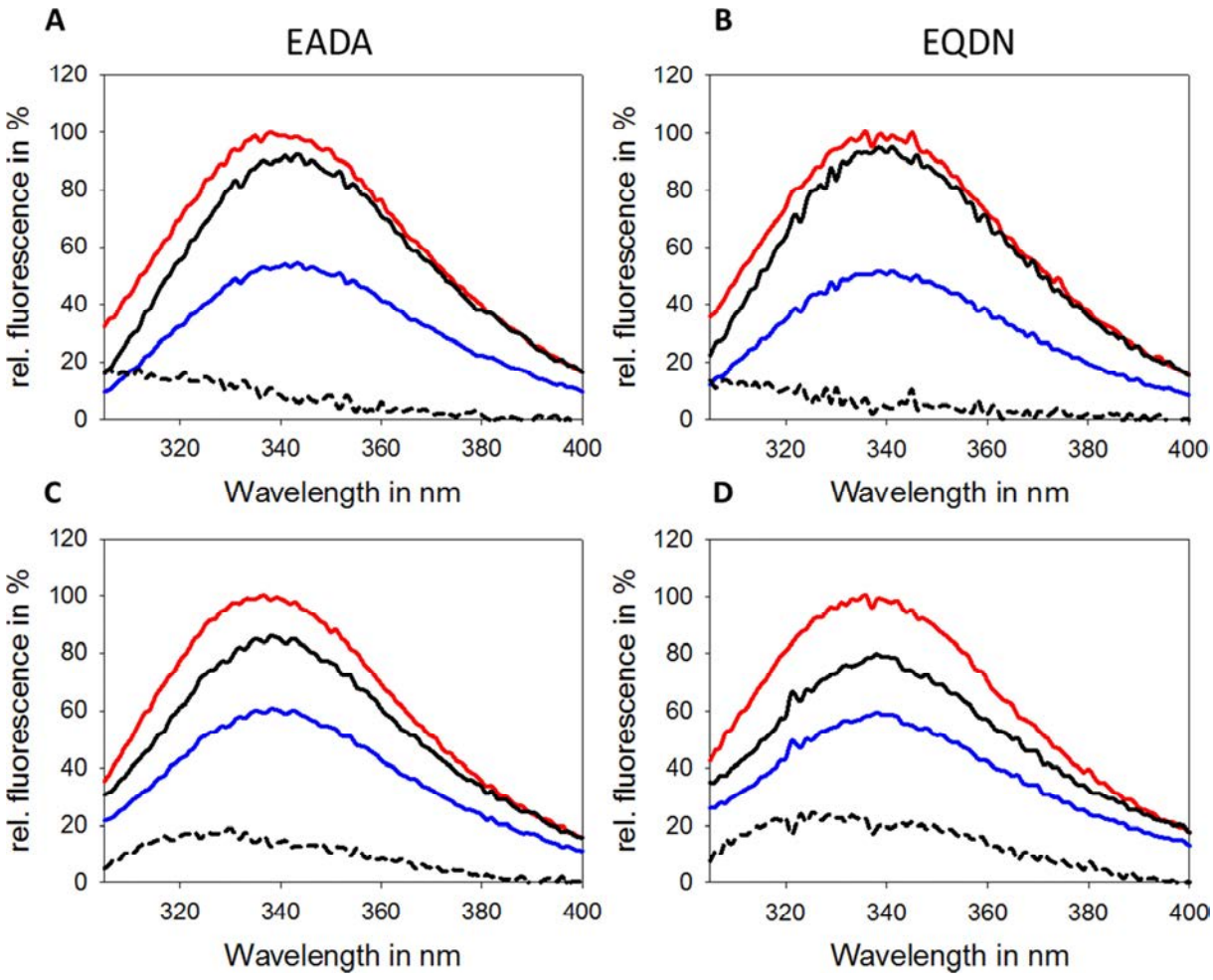

Fig. 12: Deconvolution of emission spectra from Sf6 TSP_{EADA} and Sf6 TSP_{EQDN} with excitation at 280 and 295 nm
Emission spectra were monitored from 305-400 nm with excitation at 280 (red) and 295 nm (blue). The 295 nm spectrum was normalized at 400 nm (black) to the 280 nm spectrum and the difference shows the contribution of Tyr (dashed line). The technique was applied on fluorescence data obtained for Sf6 TSP_{EADA} **(A)** and Sf6 TSP_{EQDN} **(B)** each without ligand and for Sf6 TSP_{EADA} **(C)** and Sf6 TSP_{EQDN} **(D)** with 66 µg/ml SfY PS.

The contributions of Tyr and Trp to a fluorescence spectrum can be determined by deconvolution of two obtained spectra with excitation at 280 and 295 nm. With the latter wavelength almost exclusively Trp residues are excited (Bobone et al., 2014). Deconvolution for Sf6 TSP_{EADA} spectra indicates that there is a minor tyrosine contribution of about 10 % to the spectrum at the maximum of 339 nm (Fig. 12A). Approximately the same distribution is observed for Sf6 TSP_{EQDN} with 95 % Trp and 5 % Tyr contributions to the signal, respectively (Fig. 12B). The maximum of the Tyr spectrum is at 305 nm. This maximum is shifted towards 325 nm with the addition of 66 µg/ml SfY PS for both mutants (Fig. 12C-D). Additionally, the spectrum contribution of Tyr increases proportionally with the ligand for both mutants. Sf6 TSP_{EADA} shows a partitioning of 85 % Trp and 15 % Tyr at the maximum at 336 nm, whereas the spectrum for Sf6 TSP_{EQDN} is divided into 78 % Trp and 22 % Tyr. The contribution of Tyr in the spectrum increases to a stronger extent upon the binding event for Sf6 TSP_{EQDN}.

Sf6 TSP is a trimeric protein with 17 Tyr and 5 Trp per monomer. Five Tyr residues are in 10 Å distance to the ligand and two Tyr adjacent to the elongated binding groove (Fig. 13A). Moreover, three Trp residues are found in 10 Å distance to the ligand (Fig. 13B). A comparison of binding site residue conformations in Sf6 TSP*wt* without ligand and Sf6 TSP_{EADA} with octasaccharide reveals only minor changes in side chain rotamers of binding site residues upon complex formation and mutation (Fig. 13F). Precisely, Asn340 on the middle left side of the binding site is shifted towards the ligand by 1.1 Å upon binding and Tyr400 at the lower left side by 0.7 Å. For Gln280, two rotamers were resolved without ligand, while Gln280 is exclusively found in one form with the amide group pointing towards RAM3 upon binding. These minor changes in binding site residue conformations do not explain the increase in the fluorescence in the oligosaccharide complex. However, the fluorescence blue shift suggests a removal of water molecules from the binding site resulting in an increased hydrophobicity around the fluorescent residues. The comparison of apo and ligand crystal structure shows, that most of the water molecules in the groove are removed from the binding site (Fig. 13C). Two Trp (304 and 421) and two Tyr (282 and 400) are primarily affected by removal of water molecules from their site. The two tyrosine residues exhibit hydrogen bonds to water molecules in the apo form, which are replaced by the ligand upon binding. A closer look on the surrounding amino acids of the tryptophan residues indicates, for example that for Trp304 (Fig. 13D), one Tyr, one His and two Gln are located in next proximity, which are potential quenchers of tryptophan fluorescence (Chen et al., 1996). Additionally, the negatively charged group of the enzymatically active Glu366 is in 7 Å

distance to the Trp304 indole ring. Negative charge is known to quench fluorescence of tryptophan, whereas positive charge in spatial proximity to the indole ring would increase the quantum yield (Callis, 2014). Neighboring residues with the same kind of properties are found for Trp421 with three Asp, one Asn, one His and the mentioned Glu366 (Fig. 13E). With its four charged residues, the surrounding environment of Trp421 is strongly negative, due to the carboxylate groups of Asp390/397, which are pointing towards the peptide bond of the tryptophan.

Previous studies with the single mutant Sf6 TSP_{DN} (D399N) did not show any fluorescence increase upon polysaccharide binding (Unplublished data by T. Scheidt). Contrary to this, the double mutants Sf6 TSP_{EADA} and Sf6 TSP_{EQDN} showed an increased fluorescence in the polysaccharide complex in the present work. It can be assumed that the negatively charged Glu366 quenches the fluorescence of Trp304 and 421 with the negative charge in 7 Å distance to the indole ring and only upon the change of Glu366 to Gln or Ala, the fluorescence of Trp is accessible to environmental changes. To confirm this hypothesis, additional fluorescence experiments with single point mutations (e.g. Sf6 TSP E366Q or E366A and Sf6 TSP D399A) as well as wild type with and without ligand would be required. In conclusion, complex formation of Sf6 TSP_{EADA} and Sf6 TSP_{EQDN} to SfY polysaccharide can be measured by intrinsic fluorescence titration experiments. In the following work the mutant Sf6 TSP_{EADA} was used as an inactive mutant for binding experiments.

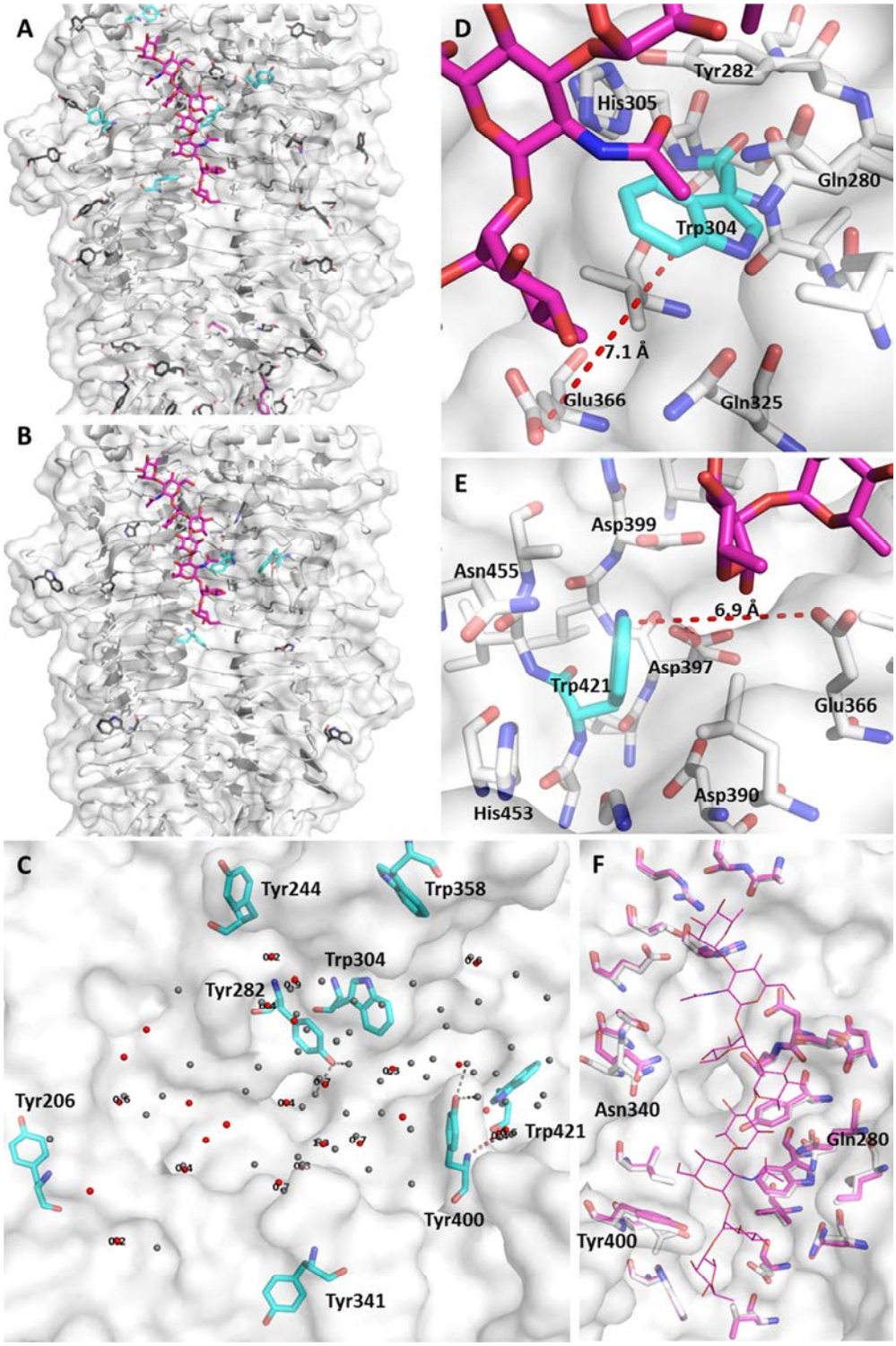

Fig. 13. Positions of fluorescent chromophores in the Sf6 TSP-oligosaccharide binding site
(A) Trimeric Sf6 TSP_{EADA} protein (pdbID: 4URR) with white backbone and white protein surface bound to the octasaccharide ligand (purple sticks) with Tyr residues as sticks (black). Tyr residues in 10 Å distance to the octasaccharide ligand (cyan) and adjacent to the elongated binding groove (pink) are colored differently. **(B)** Trp residues present in Sf6 TSP_{EADA} are depicted by the same color code as described in A. **(C)** The binding site of Sf6 TSP*wt* without ligand (pdbID: 2VBK) and Sf6 TSP_{EADA} with octasaccharide (white surface) with aromatic residues (cyan) within 10 Å of the octasaccharide ligand. Crystal water molecules within 5 Å to the octasaccharide from the apo structure (gray) build hydrogen bonds with two of the Tyr residues (gray dashes). Water molecules of the structure with octasacchairde (red) overlay with water molecules from the apo structure (Labeled numbers describe the distance between the overlaid water molecules with a threshold of 1.5 Å). **(D)** Trp304 and **(E)** Trp421 (cyan) in the binding pocket are surrounded by the depicted residues (white) and the octasaccharide ligand (pink). The carboxylate group of the enzymatic active residue E366 is in 7 Å (red dash) distance to Trp. **(F)** Binding site residues of Sf6 TSP*wt* without ligand (white) and Sf6 TSP_{EADA} with octasaccharide (pink) are superimposed in the protein binding site (white surface) including the ligand (pink lines). Residues that change their conformation upon the binding are labeled.

3.1.2 Structural characterization of the inactive mutant Sf6 TSP_{EADA} in complex with octasaccharide

The structure of Sf6 TSP_{EADA} with octasaccharide was already briefly analyzed in the evaluation of fluorescence contribution. In the following the binding contributions of the protein to the carbohydrate will be evaluated based on the influences known for carbohydrate protein interactions as hydrophobic CH,π-interactions, salt bridges, hydrogen bonds or water-mediated bridges (Davis and Wareham, 1999; Fernández-Alonso et al., 2005). Four Sf6 TSP structural models from X-ray crystallography are available in the protein data bank (PDB). Three of these are of the wild type protein with the PDB codes 2VBK, 2VBE and 2VBM (Müller et al., 2008). However, the 2VBE selenium methionine variant will not be taken into account in the present work. The model 2VBM includes one repetative unit of the SfY O-antigen, a tetrasaccharide.
In the structure of the inactive mutant Sf6 TSP_{EADA} (pdbID: 4URR) an octasaccharide ligand as natural cleavage product is bound (Kang et al., 2016). The structure 4URR was therefore chosen for comparison of experimental and simulated structures.

3.1.2.1 Contacts in the Sf6 TSP_{EADA}-octasaccharide binding site found in the crystal structure

The binding site of Sf6 TSP_{EADA} is situated between two subunits of the Sf6 TSP trimer. Twelve residues in chain A and eleven residues in chain B are in 5 Å distance to the octasaccharide ligand (Fig. 14A). These residues may form nine hydrogen bonds: seven in the upper site (*RU*1) and two in the lower site (*RU*2). Six potential nonpolar bonds are equally distributed over the binding site (Table S 2). In the Sf6 TSP_{EADA}-octasaccharide complex, 26 water molecules are found within a 5 Å distance of the ligand. Seven water molecules form hydrogen bond bridges between protein side chain residues and the octasaccharide. Two water molecules bridge between sugar hydroxyl groups (Fig. 14B, Table S 2). Twelve water-bridges are located in the upper binding site and only two in the lower binding site. In contrast, in Sf6 TSP*wt*-apo 63 water molecules are found in the binding site. These water molecules form an extended hydrogen bond network (Fig. 14D). Half of the water molecules are retained upon octasaccharide binding and half are replaced by the octasaccharide overlaying with the ligand structure (Fig. 14C, Table S 2). Water molecules are replaced more in the lower part (12 vs. 4), whereas retained water molecules are found more in the upper part (10 vs. 6).

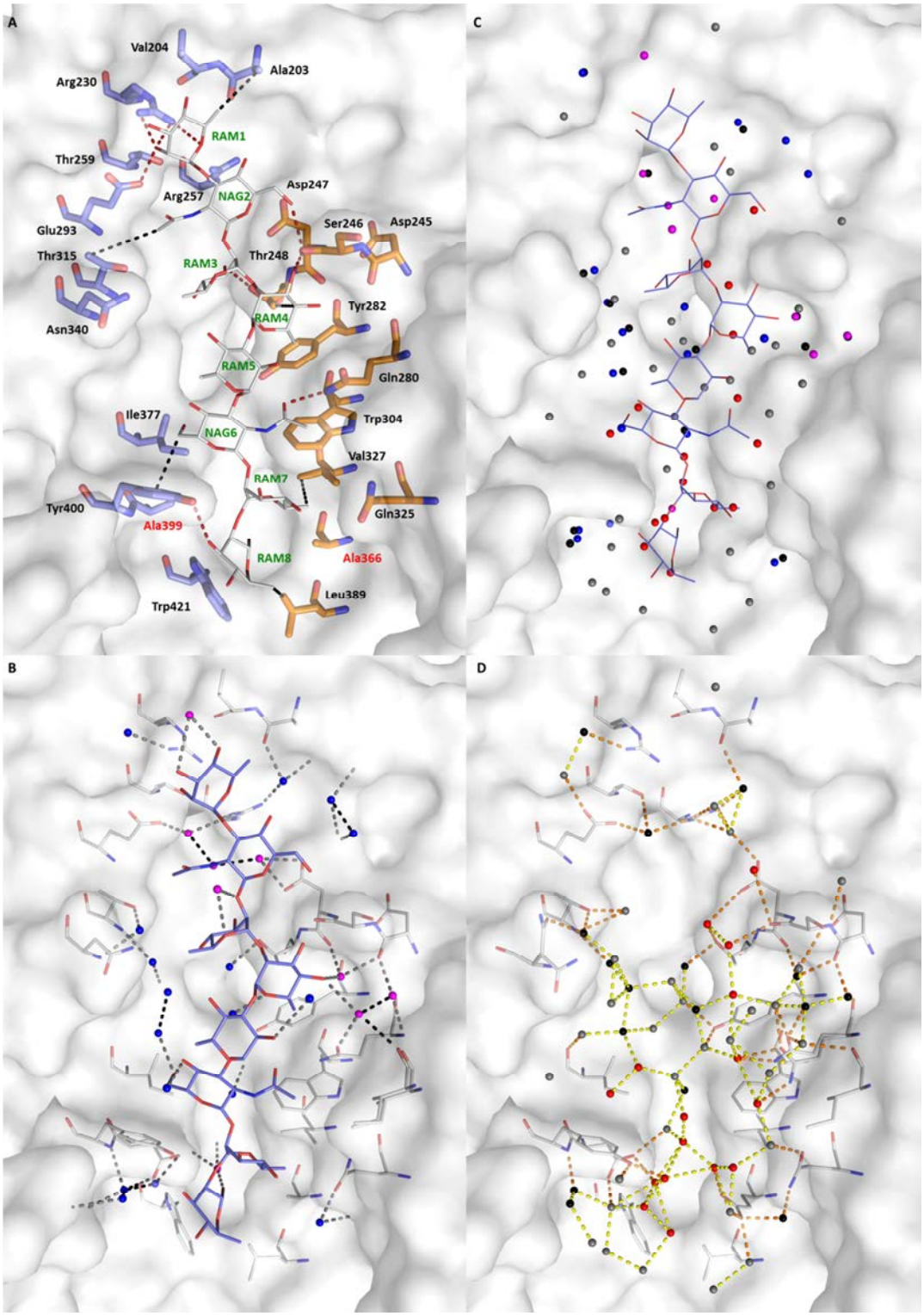

Fig. 14: Structural model of the Sf6 TSP_{EADA}-octasaccharide binding site (pdbID: 4URR) from X-ray crystallography
(A) Octasaccharide (white sticks) in the protein binding site (white surface) is surrounded with amino acids from chain A (blue sticks) and chain B (orange sticks). The amino acid residues are labeled in black with mutated residues in red. Carbohydrate moieties are labeled in green numbered from the non-reducing end of the carbohydrate (RAM for Rhamnose and NAG for N-acetylglucosamine). Hydrogen bonds (red dashes) and nonpolar interactions (black dashes) are indicated. **(B)** Structure 4URR with octasaccharide (blue lines): Water molecules in the binding site (blue spheres), water molecules bridging between ligand and protein (pink spheres), Hydrogen bonds bridging between ligand, protein and water (dashes). **(C)** Superimposed water molecules (spheres) from the apo (gray, black and red) and octasaccharide (blue and pink as in B) structure of Sf6 TSP. Water molecules from the apo structure which are in 1.5 Å distance to octasaccharide (blue lines) were considered as directly replaced by the ligand (red). Non-replaced water molecules within 1.5 Å of a corresponding water molecule (blue) in the ligand structure are shown in black. **(D)** Apo binding site with water molecules (same color code as C) with hydrogen bonds (dashes) between protein side chains (white lines) and water molecules (orange dashes) and between water molecules (yellow dashes).

3.1.2.2 Binding contacts of Sf6 TSP_{EADA} in MD simulation

So far the Sf6 TSP*wt* was simulated with octa- and dodecasaccharide ligand, even though a structural model as octasaccharide complex was only obtained for Sf6 TSP_{EADA} (Kang et al., 2016). Thus, in this work, Sf6 TSP_{EADA} with octa- and dodecasaccharide was simulated to evaluate the differences to the crystal structure and published simulations. Simulations were run with a rigid protein backbone for Sf6 TSP*wt*. However, side chain clashes were observed with the incorporation of a dodecasaccharide ligand into the binding site. Therefore, a twist on the glycosidic bond between the second and third repeating unit of the dodecasaccharide was introduced, and loops of the protein backbone were kept flexible (Kang et al., 2016). This set-up was used for Sf6 TSP_{EADA} simulations that were performed for 100 ns with TIP3P water in a box with periodic boundary conditions (Details in Chapter 2.7). The protein structure was stable throughout 100 ns simulations without detachment of the ligand (Fig. S 2).

In Sf6 TSP_{EADA} simulations, 20 hydrogen bonds for the octasaccharide and six additional hydrogen bonds for the third repeating unit in the dodecasaccharide were identified in total with occupancies higher than 10 %. For the Sf6 TSP*wt* simulations with rigid backbone and octasaccharide, 14 hydrogen bonds with a distance smaller than 5 Å were found. The third repeating unit in the dodecasaccharide was bound by 13 hydrogen bonds in Sf6 TSP*wt* with flexible loops (bonds without the two mutated residues E366 and D399 are counted only) (Kang et al., 2016). The hydrogen bonds Arg364-NH2, Asn455-OD1, Asn508-OD1 and Gln444-O observed in Sf6 TSP*wt*-dodecasaccharide simulations were also reproduced for Sf6 TSP_{EADA} simulations. Compared to Sf6 TSP*wt*, measured distances increased in simulations for Sf6 TSP_{EADA} except for the bond Tyr400OH-NAG6N2, which showed a higher distance for octa- and dodecasaccharide in Sf6 TSP*wt* (Fig. 15). The distances are particularly increased for the dodecasaccharide bound to Sf6 TSP_{EADA} between RAM3-8.

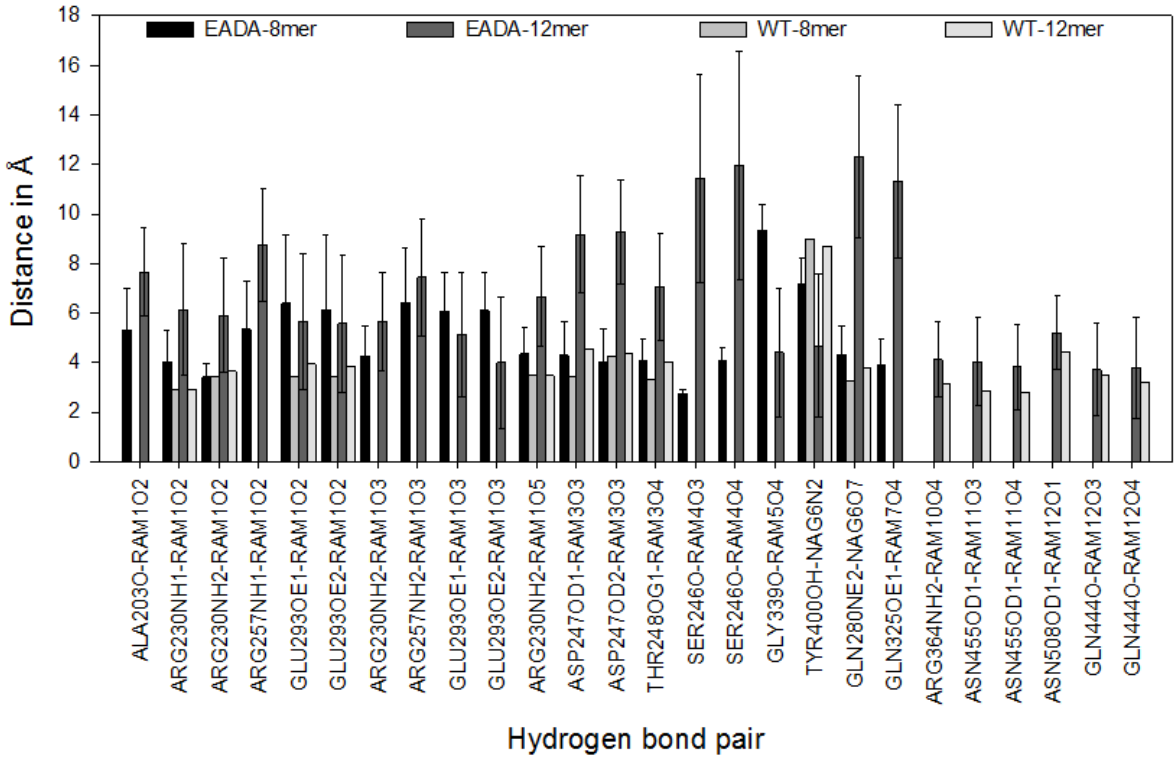

Fig. 15: Hydrogen bond distances of Sf6 TSP*wt* and Sf6 TSP_{EADA} complexes in MD simulations
Average distances with one standard deviation as error bars of hydrogen bond pairs in MD simulations from Sf6 TSP_{EADA} with octa- (8mer) and dodecasaccharide (12mer) compared to simulations of Sf6 TSP*wt* with the same ligands (Kang et al., 2016).

Apolar contacts between ligand and protein defined within a distance of 5 Å were compared between the experimental and MD simulation models of Sf6 TSP_{EADA} (Tab. 5). Five of the six apolar contacts defined by the crystal structure are also present in the simulation. The pair of Thr315-CG2 and NAG2-C8 from crystal structures was not found in the simulation as the measured average distance between these atoms was around 8 Å. The contact between Leu389-CD2 and RAM8-C6 is also increased up to 6 Å, but a large standard deviation of around 2 Å suggests that these contacts exist temporarily over the simulation time. None of the observed apolar contacts found for the first two repetative units in the crystal structure were reproduced over the course of the simulation for the dodecasaccharide ligand, as all measured distances increased to more than 5 Å with large standard deviations. Only in the starting structure of the simulation with the dodecasaccharide, seven of eight apolar contacts were found. The apolar contact analysis for the dodecasaccharide ligand suggests that there is a deviation from the starting position and a higher flexibility of the ligand in complex with Sf6 TSP_{EADA} in comparison to Sf6 TSP*wt*, similar to the hydrogen bond analysis.

Tab. 5: Apolar contacts in crystal structure and MD simulation of Sf6 TSP_{EADA}-octa- (8mer) and dodecasaccharide (12mer) complexes

Apolar contacts		Crystal (4URR)	EADA-8mer	EADA-12mer	EADA-12mer start[a]
Protein	**Ligand**	**Distance in Å**	**Distance in Å**	**Distance in Å**	**Distance in Å**
Ala203-CB	RAM1-C6	3.70	4.09±0.50	7.72±3.24	3.76
Thr315-CG2	NAG2-C8	4.80	6.95±1.09	6.84±2.71	4.44
Thr248-CG2	RAM4-ring(C4)	4.30	4.38±0.35	11.32±3.68	3.96
Tyr400-ring(CE2)	NAG6-C6	4.10	4.78±1.12	6.07±1.23	4.43
Val327-CG1	RAM7-C6	4.30	4.56±0.91	11.79±2.68	4.66
Leu389-CD2	RAM8-C6	4.10	6.21±2.25	9.20±1.81	7.54
Leu389-CD1	NAG10-C6	-	-	7.57±1.55	4.40
Tyr525-ring(CE1)	RAM12.C6	-	-	6.08±1.45	3.70

a Structure file after equilibration as the starting point of simulation of Sf6 TSP_{EADA}-12mer.

It was known from the Sf6 TSP_{EADA}-octasaccharide crystal structure that the flexibility of the ligand is increased towards the reducing end of the carbohydrate (Fig. 16B) (Kang et al., 2016). The binding state of the octasaccharide in the Sf6 TSP_{EADA} mutant in the simulation with flexible loops can be well reproduced (Fig. 16A-C). The fluctuation of the octasaccharide in the binding pocket ranges around a RMSD of 1.68±0.37 Å (Fig. 16A). The B-factor of every ligand atom is increased for the reducing end monosaccharide of the ligand (Fig. 16B). The formation of clusters from simulated ligand conformations resulted in one central structure of the octasaccharide, which nicely overlays with the crystal structure conformation (Fig. 16C) and simulations of Sf6 TSP*wt* (Kang et al., 2016).

In contrast, for the simulated dodecasaccharide in the binding site of Sf6 TSP_{EADA} the results from the wild type were not reproduced (Kang et al., 2016). The dodecasaccharide exhibited an increasing RMSD over time of the simulation with an average RMSD of 2.95±0.69 Å (Fig. 16A). The flexibility was particularly increased at the two ends (RAM1 and RAM12), while only *RU*2 showed similar fluctuations and displacements as found for *RU*2 during octasaccharide simulations (Fig. 16B). The dodecasaccharide is shifted with the first and second *RU* towards the left side of the binding groove and detachment events of *RU*1 can be observed within a 100 ns simulation. Furthermore, the third repeating unit is not linearly oriented and shows notable torsion angle deviations compared to the dodecasachharide pose used at the beginning of the simulation. The torsion angle of the L-Rha9-α(1-3)-D-GlcNAc10 glycosidic bond is increased from -45° in the starting structure to -121° in the central simulated structure (Fig. 16D). This twisted and shifted dodecasaccharide conformation seen in Sf6 TSP_{EADA} while keeping protein loops flexible in the simulation was observed for Sf6 TSP*wt* only in simulation with rigid loops (Kang et al., 2016). In contrast, simulations of Sf6 TSP*wt*-dodecasaccharide with flexible protein loops resulted in a linear conformation that superimposed well with the octasaccharide position from the crystal structure. Thus, the exchange of E366 and D399 from Sf6 TSP*wt* to alanine in Sf6 TSP_{EADA} influences the stability of the complete dodecasaccharide in simulation.

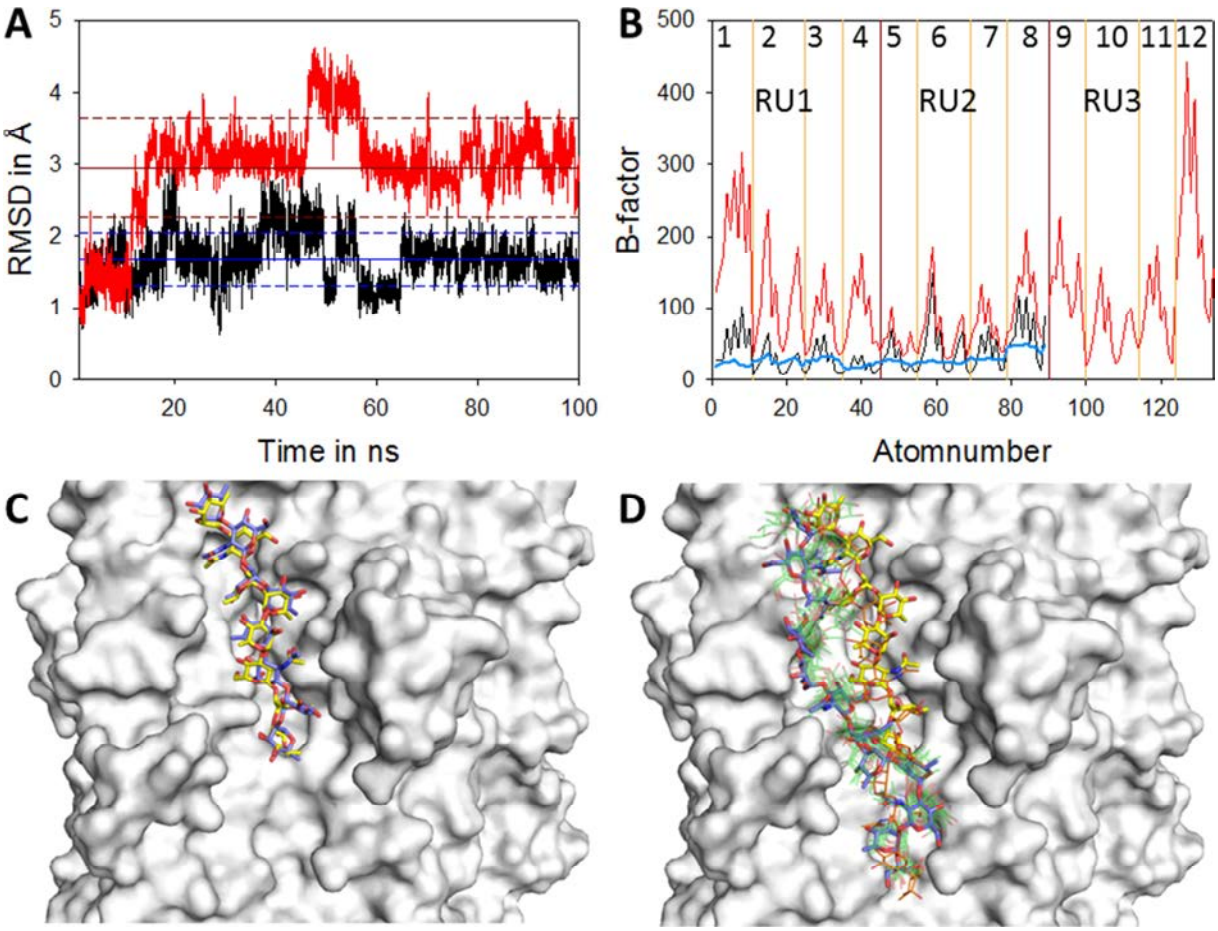

Fig. 16: Flexibility of octa- and dodedacasaccharide in the binding groove from MD simulations
(A) Trajectories of root mean square displacements (RMSD) of octa- (black) and dodecasaccharide atoms (red) in the binding site. The mean value is depicted as a solid line in blue (octasaccharide: 8mer) and dark red (dodecasaccharide: 12mer) with dashed lines in the corresponding colors for one standard deviation. **(B)** Mean B-factor of six octasaccharides in the Sf6 TSP_{EADA} crystal assymetric unit (blue) and calculated B-factor values from simulation of octa- (black) and dodecasaccharide (red). Numbers correspond to glycan monosaccharide building blocks. Yellow lines mark the glycosidic C1 as the start of a new glycan moiety and red lines indicate the end of one repeating unit. **(C)** The protein binding site (white surface) with the octasaccharide ligand from crystal structure (yellow sticks) and the central conformation of the octasaccharide (blue sticks, time frame: 91.63 ns) obtained by clustering the ligand conformations with a rmsdmin of 0.3 Å and a cutoff of 1 Å. **(D)** As in C, the dodecasaccharide with the central conformation in simulation (blue sticks, time frame 27.58 ns) superimposed with the octasaccharide ligand from the crystal structure (yellow sticks) and the dodecasaccharide starting structure (orange sticks). With the RMSD minimum and cutoff for clustering as in C, the dodecasaccharide resulted in nine clusters in 100 ns simulation (green transparent sticks).

In the crystal structures (2VBK vs. 4URR) three amino acid residues changed the conformation upon octasaccharide binding. The B-factors of binding site residues of Sf6 TSP_{EADA}-octasaccharide (X-ray) were increased for six residues in the elongated binding site (dodecasaccharide binding site) (Fig. 17A). Four of which were located within the region of *RU*3. In simulations with octasaccharide (Fig. 17B) flexibilities of protein side chains are increased when compared to those of the crystal structure. In total 16 residues were found with an increased flexibility, six of these are situated within the *RU*3 contact area. In contrast, only 14 flexible residues were found in simulations with the dodecasaccharide ligand (Fig. 17C) and a reduced flexibility for four residues in the *RU*3 contact area was observed.

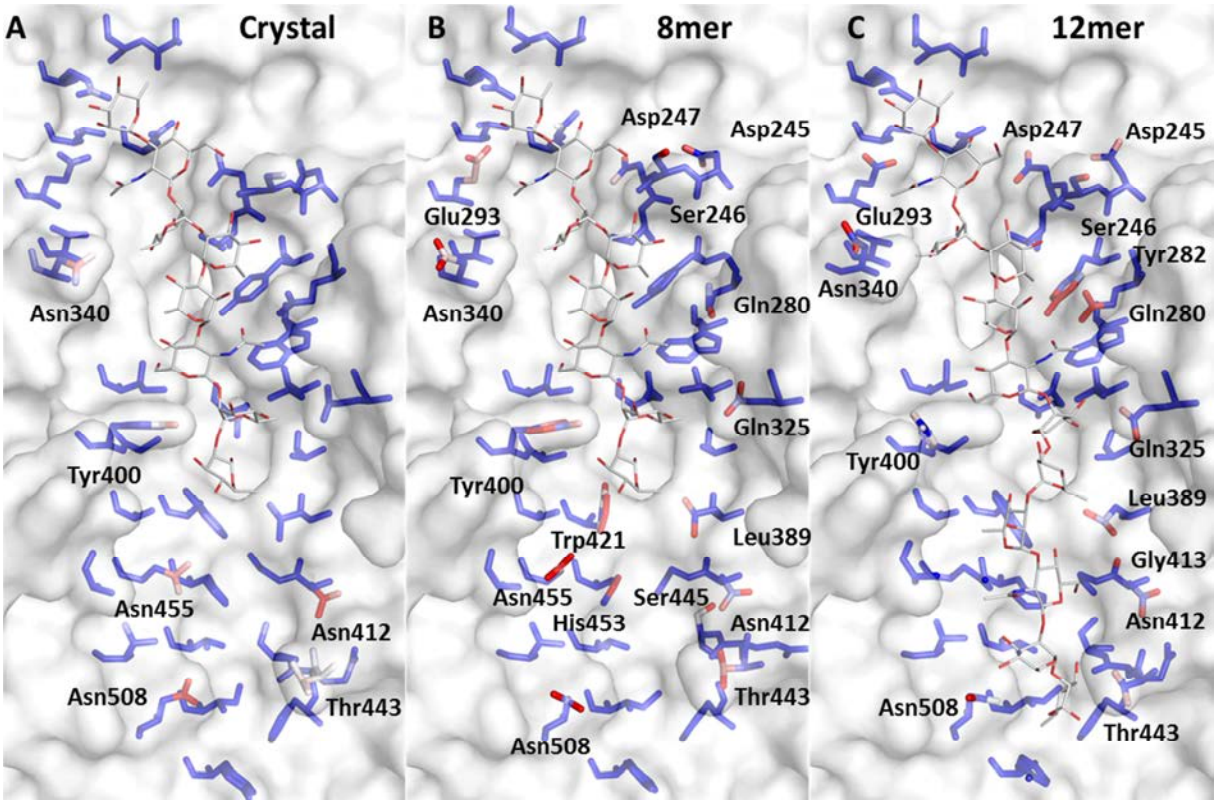

Fig. 17: Binding site residue B-factors from crystal structure and simulation
Binding site residues of Sf6 TSP_{EADA} are depicted from the octasaccharide crystal structure **(A)** and from simulation with octa- **(B)** and dodecasaccharide **(C)** colored by the B-factor of the atoms with blue for a low B-factor < 20 $Å^2$ and red for a high B-factor > 50 $Å^2$. The ligand is shown in gray thin sticks and residues with an increased B-factor are labeled.

For a comparison of the binding site residue conformations from simulation and crystal structure, an average of the simulated structures had to be defined. The tools for processing simulations provide the possibility to average given structures, but in case of Sf6 TSP, artefacts of side chain structures are produced by these tools as for example, tyrosine rings were reduced to coordinates of a straight bond (data not shown). Therefore, the tool for clustering structures *g_cluster* by Gromacs/4.6.4 was used to derive one complete cluster of the simulation. In this procedure, one simulation frame is defined as a central structure of the simulation with the smallest average RMSD to every other structure. The central structure for the two simulations of Sf6 TSP_{EADA} with octa- and dodecasaccharide were identified at 82.86 ns and 69.01 ns, respectively. A comparison of every binding site amino acid rotamer is given in Table S 3. In total, only one amino acid rotamer change occurs in the dodecasaccharide complex and none in the octasaccharide complex. In conclusion, the binding site amino acid side chain conformations and flexibilities do not vary strongly between the two different complexes, therefore no explanation can be found for the different binding modes obtained in the simulations of Sf6 TSP_{EADA} with octa- and dodecasaccharide based on the protein variations.

Water plays a crucial role as a binding partner in carbohydrate-protein interactions (Clarke et al., 2001). *MobyWat* was chosen as a tool for water position analysis in simulations (Jeszenői et al., 2015). Details of *MobyWat* usage will be given in more detail in Chapter 3.3. The *MobyWat* analysis obtained binding site water molecules from 100 ns simulations. To evaluate the best conditions for the *MobyWat* analysis, the trajectory was spliced in every, every tenth and every hundreds frame and the four lists IDa, IDe, MER and POS were derived by the tool. With every tenth frame, the best agreement rate of simulation and X-ray structure could be achieved with an IDa list top 20 % cut-off. These parameters were used for the water analysis of all performed simulations involving Sf6 TSP mutants.

For comparison of water positions, the structures were superimposed and 30 water molecules were found in the crystal structure that lie within 5 Å distance of the dodecasaccharide ligand (Fig. 18A). From these 30 positions, the octa- and dodecasaccharide simulations could reproduce 20 and 16 positions, respectively (Fig. 18A-B). In the simulations, a large number of water molecules were found in the binding site area, with 126 for octa- and 144 for dodecasaccharide. The hydrogen bond network between these water positions shows a high density of hydrogen bonds in the upper part (*RU*1 ligand area) and in the lower part (*RU*3 area) for the octasaccharide, whereas the middle part of the binding site (*RU*2 area) has a gap in the water network. In contrast, the water network assigned to the dodecasaccharide complex spans the entire binding site (Fig. 18D). Thus, there is a clear difference between octa- and dodecasaccharide simulations.

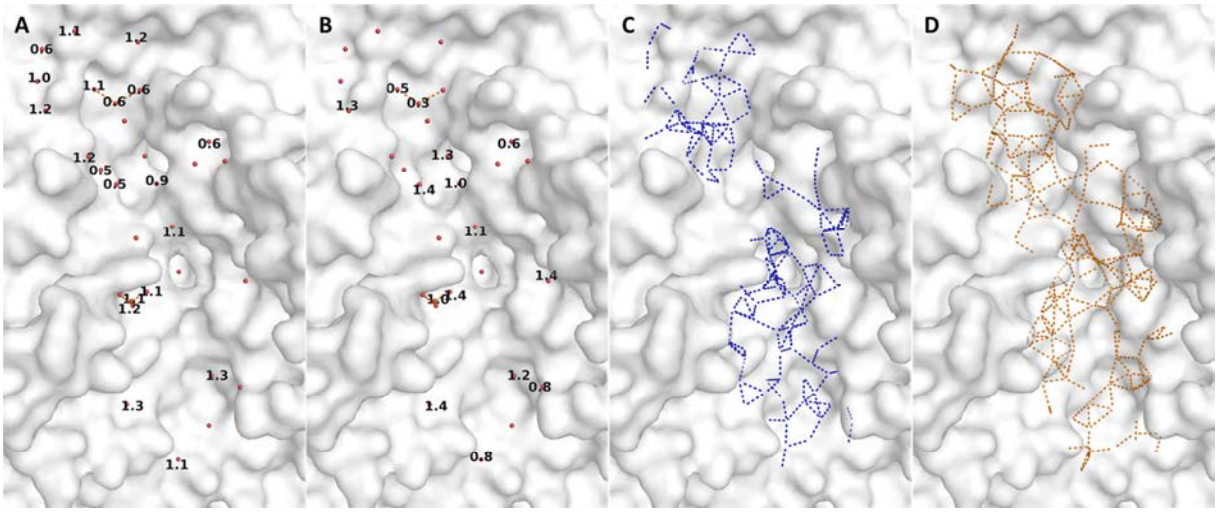

Fig. 18: Water molecule analysis of Sf6 TSP_{EADA} with octa- and dodecasaccharide
(A) Protein binding site (white surface) with crystallographic water molecules (red spheres) within 5 Å of the dodecasaccharide ligand starting structure. Potential hydrogen bonds are indicated as orange dashes. All labeled water molecules could be reproduced by simulation with the Sf6 TSP_{EADA}-octasaccharide with the indicated distances in Å. **(B)** As in A the water molecule position of the dodecasaccharide simulation which reproduced crystal water molecule positions with the corresponding labeled distances. **(C)** Water network of octasaccharide-complex simulation (blue dashes). **(D)** Water network of dodecasaccharide-complex simulation (orange dashes).

In summary, the simulated octasaccharide in Sf6 TSP_{EADA} can well reproduce the structure and binding contacts of the crystal structure and previous wild type simulation (Kang et al., 2016). Nevertheless, for simulations with the dodecasaccharide in Sf6 TSP_{EADA}, the ligand conformation found in simulations with wild type protein could not be reproduced: the dodecasaccharide has a twisted and shifted conformation to the left side of the binding cleft. Furthermore, the flexibility of the ligand is increased and detachment from the upper part of the binding site can be observed during 100 ns simulation.
In this section, the possible contributions in the binding event of Sf6 TSP_{EADA} were analyzed, namely hydrogen bonds, apolar contacts, ligand flexibility, binding site flexibility and water molecule distribution. Most of these analyses do not give a clear result as for example, apolar contact occupancies and distances or binding site residue flexibilities can be diverse, what makes a comparison complicated. Therefore in the following analyses of Sf6 TSP mutants, only hydrogen bond occupancies, ligand flexibility and water structure will be investigated.

3.1.3 Surface plasmon resonance to determine binding affinity

For the comparison of mutants which differ in their affinity to a certain ligand, a reasonable measurement set-up is necessary. The binding of Sf6 TSP to SfY polysaccharide could be measured by titration fluorescence spectroscopy experiments as described in section 3.1.1.2. However, polysaccharide is a heterogeneous mixture; the calculation of dissociation constants is not possible as the molecular mass of polysaccharide can only be estimated. Furthermore, the measurement of binding affinity by fluorescence titration experiment results in a dissociation constant in solution. Considering the natural situation of the specific binding event between bacteriophage tailspike protein and bacterial O-antigen polysaccharide, this interaction takes place at a surface with a respective dissociation constant. Therefore, the measurement of the interaction with a surface immobilized polysaccharide and the tailspike protein in the mobile phase would resemble more closely the situation during phage infection. Avidity and multivalence are known to increase the measured dissociation constant in surface plasmon resonance experiments by the factor of 150 fold compared to dissociation constants measured in solution (Lundquist and Toone, 2002). This amplification can be useful for the differentiation of only minor affinity changes, as they may occur from single point mutations that are probed in this work. SPR experiments with immobilized polysaccharide and tailspike protein in solution are connected with certain challenges, as the immobilization of polysaccharide for examples.

3.1.3.1 Immobilization techniques for bacterial lipopolysaccharide

For immobilization of lipopolysaccharide (LPS) or polysaccharide (PS) on a gold sensor chip several techniques are available, which are challenged by the heterogeneous properties of the LPS/PS structure and its non-stoichiometric modifications. The purified LPS and PS, which was used in this work, exhibits a bimodal distribution of O-antigen repeating units and the most LPS species showed 15 O-antigen repeating units (Fig. 19E). The calculated molecular mass of a *S. flexneri* LPS with 15 *RU* of O-antigen is 13.7 kDa. Without the core saccharide (1905 Da) the molecular mass of SfY PS results in 11.79 kDa, hence 12 kDa will be used as estimated molecular mass of SfY PS.

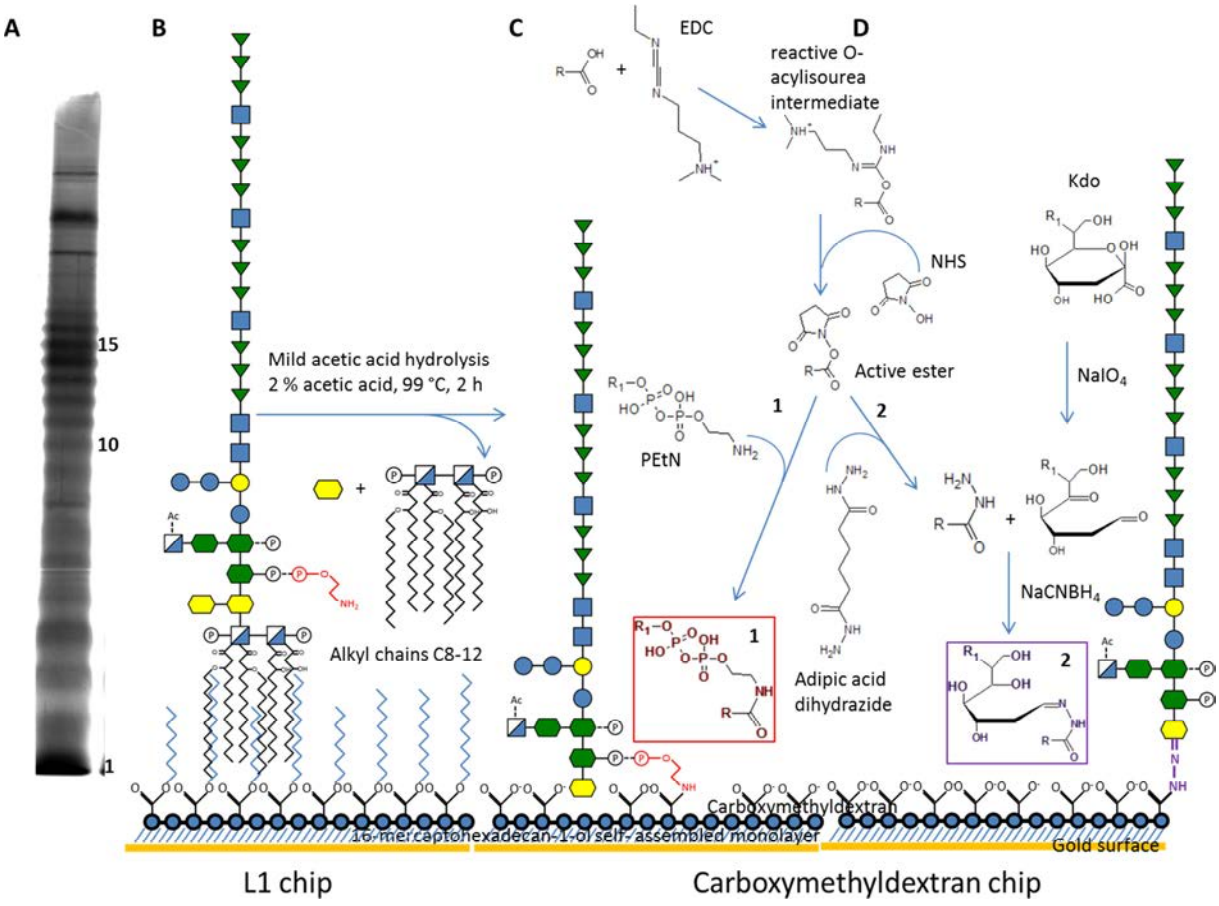

Fig. 19: LPS structure of *Shigella flexneri* Y and immobilization techniques for LPS and PS
(A) SfY LPS sample on silver stained 15 % SDS-PAGE showed the classical LPS ladder with a bimodal chain length distribution (The numbers indicate the number of repeating units). **(B)** Immobilization technique for LPS: Extruded LPS micelles built hydrophobic alkyl-chain interactions with a modified carboxymethyldextran chip decorated with alkyl chains of eight to twelve carbon atoms. **(C)** Amide coupling technique with potential non-stoichiometric phosphoethanolamine modifications of the inner core LPS **(1)** (Gedig, 2008). **(D)** Aldehyde coupling technique with EDC/NHS assisted hydrazide derivatization of carboxymethyldextran with adipic acid dihydrazide. The periodate oxidized reducing end glycan (Kdo) is coupled to the hydrazide and stabilized by sodium cyano borhydrate reduction **(2)** (Gedig, 2008).

For the immobilization of LPS or PS on a sensor chip three strategies were used in the present work: hydrophobic interaction of alkyl chains with the acyl chains of lipid A; amide coupling on a carboxymethyldextran surface via phosphoethanolamine residues of the core saccharide of PS and aldehyde coupling on the same surface via oxidized reducing ends of PS (Fig. 19B-D). The Pioneer® L1 chip from BIAcore is a modification of the carboxymethyldextran surface (Löfas and Johnsson, 1990) with attached alkyl chains with eight to twelve carbon atoms (Remmel et al., 2007). On a L1 chip vesicles can be captured intact (Cooper et al., 2000) or as continuous bilayer (Erb et al., 2000), which can also be lipid dependent (Cooper, 2004; Hodnik and Anderluh, 2010). For an *E. coli* LPS a critical micelle concentration of 14 µg/ml and micelles with 190 nm in diameter above this concentration were reported (Santos et al., 2003). Nevertheless LPS is also known to form multilamellar aggregates (D'Errico et al., 2010). The use of LPS for immobilization on a L1 chip leads to the formation of stable LPS surfaces, which could be used for interaction experiments between Sf6 TSP and SfY LPS before (Kang et al., 2016). The disadvantage of this immobilization variant is that only a limited number of interaction experiments can be performed, as the surface is unstable due to non-covalent attachment of the LPS solute (Fig. 20A).

In the amide coupling strategy a carboxymethyldextran surface without any further modifications is used. The negatively charge carboxylates are activated by EDC (1-ethyl-3-(3-dimethylaminopropyl)carbodiimide), which forms an unstable O-acylisourea intermediate that is further modified to react with NHS (N-hydroxysuccinimide) to form an NHS activated ester, which then can react with primary amino groups. The stability and lifetime of this activated ester is pH dependent for several seconds at pH 8-9 and hours at pH 4-5 (Gedig, 2008) (Fig. 19C). For the immobilization of PS the lipid A moiety of LPS can be removed by mild acetic acid hydrolysis, and the remaining polysaccharide is expected to have a undefined amount of non-stoichiometric phosphoethanolamine modification, which would serve as primary amino groups. All attempts to use this immobilization technique with SfY PS were not successful (data not shown), which suggests that not enough phosphoethanolamine modifications are present on the given polysaccharide preparation. However, fresh preparations of *Salmonella typhimurium* polysaccharide were successfully immobilized for interaction experiments with the corresponding O-antigen specific P22 TSP (Fig. 20B).

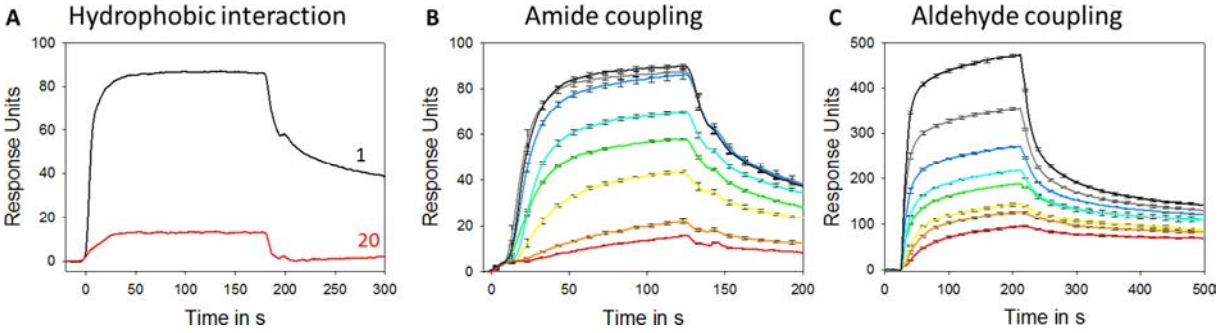

Fig. 20: SPR sensograms of TSP binding to LPS or PS surfaces
(A) SfY LPS immobilized as micelles on a L1 chip. The black curve indicates the first cycle of 100 µg/ml Sf6 TSP_{EADA} over the surface and the red curve the twentieth cycle (same surface, same concentration). **(B)** *Salmonella typhimurium* polysaccharide was immobilized via amide coupling to a carboxymethyldextran surface and interaction experiments could be performed with P22 TSP, the natural receptor of *Salmonella* polysaccharide. P22 TSP was injected in a dilution series between 8-1600 nM TSP subunits (red to black). The maximal response unit signal (RU_{max}) at 118 s was taken as the equilibrium signal for the interaction and a 1:1 binding isotherm resulted in a K_D of 80 nM. **(C)** Oxidized SfY PS immobilized on a hydrazide-activated carboxymethyldextran surface was used for interaction experiments of Sf6 TSP_{EADA} with a serial dilution of 0.08-18 µM TSP(monomer) (red to black). Every interaction curve was monitored in triplicates for B and C and the average is depicted. Error bars indicate standard deviation at every tenth value.

The coupling of aldehyde groups represents the third immobilization technique, and is based on the same basic procedure used in the amide coupling. The carboxylates on the surface are activated with EDC/NHS to active ester groups, which are modified with adipic acid dihydrazide to reactive hydrazide groups on the surface. Via these hydrazide groups, aldehyde of the reducing end of carbohydrates can be coupled via Michael addition and subsequent stabilization by reduction with sodium cyanoborhydride (Gedig, 2008) (Fig. 19D). The immobilization of unmodified SfY PS with the aldehyde coupling strategy was not successful (data not shown). Therefore, a mild oxidation of the sugar residues with sodium periodate was necessary (Gedig, 2008).

The oxidation of carbohydrates with sodium periodate is highly dependent on the chosen conditions. With higher periodate concentrations and temperatures, cleavage of C-C bonds becomes possible adjacent to hydroxyl groups (Hermanson, 2013). At approximately 8fold concentrations of periodate to glucose for example, the carbohydrate would be oxidized to formaldehyde (Avigad, 1983). SfY polysaccharide (10 mg/ml) was oxidized with 10 mM sodium periodate at slightly acidic pH for 20-30 min at 4 °C and 25 °C, respectively. The reaction was stopped with an excess of ethylene glycol and the resulting oxidized polysaccharide showed a five-fold higher amount of reducing groups compared to the control. In total ~1500 RU could be immobilized to a hydrazide-activated carboxymethyldextrane surface for Sf6 TSP interaction experiments (Fig. 20C). The polysaccharide surface was highly stable and more than 300 experiments were performed on one surface. So far, the ligand density was not attempted to be regulated on the surface. Moreover, modifications of the polysaccharide structure due to oxidation cannot be ruled out.

3.1.3.2 Analysis of TSP-polysaccharide binding affinities in SPR at equilibrium

The immobilization of SfY PS was repeated with two oxidized samples, as described in the previous section, which resulted in Chip1 and Chip2. On both of these chips, interaction experiments at different concentrations of Sf6 TSP_{EADA} were repeated twice (Measurement1-4). For the determination of dissociation constant K_D from interaction curves as shown in Fig. 20C, the reaction should reach equilibrium, which would be apparent in the sensogram by a plateau area at the end of the injection. However, no horizontal plateaus were observed in the sensograms. This continuous signal increase may occur from mass transport effects (Schuck and Minton, 1996) or unspecific binding (Copeland, 2000). Mass transport effects can be ruled out when the flow rate is increased. Furthermore, injection time can also be elongated, as the interaction might not have reached the equilibrium yet. The comparison of a series of injection times (Fig. 21A) between 1-10 min showed the same binding curve shape independent of the changed flow rate. Additionally, no equilibrium was reached within 10 min of injection time. However, the sensogram curves display a plateau with the same slightly positive slope during the injection for every concentration (Fig. 20C). Therefore, an unspecific process must influence the curve progression to the same extent for every concentration.

The equilibrium responses, which were measured at 178 s of injection, can be described by binding isotherms (Fig. 21B). A 1:1 binding isotherm fit returned a dissociation constant of 1.05 µM, but the quality of the fit only resulted in R^2=0.81. The quality of the fit was improved up to an R^2=0.99 by using a bivalent binding isotherm with two resulting dissociation constants of K_{D1}=0.086 µM and K_{D2}=25.42 µM. As already described, the concentration series of Sf6 TSP_{EADA} was repeated four times on two SfY PS coated sensor chips. The maximal responses of these interactions were not reproducible (Fig. 21D), as the first measurements on each chip resulted in 400 response units higher than the second. The dissociation constants of the four measurements fitted by a 1:2 binding isotherm vary in the micromolar range (Tab. 6). As Sf6 TSP is a homotrimeric protein, a multivalent interaction can be assumed, which would justify the use of a multivalent binding isotherm. However, unspecific processes also can be involved. The Scatchard analysis of equilibrium responses resulted in a curved scatchard plot speaking for multivalence or unspecificity (Fig. 21C).

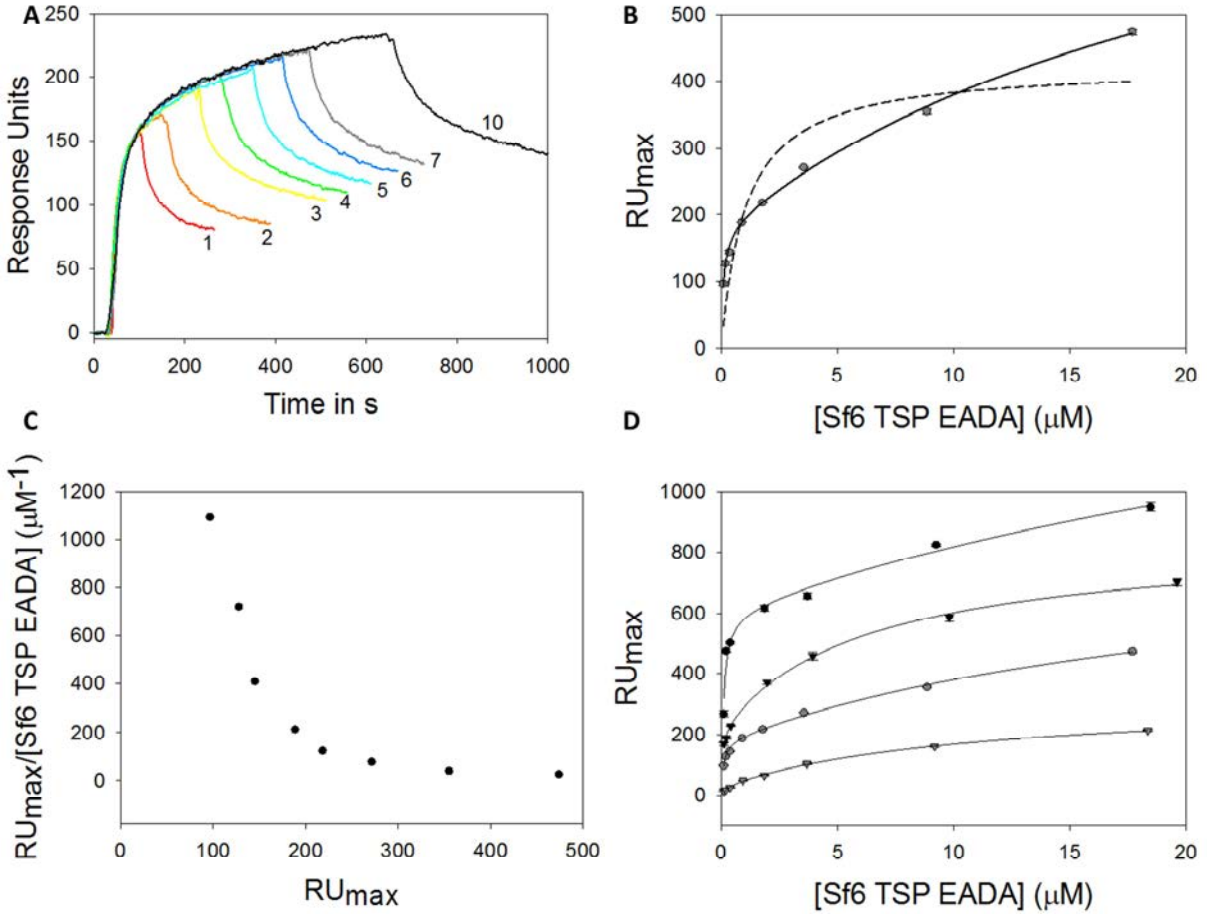

Fig. 21: Interaction analysis of Sf6 TSP_{EADA} on immobilized SfY PS with surface plasmon resonance at equilibrium
(A) Sensograms of 0.8 µM Sf6 TSP_{EADA} injected on immobilized SfY PS on carboxymethyldextran at different interaction times and flow rates: 1 min→50 µl/min; 2 min→35 µl/min; 3 min→20 µl/min; 4 min→17 µl/min; 5 min→15 µl/min; 6 min→11 µl/min; 7 min→10 µl/min and 10 min→7 µl/min. **(B)** Maximal response (RU_{max}) of the curves presented in Fig. 20C (Chip1, Measurement2) at different protein concentrations (gray dots) shown as mean of three measurements with error bars as one standard deviation. The maxima are fitted with 1:1 (dashed line) and 1:2 Langmuir binding isotherms (solid line). **(C)** Scatchard Plot analysis of the data from B resulted in a curved, triphasic distribution. **(D)** Maximal responses as mean of triplicate measurements with standard deviations as error bars against the protein concentration from the four measurements of Sf6 TSP_{EADA} on two different SfY PS surface chips with Chip1 (circles) and Chip2 (triangles): Measurement1 (black circles), 2 (gray circles), 3 (black triangles) and 4 (gray triangles).

The measured affinities on Chip2 are more diverse than the affinities on Chip1. On Chip1 and Chip2, 1514 and 1495 RU of SfY PS were immobilized, respectively. The preparation differed in the oxidation of the SfY PS for 30 min at 25 °C for the first preparation and 20 min at 4 °C for the second preparation. In the first preparation the polysaccharide chains are possibly more cross-linked due to a longer oxidation time (Maia et al., 2011). Between each measurement, the chip surfaces were treated with in total 48 cycles of Sf6 TSP mutants; some of which might have residual enzymatic activity (data not shown) and cleave the polysaccharide chains on the surface. This would explain a decrease of the overall maximal response, and an increase of the dissociation constant.

Tab. 6: Dissociation constants for Sf6 TSP_{EADA} derived from Measurements1-4.

Chip	Measurement	Binding model	Maximal response (RU_{max})	Dissociation constant (µM)	Fit quality (R^2)
1	**2**	1:1	RU_{max}=423±52	K_D=1.05±0.48	0.8128
1	**1**	1:2	RU_{max1}=603±79 RU_{max2}=1252±2491	K_{D1}=0.085±0.037 K_{D2}=47.09±100	0.9782
1	**2**	1:2	RU_{max1}=180±11 RU_{max2}=713±145	K_{D1}=0.086±0.019 K_{D2}=25.42±9.41	0.9980
2	**3**	1:2	RU_{max1}=211±28 RU_{max2}=632±30	K_{D1}=0.037±0.023 K_{D2}=6.15±1.33	0.9981
2	**4**	1:2	RU_{max1}=32±11 RU_{max2}=298±18	K_{D1}=0.273±0.181 K_{D2}=11.75±2.87	0.9989

In conclusion, the binding behavior of Sf6 TSP_{EADA} was determined as a bivalent interaction from equilibrium data with two dissociation constants of ~10^{-8} M and ~10^{-5} M on carboxymethyldextran-immobilized oxidized SfY polysaccharide.

3.1.3.3 TSP-polysaccharide binding affinities from kinetic analyses of SPR-sensograms

In the last section, SPR response curves were evaluated based on their equilibrium responses. The same data can also be adjusted to a kinetic model. The simplest model would be a 1:1 binding model, which leads to the Langmuir binding isotherm in equilibrium. As seen in the last section, the Langmuir binding isotherm did not result in an acceptable fit of the experimental data, therefore it can be concluded that additional parameters should be included in the kinetic model. Interaction experiments using surface plasmon resonance are affected by mass transport effects, as the diffusion of the analyte into the carboxymethyldextran layer is limited. The mass transport effect would result in reduced association and dissociation rates. On the one hand, the diffusion of the analyte from the bulk to the surface becomes rate limiting, which influences the association rate. On the other hand, the dissociation rate is reduced due to rebinding of the analyte (Schuck and Minton, 1996). Nevertheless in a transport limited process the equilibrium data would also be described by a 1:1 Langmuir binding isotherm (de Mol and Fischer, 2008). Thus, transport limitation cannot be the only effect influencing Sf6 TSP-PS binding. The equilibrium data could be described by a bivalent model, as the trimeric tailspike protein can bind to several ligands on the surface. The bivalent kinetic model is predominantly used for antibody binding on surfaces (Cooper and Williams, 1999). For TSP, the bivalent binding model resulted in a global fit with a poor quality quantified by Chi^2 of ~900 for Measurement2 (Fig. 22A). Therefore a bivalent binding model did not describe the data. The polysaccharide ligand was immobilized as a heterogeneous mixture on the surface, thus the multivalent binding behavior of the tailspike protein could also originate from a heterogeneous PS distribution on the surface (de Mol and Fischer, 2008;

Khalifa et al., 2001). A heterogeneous ligand parallel binding model (1:2 model) resulted in a Chi^2 of ~220 for Measurement2, which is an improved description compared to the bivalent model (Fig. 22B). The description of Measurements1, 3 and 4 within the respective model can be found in Fig. S 3 together with a description of the used models.

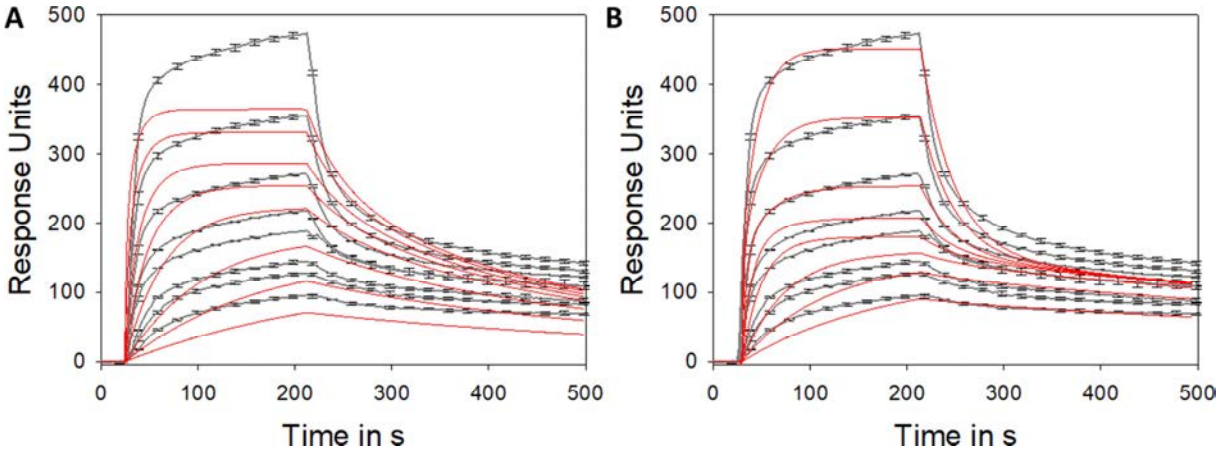

Fig. 22: Concentration series Measurement2 of Sf6 TSP_{EADA} binding to carboxymethyldextran immobilized SfY PS
The sensograms of Sf6 TSP concentrations between 0.088-18 µM (gray) as mean values of triplicate measurements (error bars show standard deviations from every 10^{th} point) were fitted (red) with a bivalent **(A)** and 1:2 binding model **(B)**.

K_D derived from the four measurements on the two chip surfaces differ by 10-20 % within the error between kinetics and equilibrium calculated dissociation constants (Tab. 7). In general, the agreement of equilibrium and kinetics evaluation is better for Chip2 compared to Chip1.

Tab. 7: Dissociation constants for Sf6 TSP_{EADA} derived from a 1:2 heterogeneous ligand binding kinetics fit

	Kinetics[a]					Equilibrium[a]		
#	$RU_{max\,1/2}$ (RU)	$k_{ass\,1/2}$ (10^3M s)	$k_{diss\,1/2}$ (10^{-3}/s)	$K_{D\,1/2}$ (µM)	Chi^2	$RU_{max\,1/2}$ (RU)	$K_{D\,1/2}$ (µM)	R^2
1	302.9±1.12	154.0±73.8	1.50±0.015	0.0097±0.0062	2155	603±79	0.085±0.037	0.9782
	712.7±1.74	5.06±5.79	24.7±0.018	4.88±18		1252±2491	47.09±100	
2	155.6±0.21	58.50±3.23	1.11±0.018	0.0190±0.0014	222.3	180±11	0.086±0.019	0.9980
	594.1±0.55	1.67±1.53	29.7±0.004	17.8±100		713±145	25.42±9.41	
3	176.0±0.51	157.0±13.6	3.09±0.004	0.0197±0.0018	381.5	211±28	0.037±0.023	0.9981
	630.0±0.32	5.07±1.59	27.3±0.004	5.38±1.87		632±30	6.15±1.33	
4	36.27±0.38	16.70±157	5.02±0.003	0.3010±0.0324	20.48	32±11	0.273±0.181	0.9989
	306.7±0.05	3.42±0.48	49.2±0.001	14.4±2.04		298±18	11.75±2.87	

a In each row: First line = RU_{max1}, k_{ass1}, k_{diss1}, K_{D1}; Second line = RU_{max2}, k_{ass2}, k_{diss2}, K_{D2}

As a comparison to the kinetic parameters obtained from SPR experiments, binding kinetics were observed in solution by tracing protein tryptophan emission increase at 336 nm (see section 3.1.1.2). The apparent rate constants derived from biexponential fits of the traces resulted in two dissociation constants: $k_{diss,1}$=0.0335 and $k_{diss,2}$=0.0066 assuming a pseudo-first order binding behavior. The second rate is considered to be unspecific due to manual mixing in the experimental set-up (Fig. 23A-B). The dissociation rate from fluorescence kinetics in an interaction experiment in solution agrees well with the dissociation rate derived

from SPR experiments depicting interactions on a surface (k_{diss2}≈0.024-0.049/s) even though the interaction experiments in solution were performed at lower temperatures. Dissociation rates in SPR experiments can be influenced by mass transport limitation and bivalent binding, which more likely occurs on a surface than in solution due to limited reaction volume.

In Fig. 23B, the apparent rate constants were fitted as a pseudo-first order relaxation process with excess ligand. The data show a clear deviation from this simple model. The apparent rate constants obtained from the fluorescence binding relaxation traces can also be used to evaluate the binding mechanism of the Sf6 TSP-PS interaction (Fig. 23C). More precisely, the binding mechanism can either follow the conformational selection model, where a conformational change occurs prior to the binding event or the induced fit model, where the conformational change occurs after binding (Paul and Weikl, 2016). Although data points at low concentrations are sparse, it may be speculated from the more sophisticated fit that Sf6 TSP-PS binding in solution is more accurately described by the conformational selection model (R^2=0.853) than by the induced fit model (R^2=0.775). Derived rate constants and the models are illustrated in Fig. S 4.

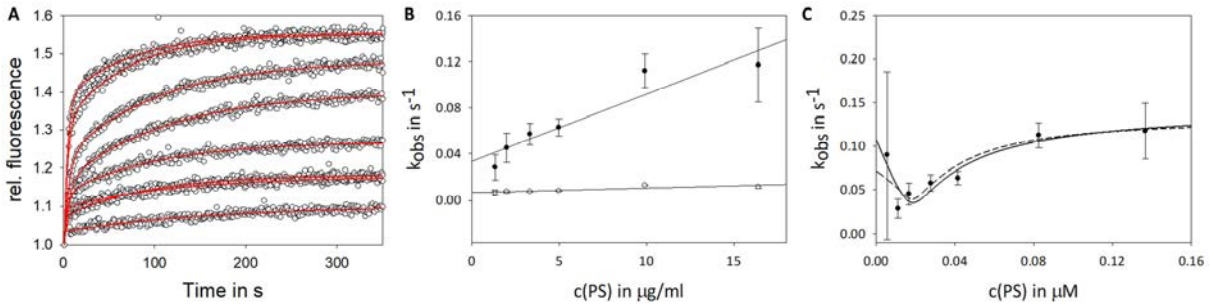

Fig. 23: Relaxation rates of Sf6 TSP_{EADA} polysaccharide complex formation analyzed by intrinsic protein fluorescence
(A) Fluorescence traces of 184 nM Sf6 TSP_{EADA} excited at 280 nm and monitored at the emission maximum of 336 nm. Traces (black circles) were monitored for SfY polysaccharide concentrations from 0-16.4 µg/ml and were fitted with a biexponential function (red). **(B)** From the biexponential fit, apparent rate constants k_{obs} were calculated and are represented at the corresponding ligand concentrations. Fits with a pseudo-first order reaction describe a linear relationship as $k_{obs} = k_{diss} + k_{ass}$ ([P] + [L]), resulting in the y-intercept as k_{diss}. The error bars represent the standard error on the given value from the biexponential fit. The linear regression was only performed for ligand concentrations between 1.33-16.4 µg/ml excluding the apparent rate constant for the lowest concentration due to the high standard error of this value. **(C)** All apparent rate constants (including the lowest concentration) are plotted against SfY PS in molar concentrations calculated with an estimated average molecular mass of 12 kDa for a PS molecule. The apparent rate constants were fitted by models describing conformational selection (solid line) and induced fit (dashed line) binding models (Paul and Weikl, 2016).

3.2 Sf6 TSP as a tool for detection of *Shigella flexneri* cells

Sf6 TSP is a highly specific binder of *Shigella flexneri* Y polysaccharide. This specificity might allow using the protein as a sensor to detect the bacterium. Already as low as ten colony forming units (cfu) of *Shigella flexneri* bacteria can cause an infection in the host, which emphasizes the need to detect very small concentrations of this pathogen. A fast detection tool would need to fulfill the following criteria: First, the potentially contaminated sample can be applied in a test without further concentration or overnight cultivation and second the tool should exhibit a high specificity so it would be applicable on bacterial mixtures as the selection and isolation of bacteria takes additional time.

3.2.1 Detection of *Shigella flexneri* isolates by ELITA

Patients infected with *Shigella flexneri* in Germany are registered and stool samples need to be sent to the National Reference Centre for *Salmonella* and other Bacterial Enterics in Wernigerode, Germany. These samples are routinely characterized by immunologic serotyping and biochemical characterization. Both methods can sometimes produce ambiguous results, for example *Shigella* and *E. coli* species are difficult to distinguish (Lan and Reeves, 2002; rki.de/DE/Content/Infekt/NRZ/Salmonellen, 05.04.2017). An alternative approach is PCR-based species characterization of specific genetic markers (Li et al., 2009).
The application of P22 and 9NA TSP in ELISA-like tailspike adsorption (ELITA) has been described for the characterization of phase variations in *Salmonella typhimurium* based on the different TSP specificities for glucosylated or non-glucosylated *Salmonella* O-antigens (Schmidt et al., 2016). In ELITA, bacteria grown to the stationary phase are adsorbed at microtiter plate surfaces. Tailspike proteins modified with N-terminal *Strep*-tag®II are bound to the O-antigen of immobilized bacteria depending on their specificity and affinity. The amount of bound tailspike proteins can be quantified by a *Strep*-Tactin® labeled horseradish peroxidase reaction. Based on the established protocol for *Salmonella* detection using ELITA assays, Sf6 TSP_{EADA} was modified with N-terminal *Strep*-tag®II and used for the detection of *Shigella flexneri* Y and 2a isolates. Four *Shigella flexneri* Y and 2a isolates were provided by the Wernigerode collection, which were characterized before as *Shigella* strains by serotyping and by PCR-based evaluation of the O-antigen gene polymerase (*wzx* gene) (Li et al., 2009) (Tab. 8). Thereby, one of the *Shigella flexneri* 2a strains (08-7230) could not be characterized unambiguously as *Shigella* strain.

Tab. 8: Characterization of *Shigella flexneri* isolates by serotyping and PCR-based gene analysis.

Bacteria Isolate	Serotyping for Y: 3,4; 2a: II, 3,4	PCR (*wzx*)=O-antigen polymerase
S. flexneri Y 99-2001	+	+
S. flexneri Y 03-650	+	+
S. flexneri 2a 03-6557	+	+
S. flexneri 2a 08-7230	+	-

+: positive; -: negative. Sera from the company Sifin were used for the serotyping. *Shigella* isolates are named based on the classification system of the Wernigerode collection. Samples and data were provided by A. Fruth.

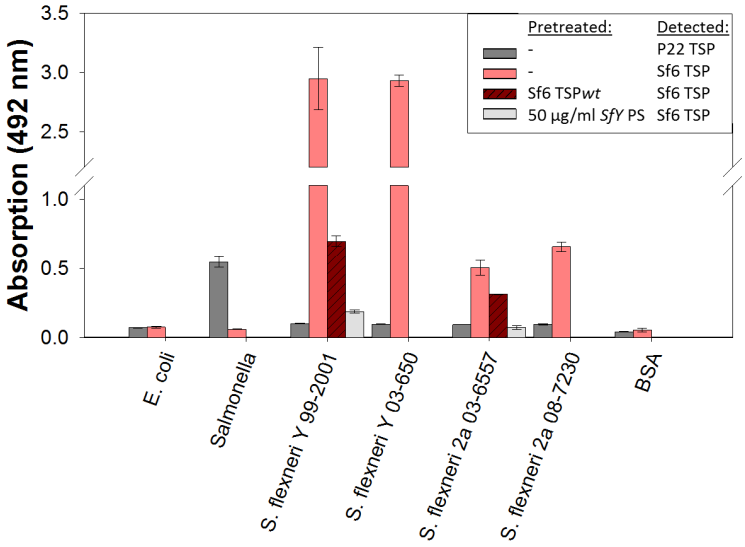

Fig. 24: Detection of *Shigella flexneri* isolates by *Strep*-tagged®II Sf6 TSP_{EADA}
ELISA-like tailspike adsorption assay with four *Shigella flexneri* isolates from the Wernigerode collection (numbers are according to the collection classifications) were tested with *Strep*-tag®II-modified Sf6 TSP_{EADA} and P22 TSP D392N (inactive variant) including the controls BSA, *E. coli H TD 2158* and *Salmonella typhimurium 06-01900* for all samples in triplicates. For *Shigella flexneri* Y99-2001 and 2a 03-6557 specificity tests were performed with pretreated cells using active Sf6 TSP*wt* and purified SfY polysaccharide.

In the ELITA, all four isolates could be identified as *Shigella flexneri* strains. Sf6 TSP could detect the isolates with absorption at the detection limit of 3 for *S. flexneri* Y and around 0.5 for *S. flexneri* 2a (Fig. 24). For *Salmonella*, detection with P22 TSP resulted in an absorption higher than 0.5, while for Sf6 TSP the reaction was negative (~0.1). In contrast, the four *Shigella* isolates were negative for P22 TSP. Detection of *E. coli* and BSA was also negative for both tailspike proteins (Fig. 24). Two of the isolates (Y 99-2001, 2a 03-6557) were further tested regarding the O-antigen specificity by pretreating the bacteria with enzymatically active Sf6 TSP*wt* to cleave of O-antigen chains from the bacteria. This indeed reduced binding signals by 76 % on the Y strain and by 38 % on the 2a strain. Furthermore the addition of purified free SfY PS together with *Strep*-tag®II-Sf6 TSP resulted in the deletion of the signal. Therefore the specificity of the signal due to *Shigella flexneri* PS binding was given. In conclusion, the ELITA protocol, which was established for *Salmonella* detection, was successfully transferable for *Shigella* detection. The test provides a high specificity, however over-night cultivation is a prerequisite.

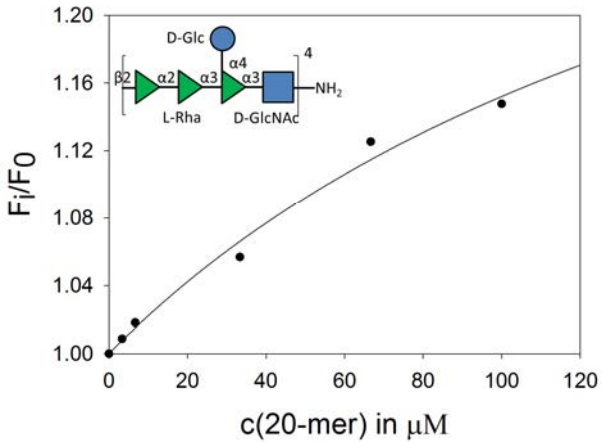

Fig. 25: Binding curve of Sf6 TSP_{EADA} with chemically synthesized Sf2a-icosasaccharide (20mer)
Fluorescence spectroscopy measured by excitation of 181 nM Sf6 TSP_{EADA} at 280 nm and emission detection at 340 nm. Synthetic Sf2a-icosasaccharide was added and the increasing fluorescence signal was fitted by the oligosaccharide binding curve (Baxa et al., 1996). This resulted in a dissociation constant of 186±92 µM. Synthetic Sf2a oligosaccharide fragments were provided by the group of Laurence Mullard at Institute Pasteur.

ELITA was also successfully for *Shigella flexneri* 2a. Preliminary results had shown that Sf6 TSP*wt* can cleave *S. flexneri* 2a O-antigen polysaccharides (Sf2a) with a reduced enzyme activity compared to SfY PS (unpublished data by Tom Scheidt). In this work, the binding of Sf6 TSP_{EADA} to a chemically synthesized icosasaccharide (20mer) containing the Sf2a repeating unit could be measured by fluorescence titration experiments and a dissociation constant of 186±91 µM was obtained (Fig. 25). Thus, the binding affinity of Sf6 TSP_{EADA} to Sf2a O-antigen oligosaccharide is by ~20fold lower than for SfY PS measured in SPR.

3.2.2 Cleavage products of *Shigella flexneri* polysaccharide

To further characterize the *S. flexneri* O-antigen structure, the isolate *S. flexneri* Y 99-2001 was chosen for polysaccharide and oligosaccharide preparation. In size exclusion chromatography of oligosaccharides obtained from Sf6 TSP*wt* polysaccharide cleavage three peaks were detected by changes in refraction indices at retention volumes of 171 (1), 184 (2) and 193 ml (3) (Fig. 26A). MALDI-TOF MS of the isolated peak fractions confirmed masses corresponding to octasaccharides for peak 3 (Fig. 26B-C), decasaccharides for peak 2 and dodecasaccharides for peak 1 (Fig. S 5). Theoretical and detected masses are listed in Table S 4. Moreover, about 14 % of the SfY 99-2001 sample was acetylated, which was not the case for the SfY polysaccharide used throughout this work. The analysis of the O-antigen structures of the residual *S. flexneri* isolates still needs to be performed.

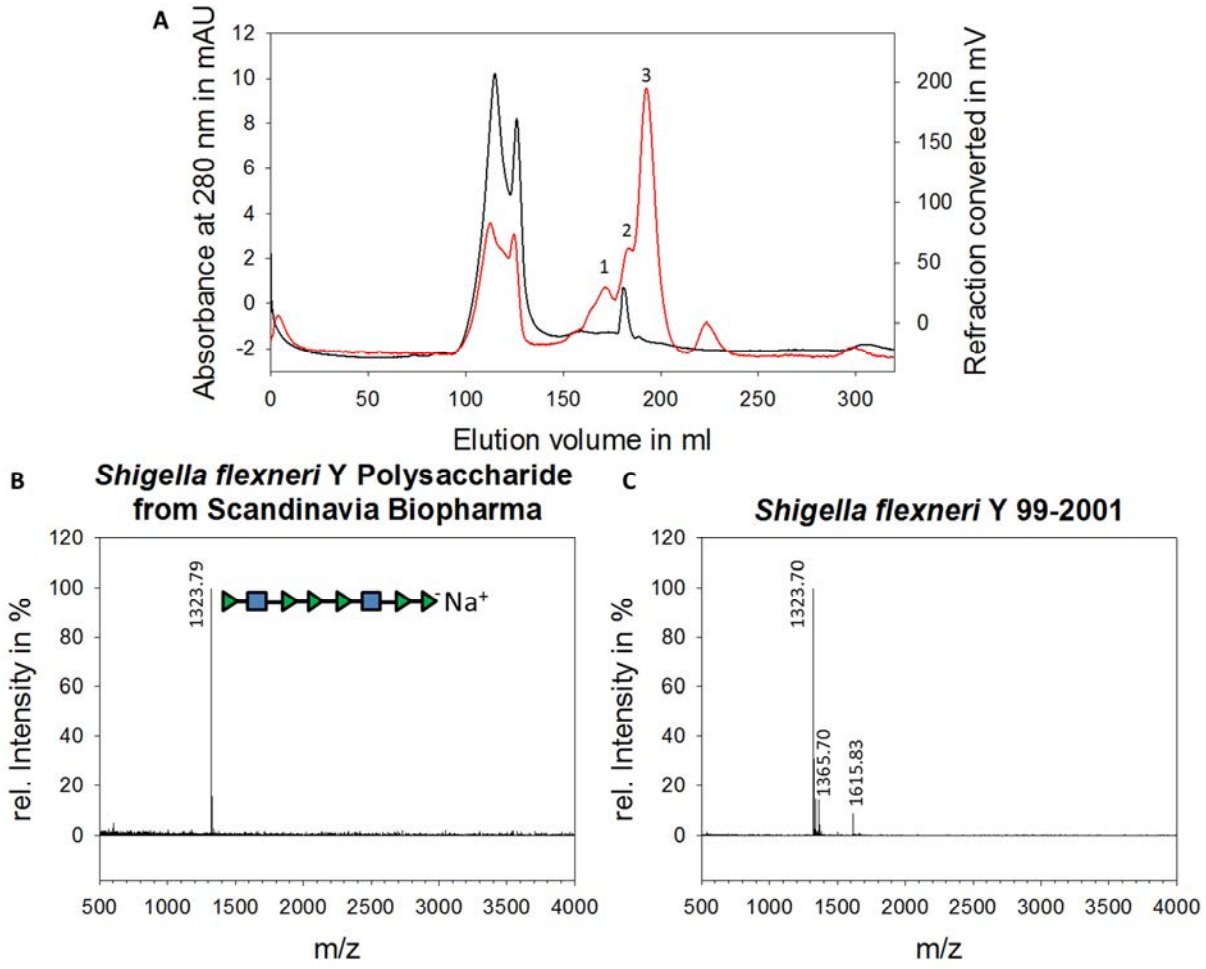

Fig. 26: MALDI-TOF mass spectrometry from oligosaccharides isolated by size exclusion chromatography
(A) Size exclusion chromatogram of SfY 99-2001 polysaccharide after digestion over night with Sf6 TSP*wt* detected by refractive index change (red). The tailspike protein can be detected by absorbance at 280 nm (black) at a retention volume of 100 ml. Oligosaccharide products were separated into three peaks depicting dodecasaccharides (1), decasaccharides (2) and octasaccharides (3). Masses of the peak species determined by MALDI-TOF MS for SfY polysaccharide **(B)** and *S. flexneri* Y 99-2001 **(C)**. Peak 3 from the chromatogram represents the octasaccharide-sodium salt with m/z of 1323.7.

3.2.3 Detection of *Shigella flexneri* polysaccharide by a fluorescent sensor

For the detection of bacteria, the pre-concentration step could be overcome potentially with the usage of a highly sensitive sensor. ELITA is based on absorbance. As fluorescence is more sensitive, the incorporation of a fluorescent label in Sf6 TSP was intended. Therefore, a cysteine mutant set with mutations in and around the oligosaccharide binding site was produced for covalent labeling with an environment sensitive fluorescent dye with emission in the visible light. Upon binding of SfY O-antigen a fluorescence increase is expected due to increased hydrophobicity around the conjugated label. This production of a fluorescent bacterial sensor for *Shigella flexneri* Y was successful. This sensor is in the patenting process,

thus no further experimental details are explained here and the fluorescent dye will only be named as "label" in the following.

Cysteine mutants of Sf6 TSP_{EADA} were thiol-conjugated with the fluorescent label and fluorescence emission spectra were recorded at different SfY PS concentrations (Fig. S 6A). One of the six labeled cysteine mutants thereby showed an increase of more than 100 % of fluorescence intensity at 540 nm upon PS binding (Fig. 27A). The equation for the calculation of dissociation constants for oligosaccharide binding to TSP described in (Baxa et al., 1996) was applied to determine the half maximal concentration of SfY PS at 1.76 µg/ml. This mass concentration can be seen as detection limit for SfY PS. Estimating the LPS mass on a single bacterial cell resulted in about 4.45 fg assuming that approximately 3.4 % of the bacterial dry mass is LPS (Darveau and Hancock, 1983; Loferer-Krößbacher et al., 1998). The 1.76 µg/ml SfY PS would correspond to a PS content of 460 cfu/ml given a mean O-antigen chain length of 15 *RU*. This detection limit would not be sufficient to detect in the necessary count range of 10 cfu/ml as infectious level of *Shigella flexneri*. To confirm these numbers the sensor would need to be tested on whole *Shigella flexneri* cells in fluorescence spectroscopy titration experiments. In conclusion a specific sensor with an increased detection limit could be produced. However, a pre-concentration of cells is still necessary. Thus, for the improvement of the detectable cell counts, the design of a high-affinity Sf6 TSP is of interest as the pre-concentration of cells could be achieved by binding to the protein for example in a flow through column system. The given Sf6 TSP-Cys label kinetic binding traces (Fig. S 6B) with SfY PS showed dissociation rates in the same range as for non-labeled non-mutated Sf6 TSP_{EADA} (Fig. 27B), which confirms that the binding affinity is not negatively affected by the fluorescent label.

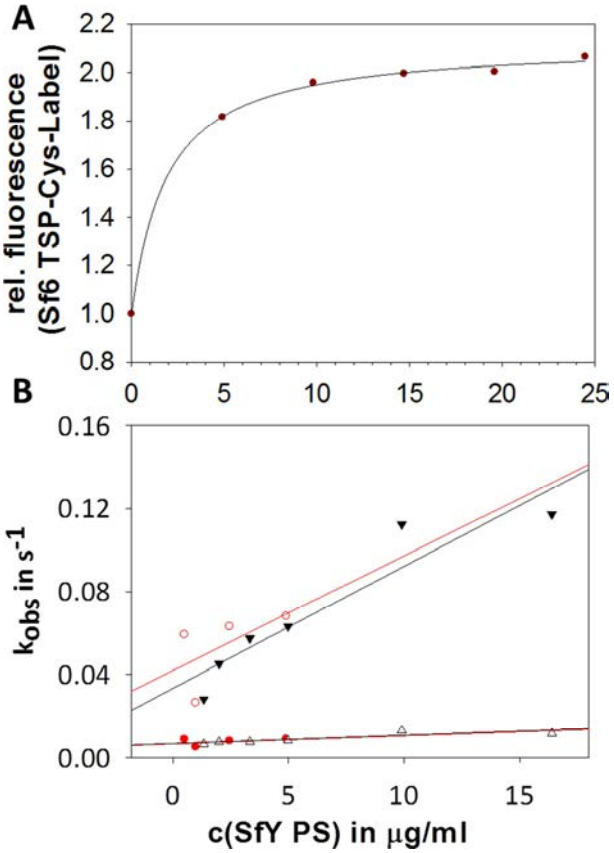

Fig. 27: Fluorescence detection of SfY polysaccharide by labeled Sf6 TSP.
(A) Labeled Sf6 TSP-Cys was excited at 492 nm and the emission was measured between 500-650 nm. With increasing SfY PS concentration, fluorescence increased to twice the signal. **(B)** Rate constants from kinetic binding traces of Sf6 TSP_{EADA} probed by intrinsic fluorescence excited at 280 nm (see Chapter 3.1 Fig. 23) (black). As in A with excitation of fluorescently labeled Sf6 TSP-Cys-Label at 492 nm and emission measured at 540 nm, kinetic binding traces were obtained. The k_{obs} were derived by biexponential fits and dissociation rates were calculated by $k_{obs}=k_{diss}+k_{ass}([P]+[L])$: k_{diss1}(Sf6 TSP_{EADA}-Cys-label)=0.0068/s, k_{diss1}(Sf6 TSP_{EADA})=0.0059/s, k_{diss2}(Sf6 TSP_{EADA}-Cys-label)=0.0419/s, k_{diss2}(Sf6 TSP_{EADA}) =0.0335/s.

3.3 Approaches to the rational design of Sf6 TSP high-affinity mutants

As depicted in the last chapter, the design of a high-affinity Sf6 TSP would have several advantages as it could improve the bacterial sensor detection limits and it could be applicable as a scientific tool in ELISA-like approaches to replace LPS or PS antibodies as well as facilitate specific surface immobilizations. The influence of mutations on the affinity towards the SfY O-antigen ligand was analyzed by computational prediction and experimental confirmation on the set of Sf6 TSP cysteine mutants from the thiol conjugation of the fluorescent sensor.

3.3.1 Influence of single amino acid exchanges to cysteine residues on Sf6 TSP binding affinity

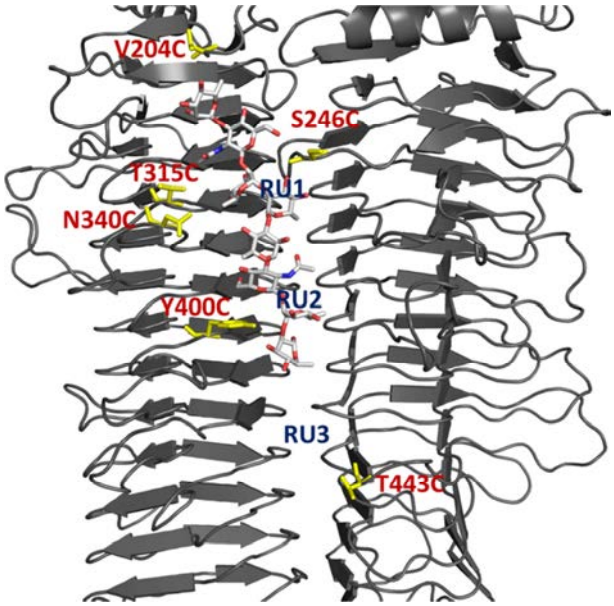

Fig. 28: Sf6 TSP binding site with cysteine mutations. Two subunits of Sf6 TSP_{EADA} (gray backbone, cartoon) are illustrated with octasaccharide ligand (white sticks) as *RU*1-2. The area of the dodecasaccharide binding site is labeled with *RU*3. All residues chosen for a cysteine mutation are shown as yellow sticks with the corresponding label in red.

Cysteines are normally excluded in design approaches due to the possible formation of disulfide bridges (Sharabi et al., 2013). Moreover cysteines are usually not preferred amino acids for participation in carbohydrate binding sites (Kulharia et al., 2009; Taroni et al., 2000). The interests in this work were the mechanisms that participate in increasing the affinity towards the polysaccharide receptor. However, the available cysteine mutants provided a test case for affinity analyses. Fig. 28 illustrates the used mutant set; all exchanged amino acid residues are located in loops adjacent to the elongated binding site.

The mutations V204C, T315C and T443C replace hydrophobic residues with cysteine as a potential hydrogen bond donor or acceptor and S246C, N340C and Y400C influence the binding as active hydrogen bond donor or acceptor in the non-mutated form. Thus, the exchange to cysteine might elucidate the influence of these residues.

3.3.1.1 MD simulations of Sf6 TSP cysteine mutants in complex with oligosaccharides

Sf6 TSP_{EADA} carrying single amino acid exchanges to cysteine were run in 100 ns MD simulations in TIP3P water with octasaccharide ligand (see Chapter 3.1.2). Moreover, Sf6 TSP_{EADA} T443C was simulated in presence of the dodecasaccharide. All simulations were analyzed with respect to ligand flexibility, hydrogen bonds and water distribution in the binding site to predict differences in binding affinity.

Ligand flexibility

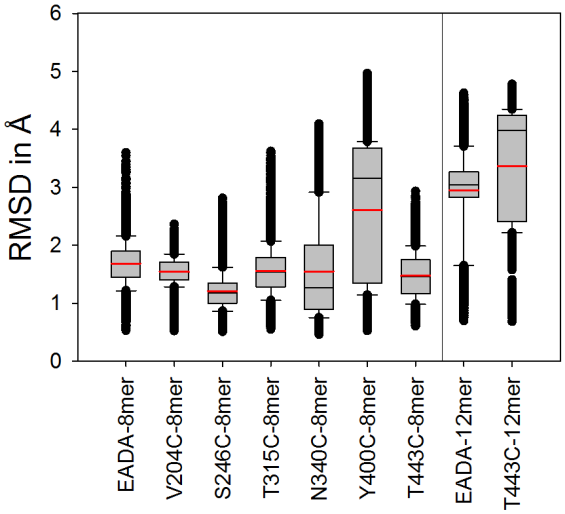

Fig. 29: Ligand RMSD traces over 100 ns simulation
Time dependent fluctuations of RMSD were calculated with *g_rms* (Gromacs). Red solid lines show the mean of RMSD. The upper, middle and lower black lines indicate 75 %, median and 25 % of the fluctuation, respectively. Error bars represent one standard deviation and black circles all outliers of the fluctuation.

Simulation traces of the simulated ligand in the Sf6 TSP_{EADA} cysteine mutant binding site are shown as mean RMSD in Fig. S 7. In Sf6 TSP_{EADA} without cysteine mutation a mean ligand flexibility RMSD of 1.68±0.37 Å was found (Fig. 29). Cysteine mutations led to a slight decrease of the ligand RMSD in four mutants. However, for Y400C the mean RMSD was increased to 2.61±1.13 Å (Fig. 29). Standard deviations of ligand fluctuations are decreased in the V204C (0.25 Å) and S246C (0.31 Å) mutants. This illustrates that in these cases the ligand is less flexibly bound which may result in increased binding affinity. By contrast, increased standard deviations of T315C and N340C mutants may indicate a decreased affinity in comparison to Sf6 TSP_{EADA}. Y400C also has an increased RMSD standard deviation and the octasaccharide detaches gradually over the simulation time of 100 ns (Fig. S 7), which strongly suggests a decrease in oligosaccharide affinity for this mutant. T443C with octasaccharide ligand shows approximately the same mean RMSD and standard deviation as Sf6 TSP_{EADA}, however with dodecasaccharide the mean RMSD and standard deviation are increased in comparison to EADA, which gives ambiguous information about the binding behavior of this mutant.

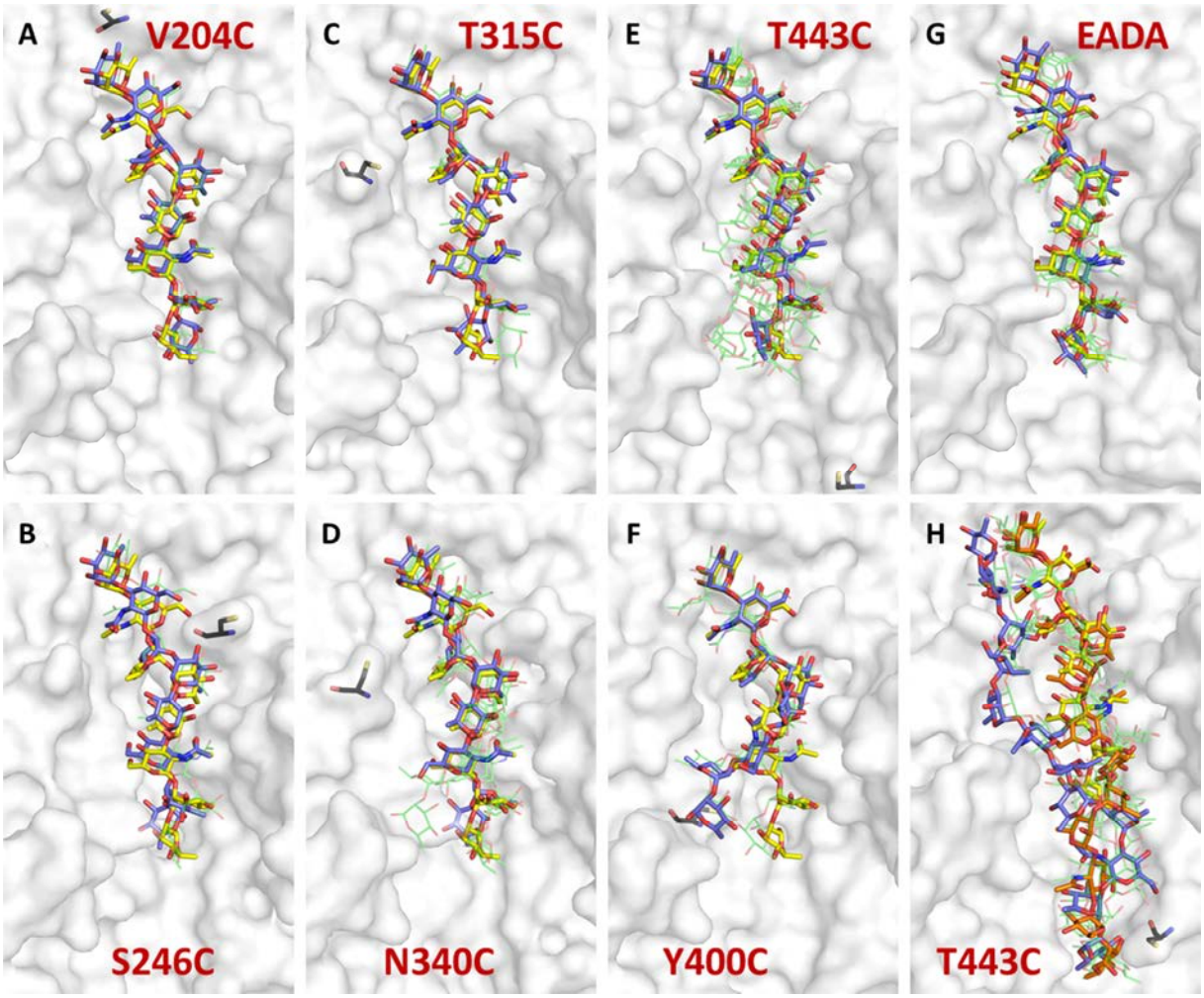

Fig. 30: MD simulation of Sf6 TSP cysteine mutant oligosaccharide complexes
Snapshots of oligosaccharide poses in the protein binding site (white surface) clustered from 100 ns simulation with a cutoff of 1.1 Å for octa- and 1.4 Å for dodecasaccharides. The crystal octasaccharide ligand is shown in yellow and the most populated central position of the simulated ligand in blue. Mutated cysteines are shown as black sticks. Additional ligand clusters during simulations are depicted as green thin transparent sticks. **(A)** V204C-octa, **(B)** S246C-octa, **(C)** T315C-octa, **(D)** N340C-octa, **(E)** T443C-octa, **(F)** Y400C-octa, **(G)** EADA-octa and **(H)** EADA-dodecasaccharide (in blue with green side clusters) and T443C-dodecasaccharide (in orange).

In a cluster analysis of the simulated ligand (see Chapter 3.1.2), a cut-off was defined resulting in five ligand conformer clusters each for EADA octa- and dodecasaccharide complexes (Fig. 30G-H). Using the same cut-off, in V204C, S246C and T315C two clusters were found that overlay well with the crystal structure ligand conformer (Fig. 30A-C). N340C, Y400C and T443C had a larger set of clusters in which tilted conformations at the reducing end occurred that bent away from the central binding groove (Fig. 30D-F). Surprisingly, T443C-dodecasaccharide resulted in only one cluster with a straight orientation, even though the mean RMSD and standard deviation of this complex was increased. Thus, ligand flexibility analysis suggests V204C and S246C to be stronger binders and Y400C to be a weaker binder. N340C, T315C and T443C are ambiguous in their predictions so far.

Hydrogen bonds

In the hydrogen bond analysis of ligands and proteins, only occupancies above 10 % were considered as a reliable interaction in 100 ns simulations (Tab. 9). Four conserved hydrogen bond forming residues are found in all mutants but their occupancy varies (marked in green in Tab. 9). Three of these four hydrogen bonds show high occupancies in the V204C and S246C mutants. In the Y400C mutant the hydrogen bond between Asp247 and RAM3 is present during most of the simulation time with 92 % occupancy. The sum of the total hydrogen bond occupancy was taken as a measure for the "binding strength". Sf6 TSP_{EADA} was used as a reference and set to a binding strength or relative occupancy of 1.00 and equivalent values were found for T315C, N340C and T443C. For the other cysteine mutants, the relative occupancies were increased up to 1.56fold for S246C.

Simulations of EADA and T443C with dodecasaccharide strongly vary in terms of hydrogen bond pairs (Table S 5). The relative hydrogen bond occupancy is higher for EADA than for T443C, which would suggest a decreased affinity for T443C.

Tab. 9: Occupancies of hydrogens bonds between Sf6 TSP_{EADA} cysteine mutants and octasaccharide ligand

Protein	Ligand	EADA-octa	V204C-octa	S246C-octa[a]	T315C-octa	N340C-octa	Y400C-octa	T443C-octa
ALA203-O	RAM1-O2	15.27 %						15.87 %
ARG230-NH1	RAM1-O2	25.65 %	56.49 %	56.89 %	38.32 %	23.15 %	43.01 %	33.03 %
ARG230-NH2	RAM1-O2	16.67 %	10.18 %	16.37 %	22.46 %	35.13 %	17.56 %	13.97 %
ARG257-NH1	RAM1-O2	12.87 %			12.48 %			
GLU293-OE1	RAM1-O2		28.44 %	31.24 %			22.85 %	10.58 %
GLU293-OE2	RAM1-O2	28.04 %	67.56 %	23.25 %			67.07 %	
ARG230-NH1	RAM1-O3					20.26 %		
ARG230-NH2	RAM1-O3	14.07 %						
ARG257-NH2	RAM1-O3	13.27 %						11.48 %
ARG230-NH2	RAM1-O5	10.08 %	26.05 %		10.98 %		18.16 %	11.88 %
SER246-O	NAG2-O6				10.48 %		18.36 %	30.44 %
ASP247-OD1	RAM3-O3	29.74 %	89.32 %	39.42 %	31.04 %	42.61 %		22.36 %
ASP247-OD2	RAM3-O3	31.64 %		32.73 %	58.58 %	37.03 %	91.72 %	50.20 %
THR248-OG1	RAM3-O4	11.48 %	11.28 %	31.94 %	21.66 %	11.48 %		
SER246-O	RAM4-O3	73.65 %	69.06 %	71.46 %	45.71 %	49.00 %	46.21 %	65.47 %
ASP245-O	RAM4-O4						29.44 %	
GLN280-NE2	RAM4-O4					16.57 %		
GLN280-NE2	NAG6-O7	17.07 %	61.58 %	68.76 %	44.61 %	32.44 %	18.46 %	32.34 %
TYR282-OH	NAG6-O7				10.48 %			
SER337-OG	RAM7-O4						13.07 %	
GLY339-N	RAM7-O4						18.66 %	
GLN325-OE1	RAM7-O4	26.35 %						19.06 %
TRP421-NE1	RAM8-O4				19.06 %			13.57 %
ASP397-OD2	RAM8-O4			75.95 %		41.52 %		
ASP397-OD1	RAM8-O3			61.58 %		32.83 %		
TYR400-OH	RAM8-O3				11.38 %			
TYR400-OH	RAM8-O2		18.16 %					
ASP425-OD1	RAM8-O1						25.75 %	
ASP425-OD2	RAM8-O1					10.68 %	32.24 %	
Relative Occupacy		**1.00**	**1.34**	**1.56**	**1.03**	**1.08**	**1.42**	**1.01**
Total number of H-bonds		**14**	**10**	**11**	**13**	**12**	**14**	**13**

a Hydrogen bonds in this mutant which are listed as Ser246 are formed by Cys246.

Thus, the hydrogen bond occupancies suggest that V204C, S246C and Y400C have a higher affinity than Sf6 TSP_{EADA}. Ligand flexibilities in V204C and S246C mutants agree well with this finding. However, the oligosaccharide bound to the Y400C mutant has more flexibility. The T315C and N340C mutants might have the same affinity towards the carbohydrate ligand as Sf6 TSP_{EADA} according to the hydrogen bond analysis. In contrast, the hydrogen bond pattern for octa- and dodecasaccharide bound to the T443C mutant is in favor for a decreased ligand affinity as already seen by flexibility analysis.

Water network

The *MobyWat* tool (see Chapter 3.4.2.2) was used to determine mobile water molecule positions on the protein surface during 100 ns MD simulations. From 30 crystallographically determined water positions, 20 could be reproduced in Sf6 TSP_{EADA} in complex with octasaccharide. Every cysteine mutant and Sf6 TSP_{EADA} with dodecasaccharide resulted in a reduced number of reproducable water positions compared to the crystal structure (Fig. S 8). The total number of simulated water position was highest for EADA and T443C, while V204C, T315C and Y400C had a reduced number of simulated water positions within the binding site (Tab. 10).

Tab. 10: Sf6 TSP mutant simulation and derived water positions by *MobyWat*

Protein mutant	Number of Water by *MobyWat*	Water in agreement with crystal structure (30)
EADA-octa	126	20
V204C-octa	106	16
S246C-octa	117	15
T315C-octa	110	12
N340C-octa	114	16
Y400C-octa	111	14
T443C-octa	126	17
EADA-dodeca	144	15
T443C-dodeca	143	16

The distribution of derived water molecule positions from the simulations did not show large differences between the mutants. The hydrogen bond network (Fig. 31) is separated into three parts in Sf6 TSP_{EADA} in complex with the octasaccharide (Fig. 31G). Two major accumulation spots of high densities of hydrogen bonds are observable in the upper part of the binding groove and in the middle left side (as indicated by green circles). No general rule exists that by default assigns favorable or unfavorable contributions of water molecules to the driving forces for ligand binding (Ladburry, 1996). However, the water network in the Sf6 TSP_{EADA} oligosaccharide complex contains dense hydrogen bond accumulation spots which suggest that these spots are important for the binding. Therefore they should also

appear in the mutant protein variants. On the contrary the appearance of dense hydrogen bond spots at other positions is most likely unfavorable. In previous simulations of Sf6 TSP mutants, it was observed that an increased number of molecules at the subunit interface might interfere with proper packing of the protein chains and result in "swelling" of the intersubunit binding site (unpublished data by Y. Kang). The dense hydrogen bond spot on the middle left site of the Sf6 TSP_{EADA} binding site can be observed in all cysteine mutants (Fig. 31), while the hydrogen bond spot in the upper part can only be observed in V204C, S246C and T315C. Hydrogen bond spots, which are not observed in Sf6 TSP_{EADA}, were not found. The water networks did not reveal strong effects of the cysteine mutations on binding site hydration. Hydrogen bond cluster distribution may suggest that V204C, S246C and T315C are equal binders as Sf6 TSP_{EADA}.

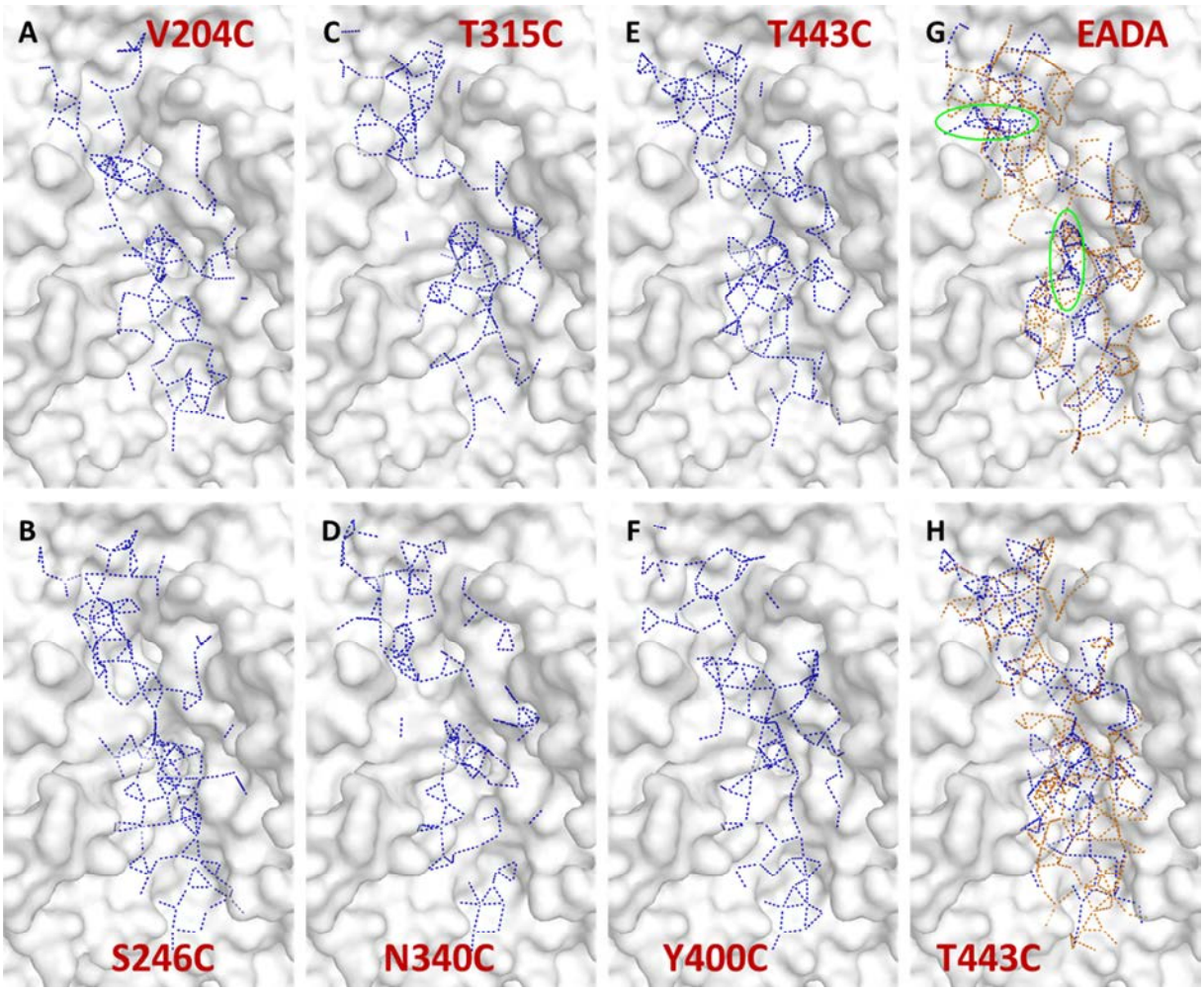

Fig. 31: Water networks from 100 ns simulation of Sf6 TSP_{EADA} cysteine mutants derived by *MobyWat*
Water networks were derived based on 100 ns simulation with *MobyWat* and the IDa list selection restricted to the top 20 % in the prediction mode (see Chapter 2.7). Resulting structures were superimposed with the crystal structure of Sf6 TSP_{EADA} and all water within 5 Å distance to the dodecasaccharide starting structures were selected. Networks are presented as dashed lines showing hydrogen bonds with maximal 3.5 Å distance. Networks in blue were derived from simulations with octa- and in orange with dodecasaccharide ligand. Green circles emphasize dense hydrogen bond regions of Sf6 TSP_{EADA}-octasaccharide simulated networks.

In conclusion, MD simulations suggest that the two cysteine mutants V204C and S246C may exhibit an increased oligosaccharide affinity, whereas the Y400C mutant most probably is a weak binder. No clear preference for increasing or decreasing affinities can be given for the residual three mutants T315C, N340C and T443C.

3.3.1.2 SPR analysis of Sf6 TSP mutant affinities towards immobilized polysaccharides

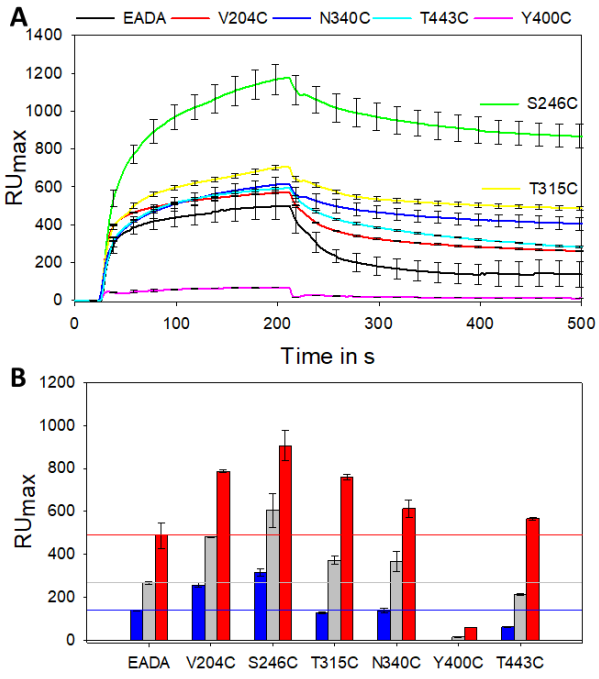

Fig. 32: Injection signal for Sf6 TSP_{EADA} cysteine mutants
(A) Sensograms of 8 µM protein injections over the surface with immobilized SfY polysaccharide (Chip1) of Sf6 TSP_{EADA} (black) and each of the cysteine mutants: V204C (red), S246C (yellow), T315C (green), N340C (blue), Y400C (purple) and T443C (cyan). **(B)** Maximal response units at 178 s of injection with Sf6 TSP_{EADA} cysteine mutants measured as triplicates for protein concentrations of 0.08 (blue), 0.8 (gray) and 8 µM (red). The horizontal lines mark the maximal response of Sf6 TSP_{EADA} in the respective concentrations.

For the experimental assessment of binding affinities of the cysteine mutant set, surface plasmon resonance interaction experiments were performed. For these experiments, the same activated polysaccharide surfaces were used as described in Chapter 3.1.3.1. For data evaluation the maximal response signals of each injection were compared (Jung et al., 2014), assuming that all mutants exhibit the same non-specific signal of around 4 % on the reference channel (more details see Fig. S 9). The cysteine mutants V204C, S246C, T315C and N340C showed a higher maximal response than Sf6 TSP_{EADA} (Fig. 32B), whereas Y400C and T443C had a lower maximal response. When the experiment was repeated on a second newly prepared polysaccharide surface (Chip2), V204C and T443C showed a decrease and increase in maximal response compared to Sf6 TSP_{EADA}, respectively (Fig. S 10B). Therefore, it is unclear from the maximal response signals, if V204C and T443C have increased or decreased affinities compared to Sf6 TSP_{EADA}, while S246C, T315C, and N340C appear with increased affinities and Y400C with a decreased affinity.

The dissociation rate from the polysaccharide surface is reduced for S246C, T315C, N340C and T443C compared to EADA, V204C depicts approximately the same dissociation rate and Y400C has an increased dissociation rate (Fig. 32A). The difference of the dissociation curves was only accessed visually as a fit with bi- and triexponential curves or any dissociation

model was not successful. Nevertheless, the association curves could be fitted with a biexponential curve to evaluate differences between the mutants (Fig. 33A-B). Thereby the results of both measurements agree with an increased apparent association rate for V204C and Y400C and a decreased apparent association rate for S246C (Fig. 33C). T315C, N340C and T443C apparent association rates are similar to the one of EADA with a slight increase regarding Chip1 and a slight decrease regarding Chip2.

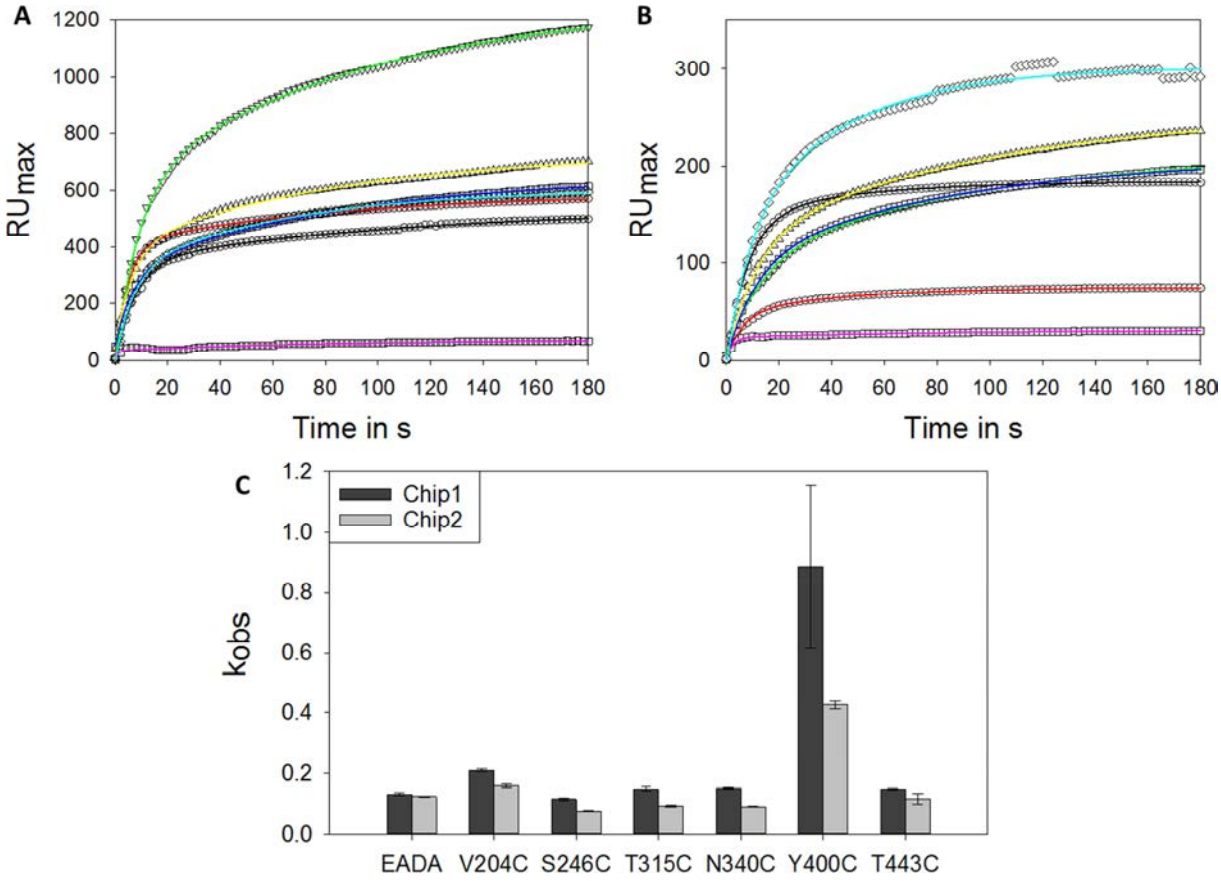

Fig. 33: Association curves of Sf6 TSP_{EADA} cysteine mutants
The association data of 8 µM were fitted with a biexponential equation for each Sf6 TSP_{EADA} cysteine mutant with measured data shown as symbols and the fits depicted as lines for EADA (black), V204C (red), S246C (green), T315V (yellow), N340C (blue), Y400C (purple) and T443C (cyan) on Chip1 **(A)** and Chip2 **(B)** as the mean of measured triplicates. All fits were calculated with R^2=0.997 or higher. **(C)** Calculated k_{obs} with standard deviations from biexponential fits of data in A and B.

As the dissociation constant is defined as quotient of off-rate and on-rate, a higher on-rate would lead to a decreased dissociation constant. The observed rates from the fitted associations still include both on- and off-rate. Therefore a direct calculation with the given data would not be possible. However, EADA and V204C showed comparable off-rates in the dissociation curves, while the on-rate from the apparent association rate for V204C is increased compared to EADA, suggesting an increased affinity also for this mutant. T443C

showed a slightly lower off-rate than EADA, but on-rates are ambiguous. Nevertheless strong differences in on- and off-rates mostly only result in small affinity changes (Nakajima, 2001). Thus, three of the six cysteine mutants (S246C, T315C and N340C) showed unambiguously increased equilibrium signals in comparison to EADA in interaction experiments on a SfY polysaccharide chip, while Y400C showed a strongly decreased signal. All results of computational prediction and experimental affinity measurements are summarized in Tab. 11. In conclusion, the S246C mutant was found to have a higher affinity towards surface immobilized polysaccharides, in agreement with the predictions from MD simulations. In contrast, Y400C could be identified as weaker binder computationally and experimentally. Nevertheless, computational and experimental results remain ambiguous for the majority of the analyzed mutant set.

Tab. 11: Comparison of simulated oligosaccharide behaviour and experimentally determined binding signals of Sf6 TSP_{EADA} cysteine mutants

Analysis[a]	EADA	V204C	S246C	T315C	N340C	Y400C	T443C
Ligand RMSD (Å)	1.68±0.37	1.54±0.25	1.21±0.31	1.56±0.43	1.55±0.81	2.61±1.13	1.48±0.38
Number of ligand cluster	5(octa) 5(dodeca)	2	2	2	4(bended)	3(bended)	7(bended) 1
Relative H-bond occupany	1.00	1.34	1.56	1.03	1.08	1.42	1.01
H-bond number	14	10	11	13	12	14	13
RU_{max} Chip1[b]	268.4±6.3	480.1±2.5	605.2±78.6	372.1±19.3	366.8±46.5	16.14±1.62	215.4±4.6
RU_{max} Chip2	102.8±3.5	38.26±0.93	191.2±37.5	158.5±25.7	228.2±27.0	58.23±5.56	165.9±8.9
Association Chip1	0.132±0.004	0.210±0.004	0.115±0.004	0.148±0.007	0.151±0.004	0.885±0.270	0.148±0.004
Association Chip2	0.123±0.002	0.161±0.006	0.077±0.002	0.092±0.002	0.091±0.002	0.426±0.013	0.117±0.017
Conclusion			Stronger			Weaker	

a Values colored in blue emphasize an increased binding propensity of oligosaccharides while values marked in red depict weaker binding in relation to Sf6 TSP_{EADA}.
b Maximal response units are listed for 0.8 µM Sf6 TSP_{EADA} cysteine mutant interactions.

3.3.2 Computational tools for amino acid selection and binding affinity prediction

The application of relatively long MD simulations of 100 ns and the simple evaluation of ligand flexibility, hydrogen bonds and water network revealed the affinity changes of two mutants. Therefore, it is an acceptable tool for affinity prediction of carbohydrate binding proteins. However, clear results are possibly only achieved for mutants with strong affinity differences compared to the reference. Thus a more sensitive computational tool would be needed together with a selection approach for amino acid exchanges.

3.3.2.1 Comparison of Sf6 TSP with *S. flexneri* O-antigen antibody binding sites

SfY polysaccharide as natural ligand of Sf6 TSP_{EADA} consists of 75 % α-L-Rha*p* and 25 % β-D-Glc*p*NAc. A search for these respective glycan moieties in Glyvicinity using α-L-Rha*p* as a ligand reveals 24 non-redundant pdb entries in comparison to β-D-Glc*p*NAc with 394 entries (Rojas-Macias and Lütteke, 2015). From the 24 pdb entries with rhamnose, four entries could be identified beside Sf6 TSP structures, which bind to a *Shigella flexneri* O-antigen structure. These are one antibody fragment (FAB) binding to *Shigella flexneri* 2a O-antigen penta- and decasaccharide (pdbID: 3C6S, 3BZ4) (Vulliez-Le Normand et al., 2008) and a FAB in complex with *Shigella flexneri* Y O-antigen tri- and pentasaccharide (pdbID: 1M7D, 1M7I) (Vyas et al., 2002). The binding sites are groove shaped including cavities which bury the branching Glc of the Sf2a oligosaccharide or the α(1-3) linked rhamnose for both O-antigens. Even though the binding site of the Sf2a Fab fragment is shallower than for SfY, both ligands are found in similar conformations when bound to the antibody fragment. If these conformers were modelled into more extended polysaccharide, stretches of right handed helices resulted (Vulliez-Le Normand et al., 2008; Vyas et al., 2002). However, these conformers do not seem to be stable and prevalent in solution as simulations of elongated fragments containing four *RU* of SfY PS in absence of a protein binding site also showed hairpin-like structures (Kang et al., 2014).

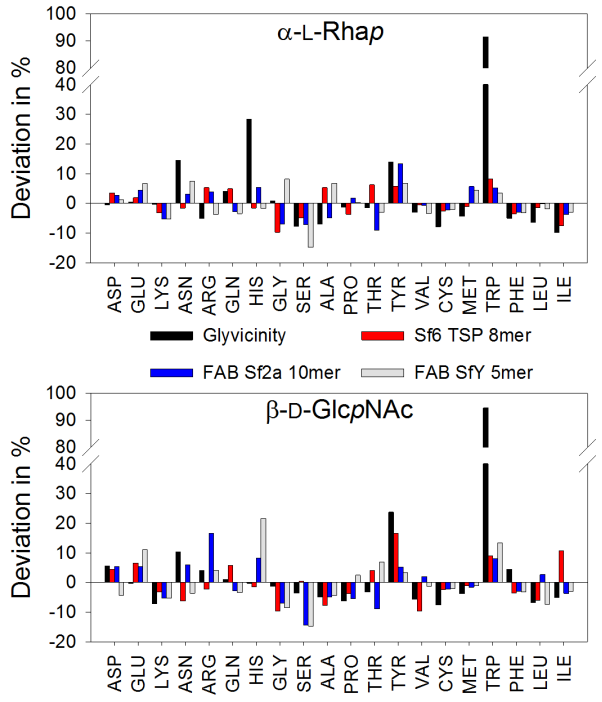

Fig. 34: Amino acid residues in the vicinity of α-L-Rha*p* and β-D-Glc*p*NAc
The percentile deviation of amino acid residues found in the vicinity of the carbohydrate residues over all structure in the PDB by Glyvicinity (black) in comparison with *Shigella flexneri* O-antigen binding proteins Sf6 TSP_{EADA} (pdbID: 4URR) (red) and FAB fragments binding Sf2a decasaccharide (pdbID: 3BZ4) and SfY pentasaccharide (pdbID: 1M7I). The deviation in % of the three proteins was calculated by the frequency of the amino acid in 5 Å distance to the sugar moiety divided by the overall frequency in the protein.

Similarities of the binding sites of the antibodies and tailspike protein were compared with the results of the Glyvicinity tool (Rojas-Macias and Lütteke, 2015), which was already used to depict frequent and non-frequent amino acid residues close to α-L-Rha*p* or β-D-Glc*p*NAc (Kang et al., 2016) (Fig. 34). In the vicinity of rhamnose the amino acids Asn, Gln, His, Tyr and Trp are highly overrepresented. In contrast Arg, Ser, Ala, Val, Cys, Met, Phe, Leu and Ile are underrepresented. The same over- or underrepresentation is observable for

Sf6 TSP_{EADA} and the antibodies for Ser, Tyr, Val, Cys, Trp, Phe, Leu and Ile. For GlcNAc Asp, Asn, Arg, Tyr, Trp and Phe were found to be overrepresented in the globel protein set in the PDB available, while otherwise the same residues as for rhamnose were underrepresented. For Sf6 TSP and the antibodies the same distributions as in the global protein set were observed for Gly, Ala, Cys, Met and Trp. In cases when all three proteins and the global protein set resulting from Glyvicinity have the same distribution, the residues are mostly underrepresented except for the aromatic residues, which are overrepresented for all four options (Fig. 34). These residues being over- or underrepresented for all four selections depict possible candidates to be introduced into the binding site. On one hand the natural binding affinity in general is low therefore introducing underrepresented residues in the binding site might lead to unexpected effects on the affinity. On the other hand aromatic residues, which are overrepresented, are known to play a role in carbohydrate-protein interactions and the inclusion of more aromatic residues might enhance the amount of CH-π interactions (Asensio et al., 2013).

3.3.2.2 Computational assessment of the influence of amino acid exchanges on oligosaccharide affinities

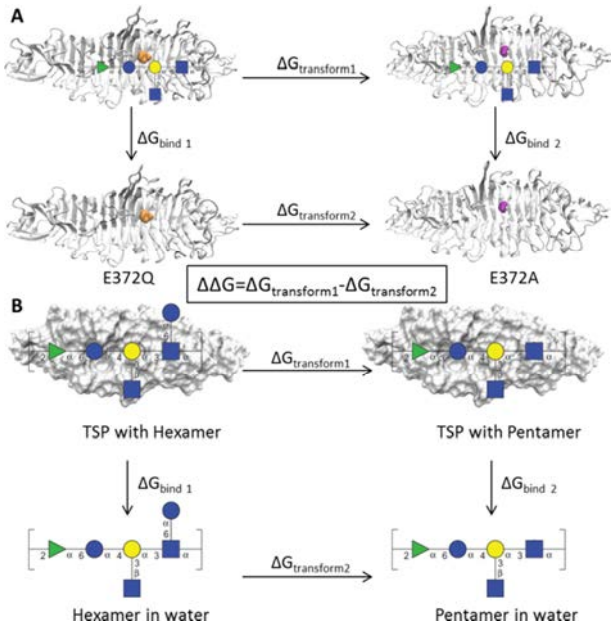

Fig. 35: IT-TI cycle for side chain or ligand exchange
(A) The protein (white protein backbone) is simulated in a box filled with explicit water molecules and with or without ligand (shown as pentasaccharide in SNFG nomenclature) in the binding site, while the side chain amino acid is mutated from glutamine (orange) to alanine (purple). **(B)** The protein variants (white surface) are simulated in a water box with ligand exchange during the calculations from hexa- to pentasaccharide. For the second term the ligand is exchanged sole in a box filled with explicit water molecules.

Besides the selection of single point mutations to influence binding affinity, a computer based technique to evaluate the affinity changes is necessary. Thus independent-trajectory thermodynamic integration (IT-TI) (Lawrenz et al., 2012) was tested on a different tailspike model as high-affinity mutants and the corresponding experimental binding free energies were available (More details to HK620 TSP are given and discussed in Chapter 3.4). The set-up included two carbohydrate ligands, on a pentasaccharide and the second the respective hexasaccharide with an additional branched glucose. The alchemical transformation of these two ligands in the

protein binding site and in solution as shown in Fig. 35B resulted in a $\Delta\Delta G_{TI}$ of -1.21±4.31 kJ/mol. This corresponds well to the experimental difference of binding free energies of the respective protein mutant with $\Delta\Delta G_{ITC}$=-2.1 kJ/mol. Nevertheless, the usability of IT-TI for Sf6 TSP was hindered by the unavailability of two different ligands resolved in the binding site. Therefore the alchemical transformation of side chains was intended, which is used in the evaluation of protein stability and folding behavior of mutants (Seeliger and de Groot, 2010). The used data set included mutations on one position between glutamate as wild type or glutamine and alanine as high affinity mutants (Fig. 35A). The data set was evaluated by implementation of available crystal structures, the change of the ionization state of nearby histidine to counterpart the changed charge modification of glutamate to non-charged residues, different protein restraints and TIP3P or -4P water. However, none of the calculated free energy differences resulted in a reasonable approximation of the experimental results. Therefore IT-TI was not applicable so far for the complexity of a tailspike protein-oligosaccharide complex system.

Tab. 12: Free energy differences calculated by thermodynamic integration (TI) on a tailspike high-affinity test case

Mutated residue	$\Delta\Delta G_{TI}$ (kJ/mol) (Pentasaccharide)	$\Delta\Delta G_{ITC}$ (kJ/mol) (Pentasaccharide)	$\Delta\Delta G_{TI}$ (kJ/mol) (Hexasaccharide)	$\Delta\Delta G_{ITC}$ (kJ/mol) (Hexasaccharide)
Gln → Glu (WT)	*63.06 ±5.12 *66.30 ±2.88 **43.32 ±4.15	15.8	*39.15 ±5.15 *28.25 ±2.94 **31.59 ±7.28 ***47.35 ±5.20	n. d.
Gln → Ala	*10.43 ±2.95 *2.46 ±2.57 *-4.98 ±3.29	-0.5	*25.09 ±2.07 *-1.06 ±3.19 *12.56 ±1.94 ****21.36±2.37	-2.1
Ala → Glu (WT)	*24.87 ±2.84	16.3	*33.68 ±2.46 *22.06 ±6.28	n. d.

*protein crystal structure, **protonation state His374, ***without backbone restraints, ****different water model TIP4P

3.4 Computational analysis of solvent contributions to thermodynamics of oligosaccharide complex formation in HK620 TSP

3.4.1 Description of the available structural and thermodynamic data sets

In the previous chapters, the attempt to design a high-affinity carbohydrate binding protein was described based on a structural approach. However, for a rational design approach, it would be useful to achieve a more general understanding on the driving forces leading to complex formation between proteins and carbohydrates to improve the success rate of chosen designs. In principle, this knowledge could then be implemented into a computer-based tool to facilitate experiments of mutant production and high-affinity screening.

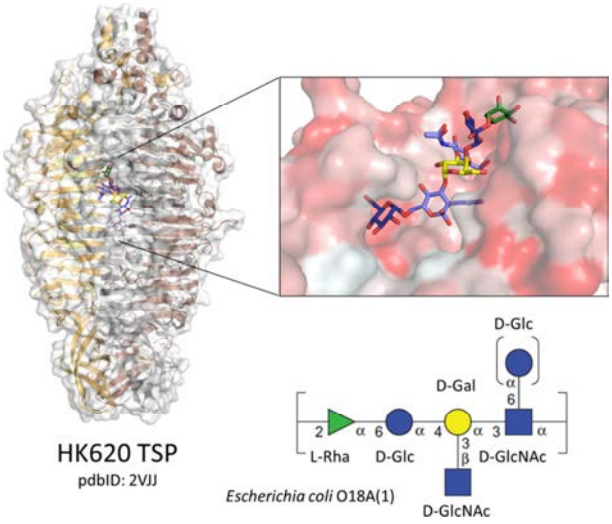

Fig. 36: HK620 tailspike protein with ligand
The homotrimeric HK620 TSP (shown as cartoon and white surface) has three binding sites located on the flat surface of each trimer subunit. HK620 TSP either binds pentasaccharide repeating units from *E. coli* O18A antigen or hexasaccharides from *E. coli* O18A1 with an additional branched glucose. The branched glucose is located in a hydrophobic groove in the binding site. The surface of the protein is colored depending on the hydrophobicity with white as polar and red as hydrophobic (pymolwiki.org/index.php/Color_h, 10.03.17; Eisenberg, 1984).

The tailspike protein from the bacteriophage HK620 was investigated in earlier studies (Barbirz et al., 2008; Broeker et al., 2013). The bacteriophage HK620 belongs to the *Podovirideae* family, as P22 and Sf6 bacteriophages (Dhillon et al., 1998; Israel et al., 1967; Lindberg et al., 1978), and infects the Gram-negative bacterium *Escherichia coli* H. HK620 TSP has an endo-N-acetylglucosamidase function and produces penta- or hexasaccharides from O-serogroup O18A or O18A1 bacterial O-antigen preperations, respectively (Broeker et al., 2013; Zaccheus et al., 2012) (Fig. 36). The hexasaccharide is bound with a dissociation constant in the submillimolar range. The two catalytic residues were identified as Asp339 and Glu372, the mutation of these residues to their non-charged amide equivalents or to alanine resulted in a loss of hydrolytic function. A set of single and double mutants of these two active site residues could be grouped according to their affinity range with a decreased affinity for D339N (DN), a mildly increased affinity by 10fold for D339A (DA) and a strongly increased tendency for complex formation by 1000fold in the mutants E372Q (EQ) and E372A (EA). Moreover, all double mutants containing the E372Q and E372A mutation,

i.e. D339N E372Q (DNEQ), D339A E372Q (DAEQ), D339N E372A (DNEA) and D339A E372A (DAEA) showed also the 1000fold increased affinity towards hexasaccharides (Broeker et al., 2013). All together, a complete thermodynamic data set measured by ITC was available for all high-affinity mutants for HK620 TSP in complex with two types of oligosaccharides (Fig. 37).

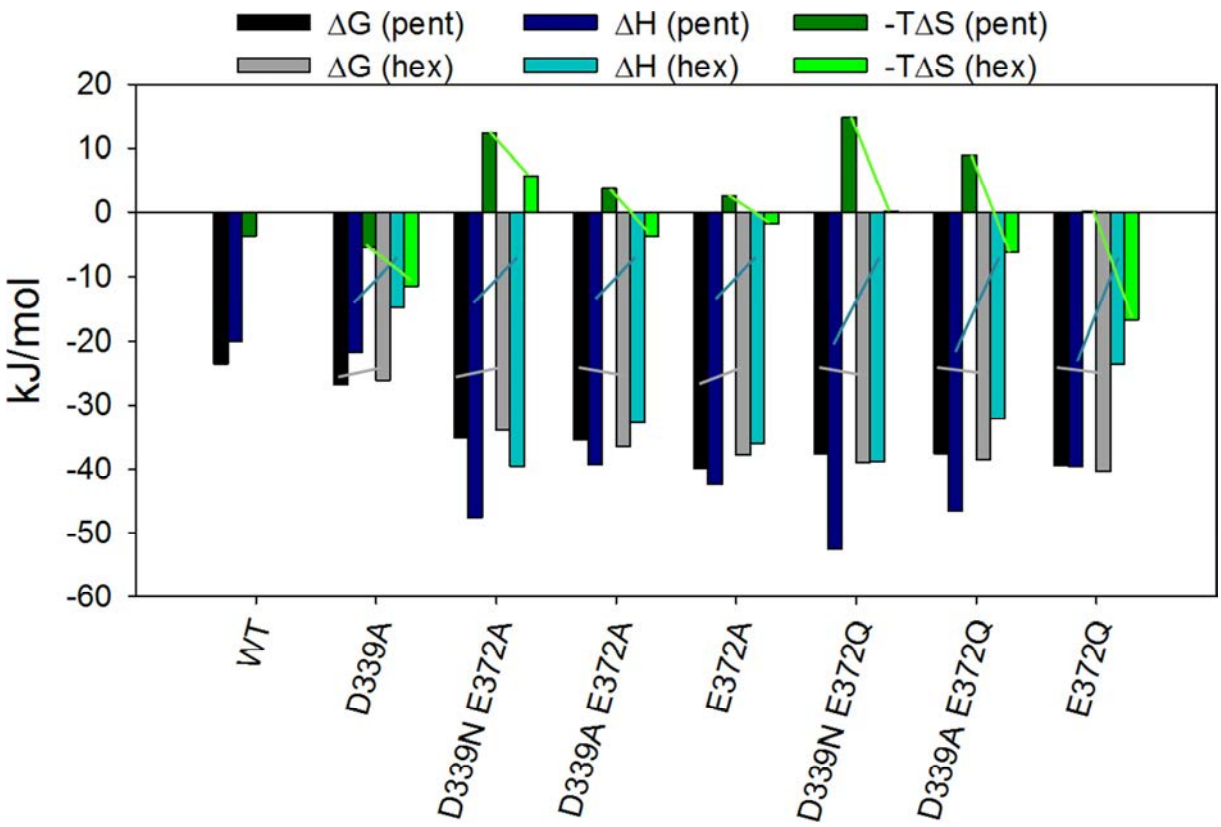

Fig. 37: Thermodynamic data for HK620 TSP mutants derived by ITC
Results of ITC measurements of the complex formation between HK620 TSP wild type or indicated mutants with one repeating unit of the *E. coli* O18A antigen (Pent) or the *E. coli* O18A1 antigen (Hex) at 25 °C. No data are available for the wild type hexasaccharide complex due to its low affinity (For the hexasaccharide, data are published in (Broeker et al., 2013)).

Structural models of the HK620 TSP-hexasaccharide complexes have been used to analyze the origins of the differences in the thermodynamic data set between different mutants and ligands (Broeker et al., 2013). In the crystal structures of WT and E372Q in complex with ligands, no differences in amino acid side chain positions could be observed, only E372Q can form an additional hydrogen bond to the ligand. However, in structures without ligand, the mutated Q372 is rotated by 110° concerted by a rotation of Y344. Thereby H374 is shifted, a salt bridge disrupted and an additional water site is formed, which in the WT binding site would be occupied by the carboxylate residue. Therefore, it was concluded that the affinity increase of 1000fold for E372Q derives from the combined side chain movements and from the expelled water upon binding due to an increase of entropy and enthalpy (Broeker et al.,

2013). These changes explained the affinity increase of E372Q mutants, but not in the E372A mutants, as alanine cannot form hydrogen bonds and also would not provide a connection for additional water positions. Furthermore, it remained unclear whether other binding modes were present in the double mutants. To address these questions, all single and double mutants were crystallized without ligand or with penta- or hexasaccharide before this work resulting in 27 structures for eight mutants and the wild type. Six of these structures were analyzes in earlier work (Barbirz et al., 2008; Broeker et al., 2013). Thus together with a thermodynamic data set provided for all mutants in complex with pentasaccharides, a complete experimental description of the HK620 TSP binding site was available that will be summarized briefly in the following (Kunstmann et al., *in preparation*).

Isothermal titration calorimetry (ITC) with penta- or hexasaccharide resulted in similar free enthalpies of binding to HK620 TSP ($\Delta\Delta G_{ITC}$=2 kJ/mol) (Fig. 37). However, an enthalpy and entropy difference between penta- and hexasaccharide binding was observed ($\Delta\Delta G_{ITC}$=6 kJ/mol) for E372A single and double mutants as well as for the D339A mutant. For E372Q single and double mutants, this difference increased to 14 kJ/mol. In both cases, the binding of the pentasaccharide is more enthalpically driven than for the hexasaccharide. Both effects might be assigned to solvent given the overall structures of the oligosaccharide complexes. In the crystal structures, the additional glucose of the hexasaccharide is enclosed in a hydrophobic binding groove (Fig. 36). If empty, *i.e.* without ligand or with pentasaccharide bound, five distinct water positions were observed in this pocket. These water molecules take the position occupied by glucose hydroxyl groups upon hexasaccharide binding (Fig. 38). In the present work it was therefore aimed to study the influence of the enclosed or replaced water in the glucose binding groove on the thermodynamic driving forces of penta- or hexasaccharide binding (Kunstmann et al., *in preparation*).

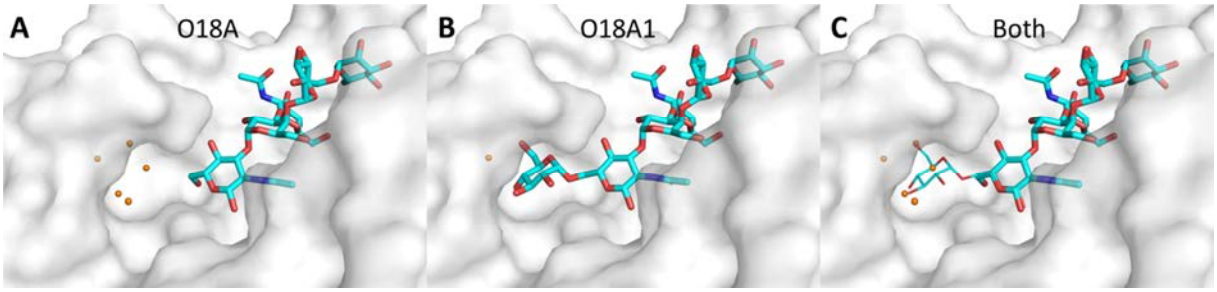

Fig. 38: Oligosaccharide ligands in the HK620 TSP binding site
(A) The pentasaccharide of O18A serotype O-antigen (blue sticks) is bound in the binding pocket (white surface) with five water molecules (orange) in a hydrophobic binding pocket, while the hexasaccharide from O18A1 O-antigen replaces four of these water molecules **(B)**. **(C)** Glucose hydroxyl groups overlay with the positions of the replaced water molecules. Pentasaccharide represented by thick sticks, hexasaccharide represented by thin sticks.

3.4.1.1 Analysis of binding site rearrangements upon ligand binding observed in the crystal structures

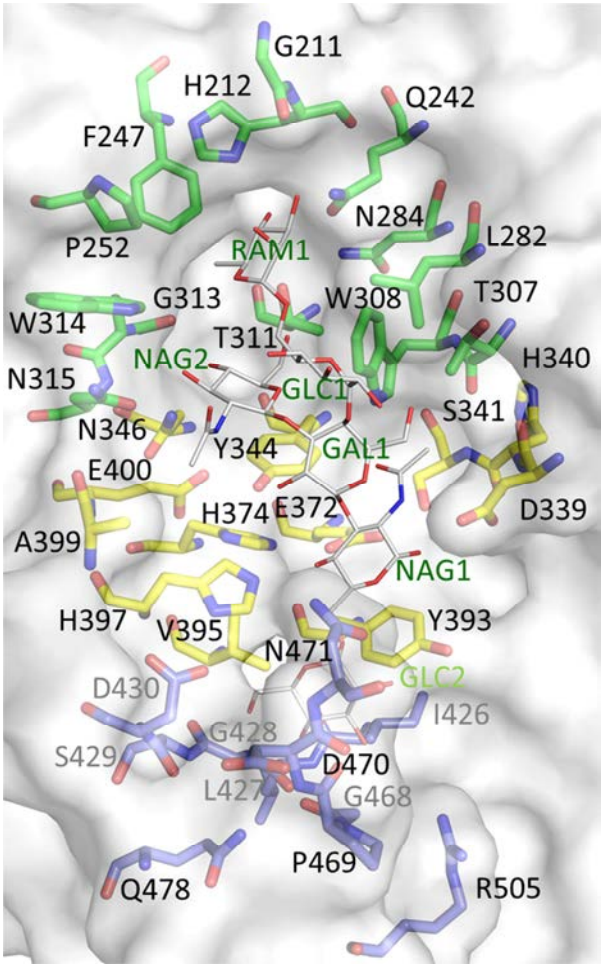

Fig. 39: Binding site residues of HK620 TSP
The protein binding site (white surface) with residues in 5 Å distance to the hexasaccharide (white sticks, green labels) were divided in upper (residue 211-315, green sticks), middle (residue 339-400, yellow sticks) and lower part (residue 426-505, blue sticks). The binding site residues are labeled in black and gray.

The side chains belonging to the binding pocket were defined as all residues within 5 Å distance to the hexasaccharide ligand in the wild type crystal structure (pdbID: 2X85). The residues of the binding site were compared and separated in an upper (residues 211-315), middle (residues 339-400) and lower part (residues 426-505) (Fig. 39). *Endo* and *exo* conformations of P252 or P469 were not taken into account, as the pyrrolidine ring shows a high flexibility in general (Deslaurier et al., 1974).

Most of the residues were populated in just one or two rotamers, a bound and an unbound position. The most populated position for a specific side chain residue is called the bulk state and if not stated differently the WT-hexasaccharide (WT-hex) structure defined the bulk state. All changes in side chain rotamers are listed in Table S 6 and Table S 7. An overview of all hydrogen bonds can be found in Table S 8 and Table S 9. Five residues with the most pronounced changes are described in the following.

The most diverse side chain positions in the upper part of the binding site were observed for N315 with four different arrangements (Fig. 40A). DNEA-apo and EA-apo exhibited two resolved forms; one of these twisted around χ1 from -66.4° (WT-hex) to -150° and shifted by 1.8 Å and 3.9 Å for ND2 and OD1 of the side chain, respectively. Additionally, the other conformation of DNEA-apo and EA-apo overlaps with DAEA-apo with a χ2 of 8-18° (-2.3° for WT-hex). For the amide group this resulted in a twist by 24° from the bulk position. DA-apo and the single and double mutations of E372Q in apo showed a turn of χ2 of up to 9.8° to the bulk. N315 builds three hydrogen bonds without ligand: N315N-N315OD1, N315ND2-

A399O and N315OD1-N346ND2. The latter is established in all variants upon every binding state, while the bond between N315ND2-A399O is only formed in the apo and not in the bound states. In WT this bond is not formed in all three states. The first hydrogen bond N315N-N315OD1 is only formed in the apo forms of DN, DA, EA, DNEQ and EQ.

In the middle part of the binding pocket, D339 is populated in four states (Fig. 40B), including a bound and unbound state for aspartates and asparagines each (Alanine mutations were not analyzed). The aspartate residues are turned by 25° upon binding at χ2 versus the lower part of the binding pocket resulting in a position between the two states of the amide group of asparagine. The amide group is turned by 66° at χ2 upon binding to the lower part of the binding pocket. This twist did not occur in DN-hex, the amide stayed as unbound rotamer.

The second mutated residue 372 showed the same variation as residue 339 with four major positions: two bound and unbound for glutamate and glutamine each (Fig. 40C). For WT-apo, DN-apo, DN-hex and DA-apo the residue is turned around χ3 by 20° upon binding of the ligand, what results in a shift of the carboxyl group with up to 0.9 Å. DN-hex again did not change upon binding. The glutamine on position 372 is strongly turned upon binding with χ2 80° and χ3 120° consequently the amide group is shifted by 2 Å. For EQ-pentasachharide (EQ-pent) the amide group is turned by 180° around χ3 compared to the other bound states, which might be a wrong assignment of the electron densities. For E372 with hexasaccharide only one direct hydrogen bond is formed between E372OE2-NAG1O1, while with pentasaccharide additionally E372OE1-NAG1O6 is formed. For Q372 with hexasaccharide two hydrogen bonds are formed (Q372OE1-NAG1O1 and Q372NE2-NAG1O5), with pentasaccharide even three hydrogen bonds exist with Q372NE2 also bound to NAG1O6. A372 in apo forms a hydrogen bond between A372N-S341OG.

H374 exhibits one of the strongest variations upon the 27 crystal structures with up to seven different rotamers in the binding pocket with varying bound forms of the WT, DN and DA (Fig. 40D). WT-pent, DN-pent/hex and DA-pent bound states were turned by 56° around χ2 to the bound state of all E372A/Q mutants, which is the bulk state. Close to the WT-pent rotamer, E372A mutants in apo are located with a turn by χ2=75° to the bulk state. The apo forms of DN/A were located in between with a χ2 of 40° to the bulk state. DA-hex shared the position with WT-hex as they are turned towards the bound positions of the E372A/Q mutants by χ2 of 20°. This bound state was placed parallel to WT-apo which was shifted towards D339 by 0.6 Å. The last rotamer of H374, E372Q mutants in apo exhibited an own position, which was parallel to their bound state and shifted towards Y344 by 1.0 Å. In the pentamer bound state, DA, DNEA and all E372Q mutants built a hydrogen bond to NAG1O6.

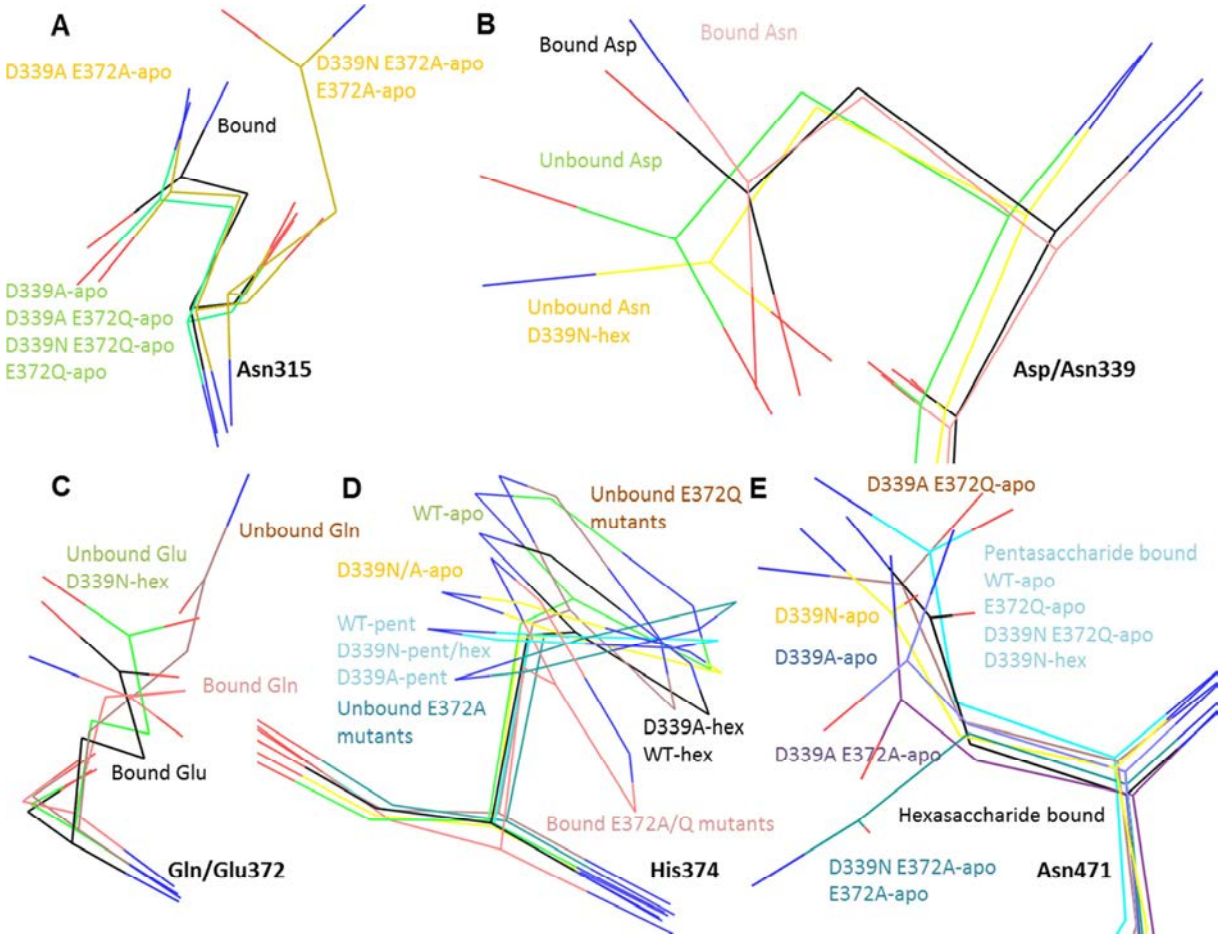

Fig. 40: Side chain rotamers found in different HK620 TSP crystal structures
The residues 315 **(A)**, 339 **(B)**, 372 **(C)**, 374 **(D)** and 471 **(E)** depicted from the 27 crystal structures. The related color coding describes the depicted structure.

The most changes in the lower binding site part are exhibited by N471 with seven populated states (Fig. 40E). The bulk position is occupied by all pentasaccharide bound forms in addition to WT, EQ and DNEQ in apo and DN-hex. All hexasaccharide complexes except for the DN mutant show N471 positioned in a χ2 of 20° to the bulk with the carbonyl group pointing outwards of the glucose binding pocket and DN-apo shifted parallel outwards from this position by 0.7 Å. DAEQ-apo is turned at χ2 by -40° with the carbonyl group pointing inwards the glucose binding pocket, which would close the glucose binding pocket. In contrast with an opening of the glucose binding pocket the apo structures of DA and DAEA were turned by χ1 of 40° and χ2 of 85° and 40°, respectively, together with a shift of 1.2 Å of Cg. The complete opening of the glucose binding pocket N471 was exhibited by the apo structures of DNEA and EA, as the whole residue is turned on χ1 by 110° with a shift of the side chain by 3.2 Å pointing outwards.

In conclusion, the conformational changes of the side chain rotamers can partly be grouped into E372Q or E372A mutant related changes. However, the total number of hydrogen bonds does not vary much and none of these arrangements truly explains the increased affinity for the E372A mutant forms. As a consequence the observation of the "dry" binding pocket surface did not give an explanation for the thermodynamic driving forces and the inclusion of the crystalized water molecules is necessary for a further explanation.

3.4.1.2 Distribution of crystal water in the HK620 TSP binding site

The previous work showed that in the HK620 TSP E372Q mutant, Q372 is turned in apo together with the two adjacent residues Y344 and H374 upon binding of the hexamer. The glutamine residue expels a water molecule from the binding site. This water replacement was accounted for 8-10 kJ/mol contribution to the free enthalpy complex formation (Broeker et al., 2013) (Fig. 41A-B). The same area of the binding site when including the mutated residue E372A exhibits less conformational changes upon the binding event. However, with an alanine at position 372 in direct proximity much more water molecules could be resolved in the crystal structure, which are diversely distributed for the apo form, but are more concentrated on two specific positions for the ligand bound forms of E372A mutants (Fig. 41C). These two well localized water positions of E372A mutants are not present in a WT ligand bound crystal structure because their positions would overlap with the removed carboxylate oxygens (Fig. 41D).

Therefore, the binding leads to the immobilization of two water molecules in the central binding site which are placed on positions of the former wild type glutamate carboxylate oxygens. Immobilizing these water molecules would first of all result in an entropy loss. However their position suggests a hydrogen bond bridge, which compensates for hydrogen bonds lost by the mutation. The obtained space by the mutation to alanine is filled by water molecules. Most double mutants with either E372A or E372Q background also had water molecules at the described position (In Fig. 41 water molecules are depicted for EA/Q, DNEA/Q and DAEA/Q). In conclusion, the affinity in E372Q mutants is derived from a removal of one water molecule, while for E372A mutants the affinity is derived from a fixation of two water molecules in the central binding site. For the high-affinity double mutants DNEA, DNEQ, DAEA and DAEQ, the binding mode is not changed in the 372 residue area compared to the single mutants EQ and EA.

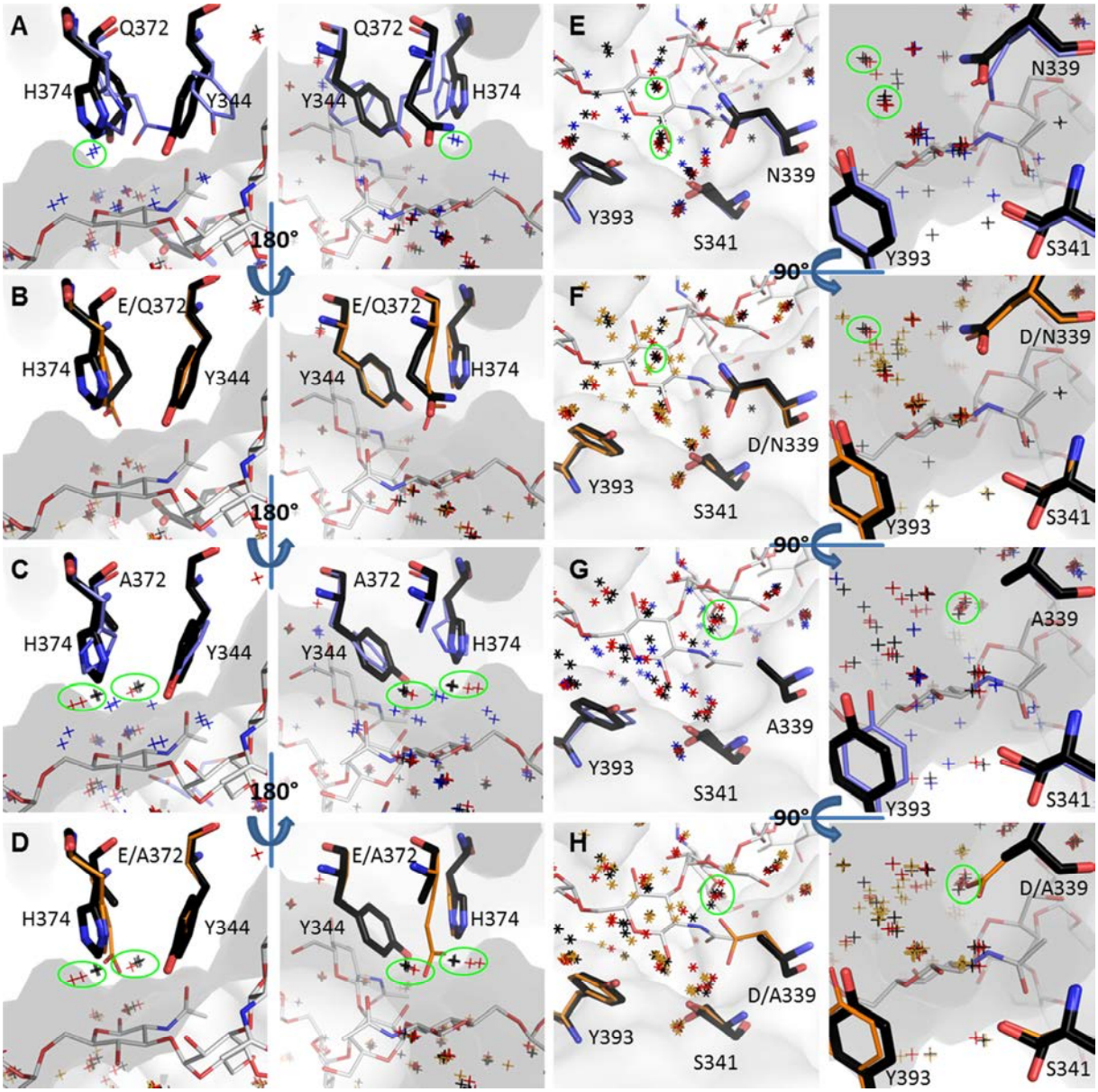

Fig. 41: Water arrangement around high-affinity mutated residues resolved from X-ray crystallography
(A) The mutated Q372 with the two adjacent binding site residues H374 and Y344 with hexasaccharide bound (white sticks) in the E372Q-hex (black sticks) and the E372Q-apo conformation (blue sticks). Water molecules (crosses) from all crystal structure with E372Q mutation are colored according to the binding state with apo=blue, penta-=red and hexasaccharide=black. **(B)** The same residues as in A are depicted in black for E372Q-hex overlaid with the same residues from the WT-hex structure (orange sticks). Water molecule positions from bound conformations of E372Q mutants (penta-=red, hexasaccharide=black) are shown in combination with water positions from WT structure with ligand (yellow crosses). **(C, D)** The mutated A372 is presented as in A and B with the same color coding and water positions from all E372A mutants. For A-D, the right panel is turned by 180° perpendicular to the left penal. Green circles emphasize water molecule positions discussed in the text. **(E,F)** The mutated N339 with the two adjacent binding site residues S341 and Y393 in comparison of apo to hexasaccharide form **(E)** and between WT and mutated forms **(F)**. **(G,H)** The same is depicted as in E and F for the A339 mutations. For E-H, the left panel is rotated by 90° to the right panel. The color coding stays the same in all included figures.

The double mutation might influence the most the binding site area around the varying residue 339. D339 is located adjacent to S341 and Y393 in the binding site. The conformational changes of these residues upon ligand binding are minor. Only a slight shift of the asparagine out of the binding pocket with ligand is observed in the crystal structures.

Nevertheless water positions change upon ligand binding for N339 mutants: Without ligand, water molecules preferably occupy positions which are replaced by the hydroxyl groups of the oligosaccharide, a phenomenon often encountered in carbohydrate binding sites (Modenutti et al., 2015). In presence of the oligosaccharide ligand, two positions of water clusters can be identified which are bridging a potential hydrogen bond network above the ligand that would lead to an enclosure of the ligand between protein binding site and water (Fig. 41E). From these positions only one is specific for N339 mutants as it is not observed in the bound WT or single mutations EA and EQ (Fig. 41F). In A339 mutants, the same effect upon ligand binding can be observed, whereby the specific localization of the water clusters is not as strong as for N339 (Fig. 41G). Thereby, an additional water position appears in A339 mutants in ligand bound form which is positioned adjacent to the original position of aspartate (WT/EA/EQ) or asparagine (DN/DNEA/DNEQ) (Fig. 41H). The entropic cost of the binding increases from the single mutants EA and EQ to the D339A mutants to the D339N mutants (Fig. 37). The water molecule distribution in the bound state of D339N mutants is more ordered around N339 than observed for A339, which might explain this entropic cost. As an alanine cannot form hydrogen bonds by itself, the more diverse distribution of water molecules around it would be expected. However, the number of immobilized water clusters is the same for D339N and D339A mutants. The removal of water around the charged residue aspartate (D339) is connected with an entropic gain, but would be enthalpically expensive (Biela et al., 2013). With an alanine mutant, the freedom of immobilized water positions upon ligand binding is still moderate compared to water molecules fixed around the polar asparagine. In conclusion, the thermodynamic contributions of the double mutants can be partly encountered to the removed or present additional charge.

Five to six water positions can be observed in the glucose binding groove of pentasaccharide bound crystals. The enclosure of these water molecules may account for the entropic penalty found when the ligand is inserted into the binding site. In hexasaccharide bound complex structures these water molecules are released. The difference between E372A and E372Q mutants may therefore result from a connection of the groove water molecules with the residual binding site in E372A mutants observed at the entrance of the groove (Fig. 42A).

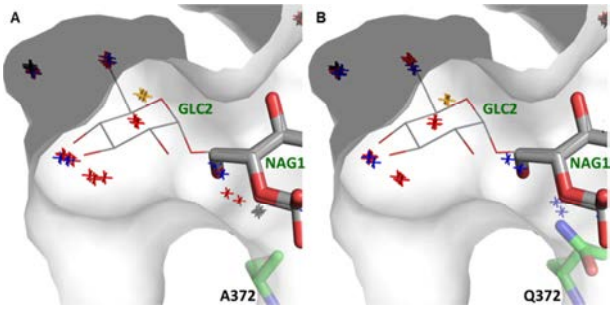

Fig. 42: Water molecules in the glucose binding groove from crystal structure analysis
The glucose binding groove with hexasaccharide (thin white sticks) and pentasaccharide (thick white sticks) with all crystallographic water positions (crosses) from E372A **(A)** and E372Q mutants **(B)** in apo (blue), pentamer (red) and hexamer (black). One position is only occupied in two of the three protein pentasaccharide variants each (yellow).

In E372Q these connecting water molecules are not found in any of the ligand bound structures as the glutamine is located at the entrance of the groove and can form a gate for water exchange between the rather flat oligosaccharide binding site and the occluded glucose binding groove (Fig. 42B).

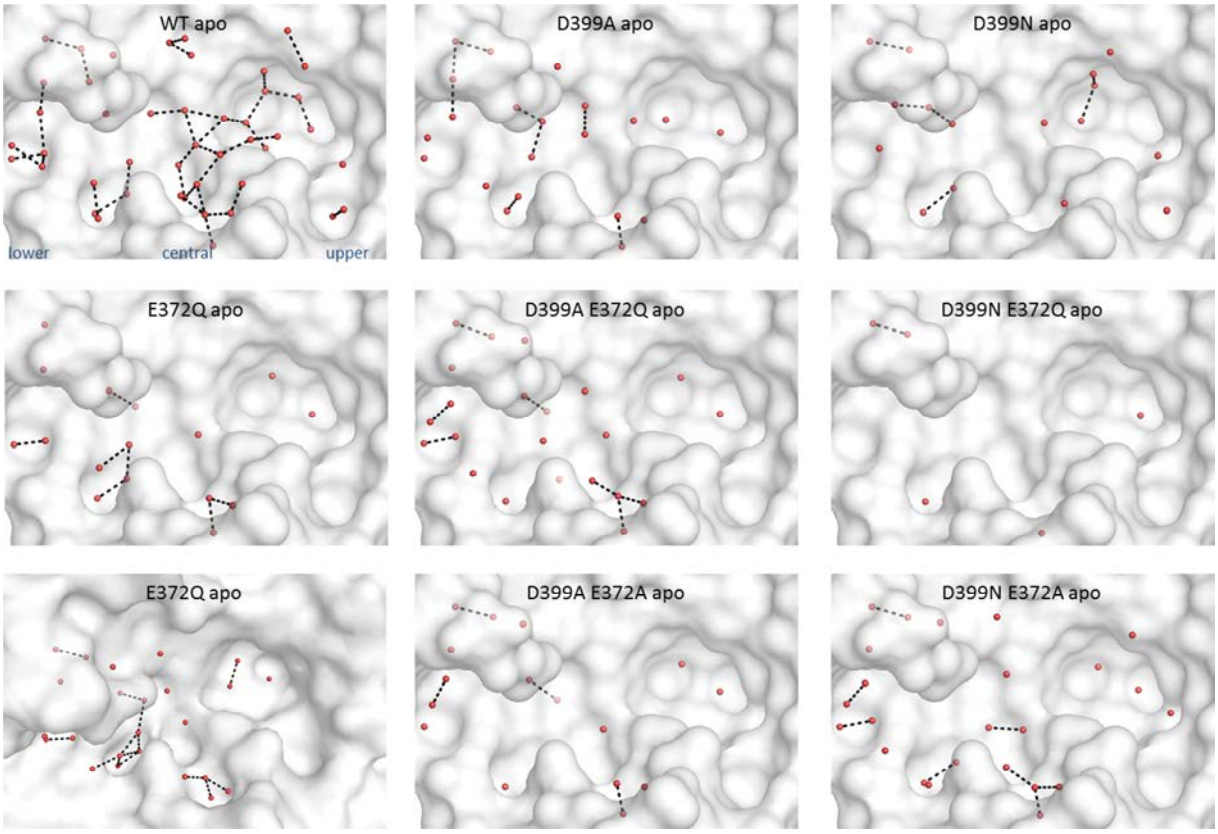

Fig. 43: Crystallographic water molecules in apo crystal structures.
Binding site surface (white) of various mutations and wild type of HK620 TSP depicted with all water molecules within 5 Å of the WT-hexasaccharide ligand. Water molecules are depicted as red spheres and black dashed lines indicate a hydrogen bond with maximal 3.5 Å distance.

In conclusion, the high affinity behaviour of EQ/EA derives from either a fixation or a removal of water in the binding site. The connection or isolation of water molecules in a deep hydrophobic pocket leads to different enthalpic and entropic contributions for the binding of two oligosaccharide ligands. Therefore different solvation patterns play a crucial role in the binding event. However, among the 27 crystal structures resolved for HK620 TSP, no water network could be assigned in the mutants (Fig. 43) and not all double mutants confirmed the discussed relevant water positions around residue 372 and 339.

3.4.2 MD simulations for HK620 TSP binding site and water flexibility analyses

The static picture of X-ray crystallography can be overcome by using MD simulations. Therefore, HK620 TSP wild type and the high affinity mutants EQ and EA without ligand and with penta- and hexasaccharide were simulated for 10 ns in TIP3P water in a box with periodic boundary conditions with HK620 TSP as a truncated form of one monomer including only the amino acids 205-521: a 12 Å cut around the oligosaccharide binding site to reduce computational costs. In total, nine simulations were used for analysis of binding site and ligand flexibility, for the assignment of water networks on the protein surface by *MobyWat* and as the basis of free energy calculations of specific crystallographic water molecules.
Furthermore, for the evaluation of conserved water positions, elongated simulations of E372A without ligand (50 ns) and with penta- and hexasaccharide (100 ns) were run for one complete monomer of HK620 TSP in each TIP3P, -4P and -5P, as the water model has an influence on carbohydrate-protein interactions in simulation (Fadda and Woods, 2011).
These 18 simulations showed a stable protein conformation over the simulation run time without detachment of the ligand from the binding site. Truncated forms of the β-helix sheet of HK620 TSP still exhibited the structural backbone conformation states after simulation, as full monomer structures (Fig. S 11).

3.4.2.1 The ligand and binding site residue flexibility

The oligosaccharide ligand and amino acid side chains that constitute the binding site can exhibit different levels of flexibility and movements, which can contribute to the overall thermodynamic signature of complex formation. Comparison of the 27 crystal structures showed major movements in the central part of the binding site (Section 3.4.1.1). From the 10 ns simulations of WT and EQ/EA in their three binding states, root mean square fluctuations (RMSF) of every binding site residue were calculated. An error of 10 % on each value was assumed to be a measure between the differences of the calculated RMSF. Residues were considered as different in their flexibility when these 10 % errors did not overlap with the RMSF calculated for the apo form of the residue. The resulting differential RMSF bar diagram is derived by subtraction of the apo form RMSF (Fig. 44, the figure only contains residues that markedly changed in flexibility during complex formation. All residues are included in Fig. S 12).

In the upper part of the binding site (residues 211-315) WT and EA are more flexible than EQ. In WT-pentasaccharide (pent) H212 increased in flexibility and in WT-hexasaccharide (hex) N315 decreased. For EQ only one strong decrease in flexibility can be observed for N315 with pentasaccharide. For EA the flexibility increased for H212 and strongly decreased for N315 for both ligands. Additionally, L282 and W308 decreased in flexibility in EA. Nevertheless, the upper part includes the smallest number of flexibility changes of the whole binding site. The flexibility changes in the central part (residues 339-400) are the most prominent. For WT-pent the flexibility of five residues increases, while for EQ-pent six residues are increased and two residues are decreased upon binding. For EA-pent in total seven residues have a decreased flexibility. In contrast for WT-hex, seven residue flexibilities are decreased and V395 has an increased flexibility. EQ-hex shows the least flexibility changes with an increase for Y344 and H397 and a decrease for V395 and E400. Five residues in EA-hex are decreased in their flexibility and only S341 is increased. In the lower part of the binding site (residues 426-505) for WT-pent only an increase of D430 and a decrease for D470 are observed, while five residue flexibilities are decreased with hexasaccharide. In EQ-pent L427 and N471 are increased in their flexibility and D430 is decreased. EQ-hex shows only two decreased flexibilities for S429 and D430. In EA L427 is increased and S429 is decreased in flexibility for both ligands. Additionally there is a decrease of D430 and an increase of N471 in the pentasaccharide bound form.

These differences could also be derived by varying flexibility residues in the apo forms. By comparison of these apo states (Fig. S 12), only differences in the central and lower part of the binding site are observed. For the WT, in total four residues have an increased flexibility and two are reduced, thus exhibiting the most changes compared to the mutations. For E372A, the flexibility of the three residues Y344, H374 and Y393 changed, which are adjacent to the mutated residue. For E372Q the residue Y393 is reduced in flexibility and H397 and D430 are increased, which are oriented towards the lower part of the binding site between the glucose binding groove and the mutation.

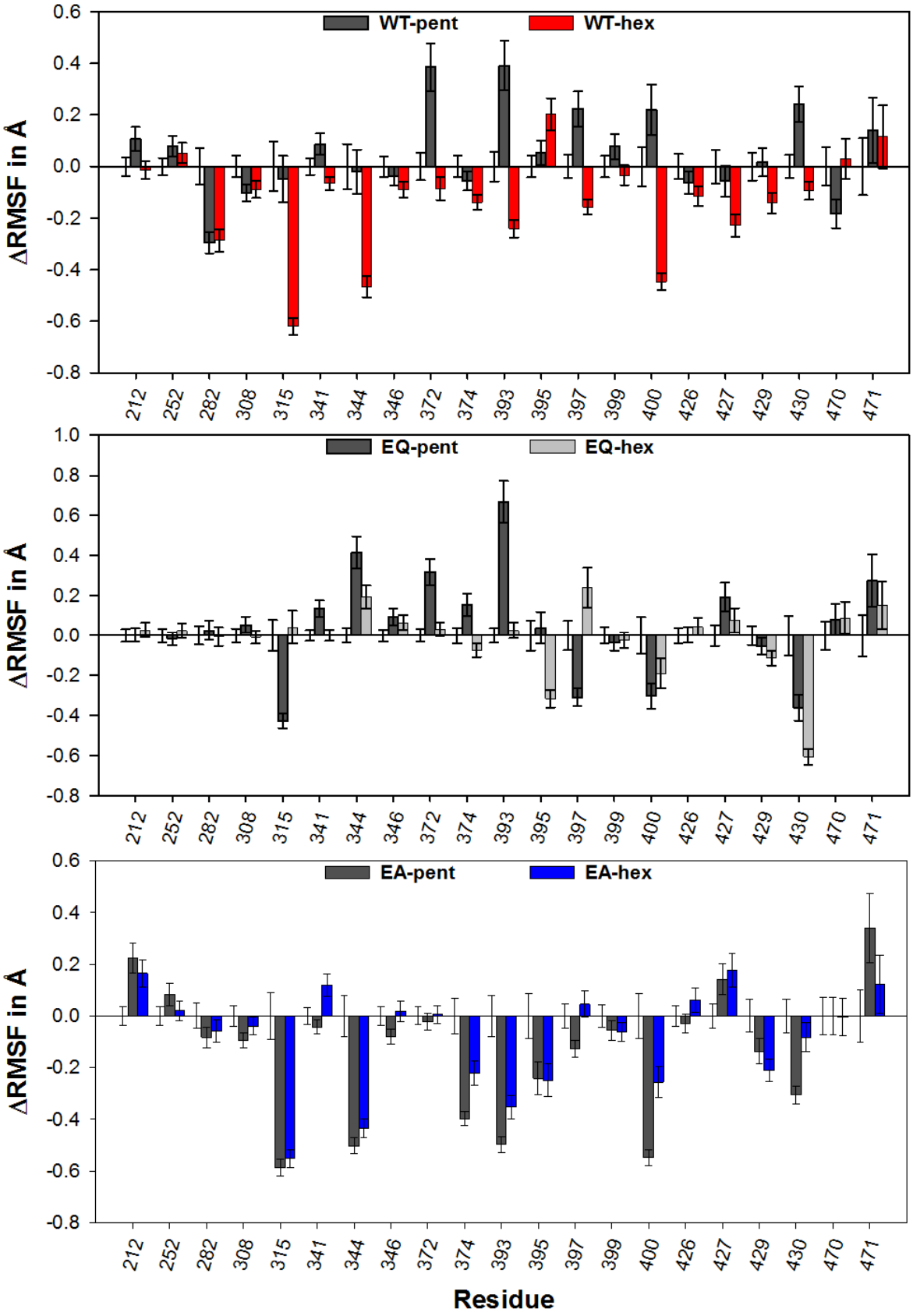

Fig. 44: RMSF differences of residues in the ligand binding site.
RMSF were derived from 10 ns simulations for whole residues with *g_rmsf* (Gromacs) for all binding site residues. RMSF differences were calculated by subtraction of the RMSF of the apo form. Error bars correspond to 10 % of the RMSF.

In crystal structures, the lid of the glucose binding pocket (residue 471-472) appears closed when a ligand is bound, based on the turn of N471 towards the protein surface (Fig. 46). The closed lid is observed in WT and EQ-apo, while in EA-apo the lid is open in the crystal structures with N471 turned outwards away from the protein surface. The fluctuation of root mean square displacement (RMSD) of the glucose binding groove lid (residues 468-472) were tracked during 10 ns of simulation, whereby the first nanosecond was excluded to ensure proper equilibration (Fig. S 13). The mean of the RMSD shows a different behavior for each of the protein variants (Fig. 45). For WT the RMSD decreases upon pentasaccharide binding and increases slightly upon hexasaccharide binding. In contrast for the high affinity mutants, the RMSD for the apo form is the lowest for EQ and increased to the same level with penta- and hexasaccharide. For EA the RMSD increased with pentasaccharide and remained unchanged for apo and hexasaccharide. By clustering the movement of residues 471-472 from the simulation, the lid opening and closure is depicted (Fig. 46). Interestingly in simulations N471 is turned outwards away from the protein surface in all main central populated conformations. For the apo and pentasaccharide form, the WT glucose binding groove lid is opened to the greatest extent as the backbone O of N471 is simultaneously shifted outwards from the binding site, while for EQ-apo this shift is the smallest. EQ-pentasaccharide exhibits the most conformations, as N471 oriented outwards is still able to form a hydrogen bond to the pentasaccharide. In contrast, the lid is more closed for WT-hexasaccharide than for the other forms and the strongest opening is observed for EQ-hexasaccharide. EA-apo and pentasaccharide agree with the crystal structure and EA-pentasaccharide includes one conformational state, where N471 is turned towards the binding site. Additionally EA-hexasaccharide forms a potential hydrogen bond from the backbone O of N471 to Glc2. In conclusion, an open conformation of the glucose binding groove lid was obtained from simulation in contrast to a closed conformation in crystal structures. Nevertheless the differences between the protein variants and binding modes are helpful to explain the thermodynamics.

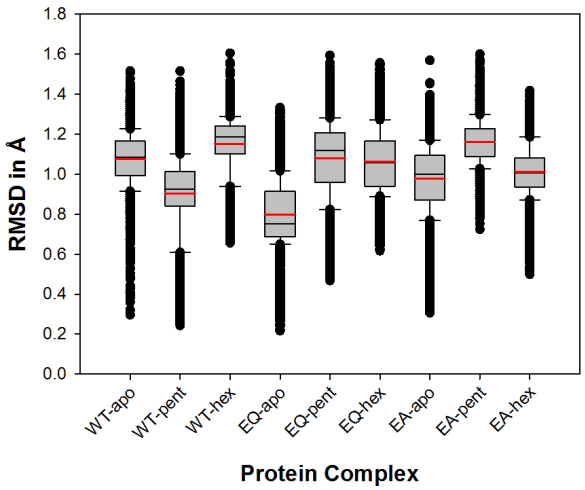

Fig. 45: Flexibility of the glucose binding groove lid. Time dependent fluctuations of RMSD were calculated with *g_rms* (Gromacs/4.5.5). One nanosecond at the beginning of each simulation was removed to ensure proper equilibration of fluctuations. Red solid lines show the mean of the RMSD. The upper, middle and lower black lines indicate 75 %, median and 25 % of the fluctuation, respectively. Error bars represent one standard deviation and black circle all outliers of the fluctuation.

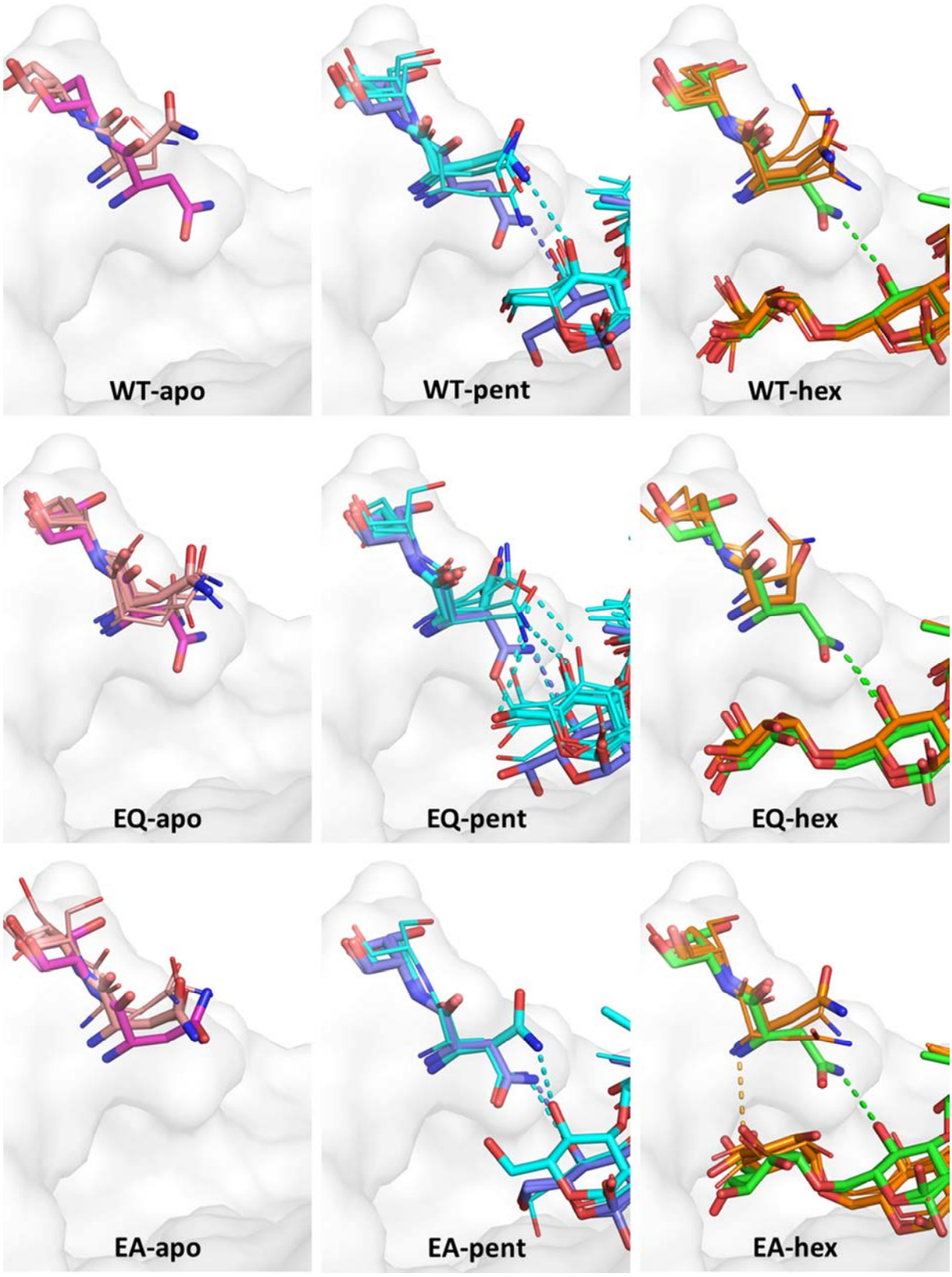

Fig. 46: Glucose binding pocket flexibility from 10 ns simulation with clustered conformations of residues 471-472. Superposition of E372A/Q and WT crystal structures with frames of 10 ns simulations of the named structures with crystal structure depicted in pink for apo, blue for penta- and green for hexasaccharide structures on the protein surface (white). Depicted frames were selected by a cluster analyses with *g_cluster* (Gromacs) and a cutoff of 0.45 Å. Selected frames are colored in rose for apo, cyan for penta- and orange for hexasaccharide simulations. Thick sticks show the highest central populated cluster and the reference crystal structure. Dashed lines represent hydrogen bonds with maximal distance of 3.5 Å. Number of depicted clusters: EA-apo=4; EA-pent=2; EA-hex=3; EQ-apo=6; EQ-pent=7; EQ-hex=2; WT-apo=2; WT-pent=4 and WT-hex=5.

Beside protein flexibility, ligand flexibility within the binding site may also depict interaction dynamics in different regions of the binding site. Therefore the RMSF of the ligand was analyzed equivalently to the binding site residues for every atom (Fig. 47). Hereby a strong increase of the flexibility for the pentasaccharide compared to the hexasaccharide is observed, while this increase is the most apparent for the non-reducing end rhamnose. EQ showed the most flexible ligand and the difference between penta- and hexasaccharide is the strongest, while in EA the least flexible ligand could be found. As exception, Glc2 is most flexible in EA and least flexible in WT.

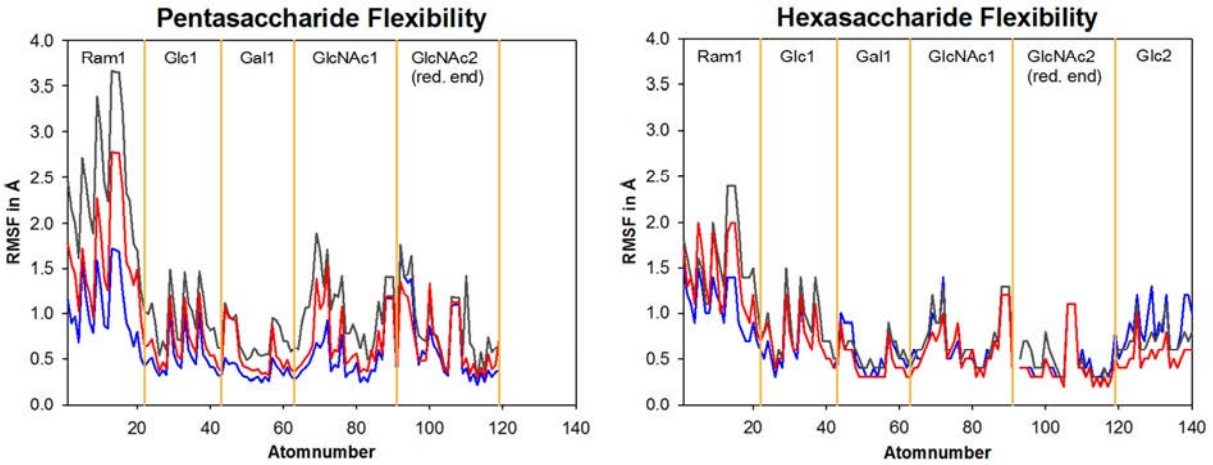

Fig. 47: Flexibility of the ligand in the binding site.
RMSF were derived from 10 ns simulations with *g_rmsf* (Gromacs) for all atoms of the ligand. Atoms were sorted equivalently for penta- and hexasaccharide. Solid lines are colored red for WT, gray for E372Q and blue for E372A. Yellow vertical lines mark the border between to monosaccharide building blocks.

These observations can be visually confirmed by a cluster analysis of the ligand (Fig. 48). The crystal structure of WT-hexasaccharide and the simulated central position overlay well. The WT-pentasaccharide does not overlay well at RAM1 and NAG2. Also in agreement with the high RMSF, the pentasaccharide bound in EQ exhibits the most conformations especially in the region of RAM1 and GLC1, while the hexasaccharide in EQ is only strongly varying at RAM1. The pentasaccharide in EA is in a good agreement with the crystal structure as only three additional clusters could be calculated with a rotation of RAM1. The EA-hexasaccharide conformations are the least overlaid at GLC2 and RAM1. In total this showed the highest flexibility of the pentasaccharide in the EQ and WT, which is strongly reduced for the hexasaccharide. Almost equivalent ligand flexibilities occur in E372A.

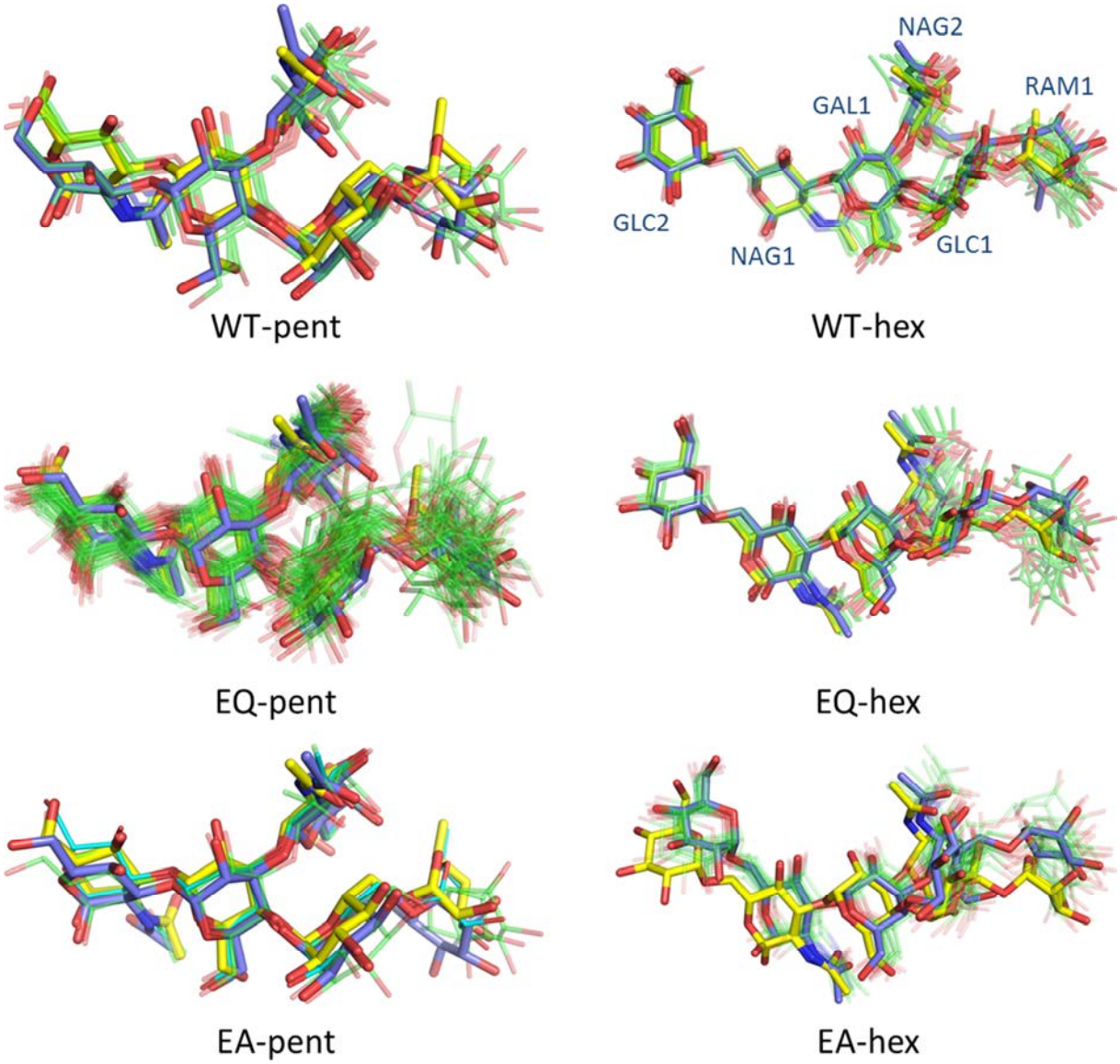

Fig. 48: Ligand flexibility in 10 ns simulations with cluster derived ligand conformations.
E372A/Q and WT crystal ligand conformers are superimposed with frames of 10 ns simulations of the corresponding ligand with the crystal ligand in yellow. Depicted frames were selected by cluster analyses with *g_cluster* (Gromacs) and a cutoff of 0.48 Å. The central position of the main populated cluster is indicated by blue thick sticks. Green transparent thin sticks show low populated clusters. Number of depicted clusters: EA-pent=4; EA-hex=15; EQ-pent=46; EQ-hex=13; WT-pent=5 and WT-hex=14.

3.4.2.2 Methodology of water molecule position determination on protein surfaces

The flexibility of oligosaccharide and binding site side chain residues derived from MD simulation helped to obtain a more detailed view on the binding behavior of HK620 TSP WT and EA/EQ. However, the thermodynamic contributions of complex formation could not be explained without the consideration of explicit water molecules. Three methods were used for the eluciation of explicit water molecule positions on the protein surface by MD simulation.

The *MobyWat* tool was used for a mobility based prediction of the hydration structure (Jeszenői et al., 2015). The algorithm of the program uses either the identity or the position of

water molecules for the generation of water pool lists. The resulting water positions on the protein surface can be rated by the success rate SR_x. The success rate indicates the number of water positions which agree with the crystal structure water positions (cutoff 1.5 Å), divided by the total number of crystallographic water positions in percent. The developers of the program showed that they achieved a reasonable success rate above 80 % with 10 ns simulations (Jeszenői et al., 2015). Simulations were analyzed in the prediction mode and the four derived lists were compared with the crystal structures: all-inclusive identity based (IDa), elitist identity based (IDe), position based (Pos) and merged list (Mer). The overall achieved success rate was evaluated in comparison with the agreement rate. The agreement rate describes the number of MD derived water positions agreeing with a crystallographic water position (cutoff 1.5 Å) divided by the total number of MD derived water positions in percent. The best compromise of a good success and agreement rate was found at the top 25 % of the list for HK620 TSP (Fig. 49A). Additionally as already suggested (Jeszenői et al., 2015), the identity based lists achieved the highest success rates with up to 86 % of the 25 % top list (Fig. 49B).

The water molecule network derived by *MobyWat* includes a high number of relatively mobile water positions. The water positions might be occupied during the whole simulation even though there is a high exchange of different water molecules. Thus, two additional methods were used to determine more stable or so called conserved water molecules by elongated 100 ns simulations of E372A.

For the end point method, trajectories were treated with *trjconv -nojump* (Gromacs/4.5.5) leading to a trajectory that is reduced every frame by every molecule that leaves the box of the periodic boundary conditions. As a result, the last frame of the simulation only includes molecules, which did not leave the simulation box. Thus, highly conserved water molecules remain on the protein surface. This method was used on Sf6 TSP (Kang et al., 2016), which resulted in a good agreement of crystal and simulated water positions.

For the hydrogen bond based method, positions of water molecules were located where a hydrogen bond between binding site amino acid residues and water molecules was formed with occupancies higher than 2 %. At this position the exchange of water would be monitored, which was quantified by the Q-factor defining the number of water indices on a given position divided by the number of frames that the position was occupied by a water molecule. Finally the averaged position of these different occupied frames was calculated.

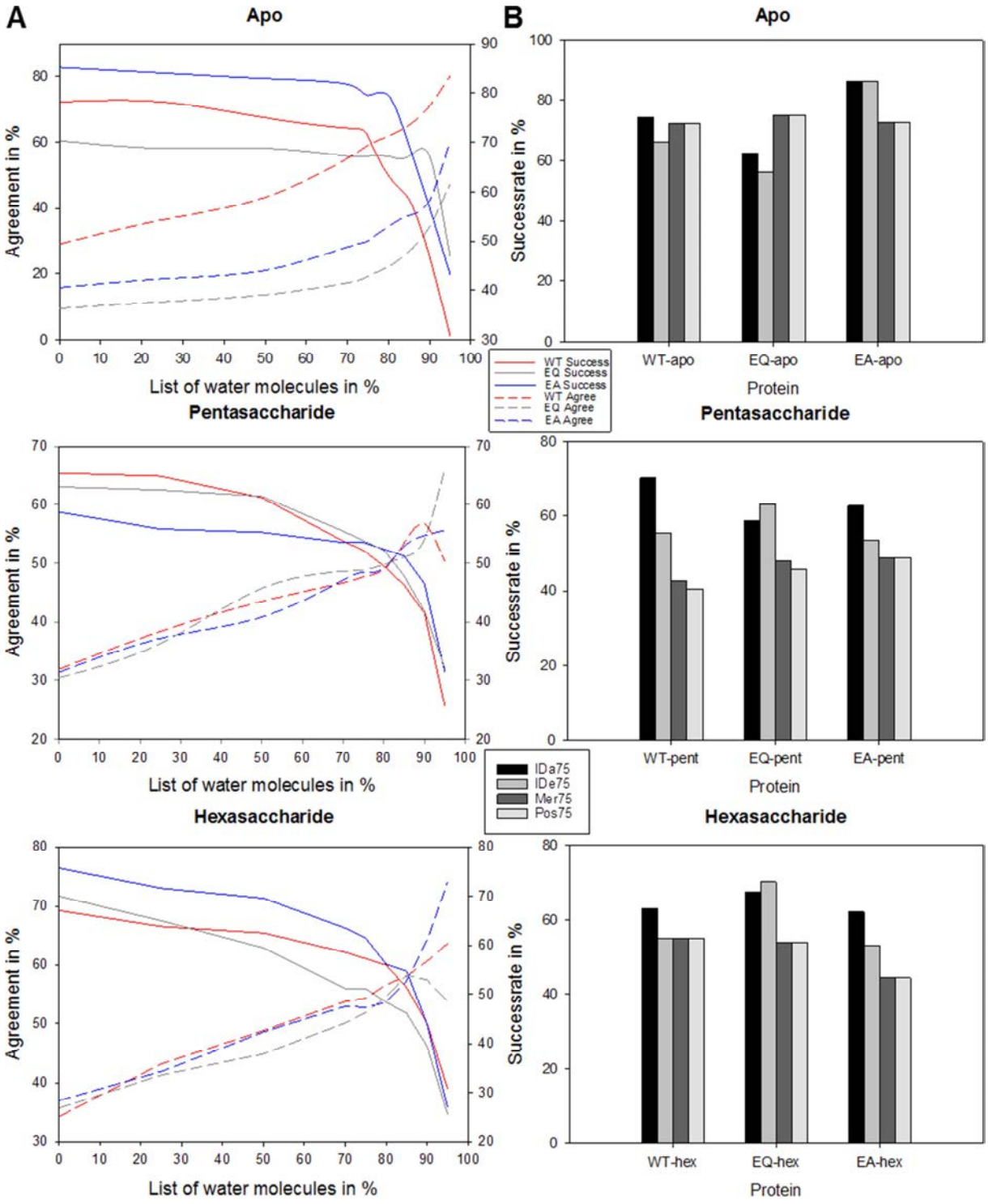

Fig. 49: Validation of success and agreement rate by *MobyWat* derived water positions.
(A) Agreement rates in % (dashed lines, WT=red, E372A=blue, E372Q=gray) derived by the quotient of the number of reproduced crystal water positions and the total number of produced water positions by *MobyWat*. Success rates in % (solid lines, WT=red, E372A=blue, E372Q=gray) derived by the quotient of number of reproduced crystal water positions by *MobyWat* and the total number of crystal water positions in the binding site. Water molecules were automatically sorted by *MobyWat* depending on their mobility. Depicted values are the mean of four derived rates of IDa (all identity), IDe (elitist identity), Pos (position) and Mer (merged) based lists. **(B)** Bar diagram illustrates success rates from the four lists derived by *MobyWat* of the top 25 % of water molecules in the list. Top 25 % corresponds to 75 % on the left.

3.4.2.3 Assignment of water networks and conserved water molecules by MD simulation

The water networks derived by *MobyWat* span the entire binding site (Fig. 50). Thus, the extension of the networks decreases with increasing ligand size. The largest network is observed in the WT-apo simulation. Here, the network is interrupted between the upper and central parts of the binding site. This disconnection is not observed for the high affinity mutants EA and EQ. With binding of the pentasaccharide this network gap is extended for all protein variants. In the WT-pentasaccharide complex, a more extended network is apparent within the glucose binding groove compared to the high affinity mutants. In the presence of the hexasaccharide, the networks of WT and EQ become smaller and only span the area at the entrance of the glucose binding groove, while for E372A the network remains the most extended.

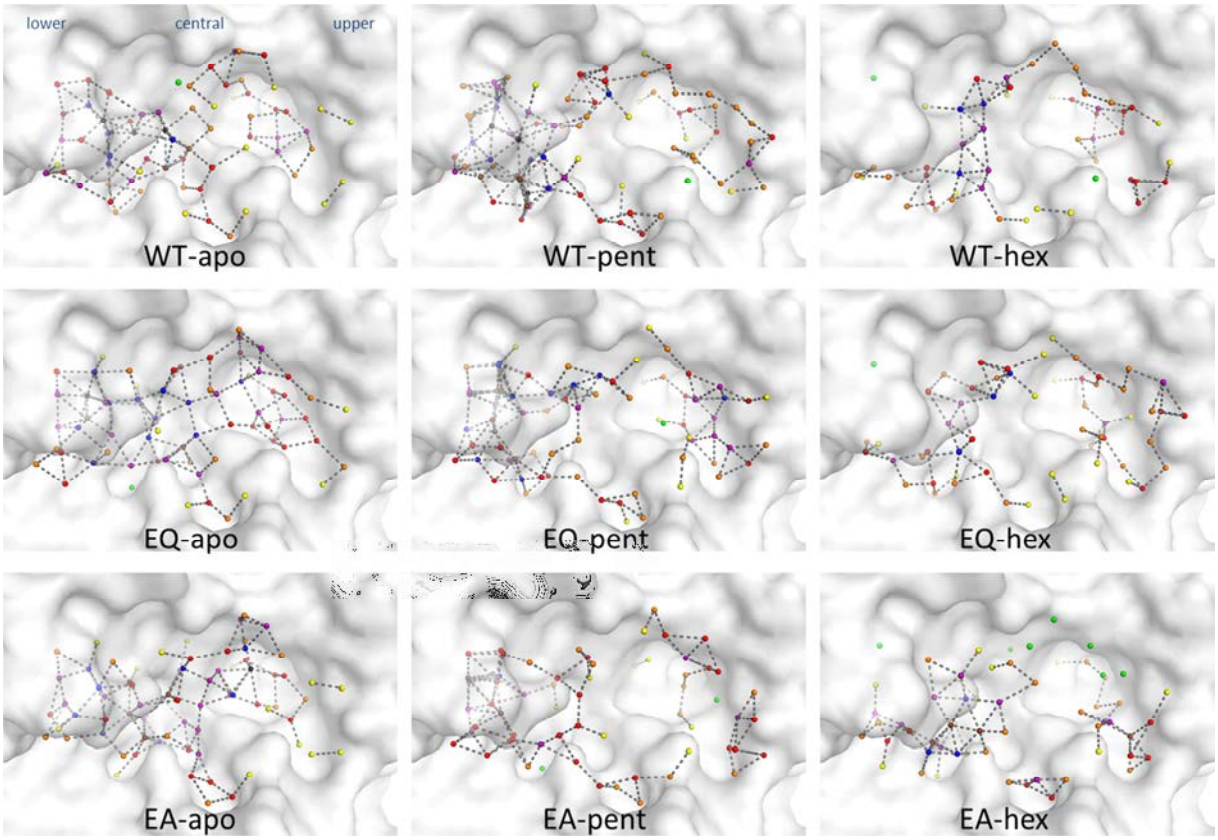

Fig. 50: Potential water network from 10 ns simulations derived by *MobyWat*
Binding site protein surface (white) with hydrogen bonds between water positions from the top 25 % IDa list are represented in gray dashed lines with maximum 3.5 Å distance. Water molecules (spheres) were colored depending on the number of hydrogen bonds: 0=green, 1=yellow, 2=orange, 3=red, 4=purple, 5=blue, 6=brown, 7=black.

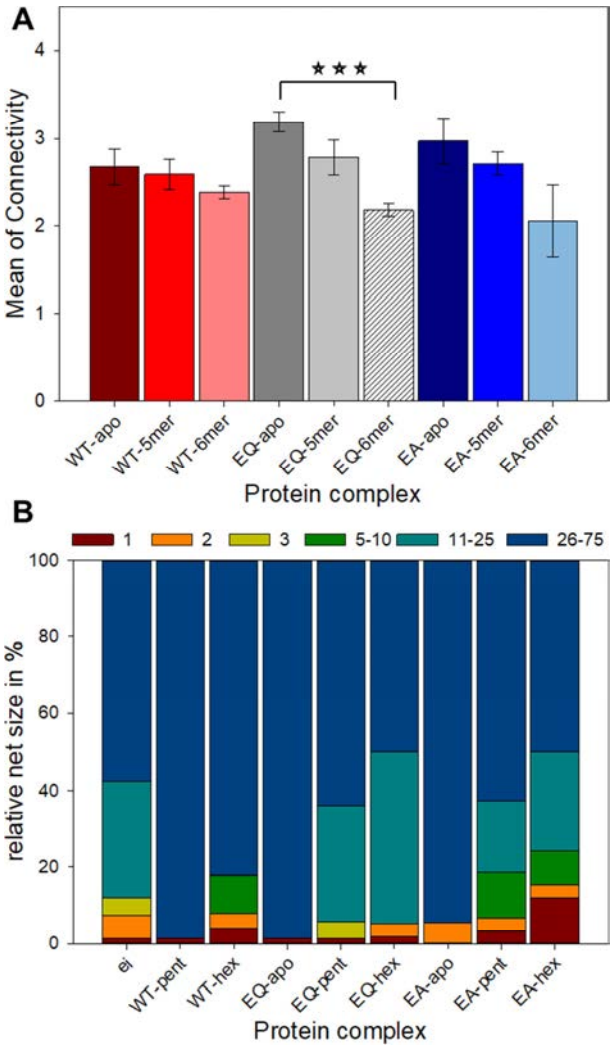

Fig. 51: Connectivity and net size of water networks obtained in MD simulations.
(A) Mean of connectivity was determined by the maxima of Gaussian curves fitted on the distribution of the number of connecting water molecules from derived water networks. Error bars depict one standard deviation of the Gaussian curves. Statistical differences are indicated by stars and were evaluated by Holm-Sidak-Test with P≤0.001 and an overall significance level of 0.05. **(B)** Continuous net size in percent calculated from 10 ns simulations with number of water molecules in the continuous net: 1=red; 2=orange; 3=light green; 5-10=dark green; 11-25=cyan; 26-75=blue.

The quantitative evaluation of networks is generally complicated, therefore two parameters were chosen for a comparison: the connectivity and the continouos net size. The mean connectivity was determined by counting the number of hydrogen bond partners for each water position of the network (as colored in Fig. 50). The distribution of the number of hydrogen bond partners was fitted using a Gaussian curve, in which the maximum of the curve describes the mean connectivity of the network (Fig. 51A). From previous visual inspection it is clear that the mean connectivity decreases with increasing ligand size from apo to hexamer for all three protein variants. However, only HK620 TSP E372Q showed a statistically significant difference.

The continuous net size was determined by the number of water positions within a continuous network without interruptions (Fig. 51B). The size of the net decreases with increasing ligand size for the high affinity mutants EA/Q whereas for WT the smallest net is found for apo and the largest net for the pentasaccharide bound form. In conclusion, the *MobyWat* analysis provided water networks spanning over the entire binding site to different extents among HK620 TSP variants.

This mobile network has an impact on the binding site hydration but thermodynamics of specific water molecules cannot be deduced so easily from it. E372A lacks the charged carboxylate group found in the wild type central binding site. Thus binding thermodynamics in E372A is possibly influenced differently by water molecules than for the other HK620 TSP mutants.

With the end point based method no conserved water molecule positions were obtained in simulations of the E372A apo structures and only up to 4 water molecules could be stably located on the protein surface binding site in the presence of the ligand (Tab. 13). Thus, this method alone was not as suitable for the determination of conserved water molecule positions on HK620 TSP E372A.

Next, the hydrogen bond based method was employed to find more mobile but still conserved positions in the binding site. Here, the number of stable occupied water positions was on average around twelve positions in the apo structure (Tab. 13). For pentasaccharide complex simulations the number of positions increased for all three water models. In contrast, for hexasaccharide simulations the number of water molecules is reduced in comparison to the pentasaccharide complexes. TIP5P has been shown to be less mobile than TIP4P and -3P in a carbohydrate binding site (Fadda and Woods, 2011). This observation cannot be drawn with consistency from the obtained data obtained here.

Tab. 13: Number of conserved HK620 TSP binding site water positions

Protein	Crystal	End point method			Hydrogen bond method			Conserved	Unique
		TIP3P	TIP4P	TIP5P	TIP3P	TIP4P	TIP5P		
EA-apo	24	0	0	0	12	11	12	9	3
EA-pent	43	0	1	2	25	19	23	22	15
EA-hex	44	4	4	2	14	14	20	19	13

The end point and hydrogen bond based water positions from simulations using the TIP3P, TIP4P and TIP5P water models, along with all positions experimentally defined in the crystal structure, resulted in seven different sets of water positions for each binding state of E372A: apo, penta- and hexasaccharide (Fig. 52). In overlays of these seven sets, water positions were counted as conserved when at least two water molecules from different sets were found within a distance of 1.5 Å (Tab. 13). This resulted in nine conserved water positions for EA-apo, whereof three positions are replaced by ligand binding (Fig. 52A). For the pentasaccharide bound state 22 conserved positions were found, with more than the half of the positions not present in the apo form (Fig. 52B). For the hexasaccharide the same effect is observed as for the pentasaccharide (Fig. 52C). There is an increase of conserved positions upon ligand binding. Two additional conserved positions were found for penta- compared to hexasaccharide as Glc2 replaces these water molecules in the glucose binding groove. In conclusion, the *MobyWat* analysis of mobile surface water molecules showed a decrease of water molecules with the increasing size of the ligand, while the analysis of conserved surface water molecules by end point and hydrogen bond based method showed an increase of conserved water molecules with the increasing size of the ligand.

The two water molecules found at the mutated A372 upon ligand binding could be reproduced in simulations as well as the connecting water in front of the glucose binding groove. These water molecules were observed in the crystal structures of E372A mutants. Thus, simulations confirm the exchange of glucose binding groove water molecules with the residual binding site.

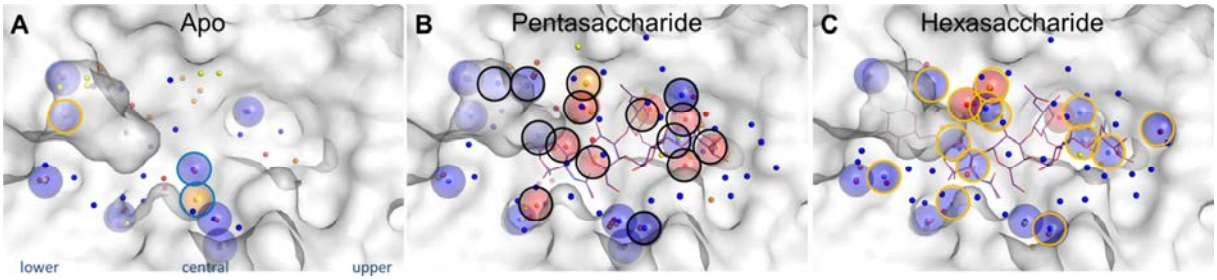

Fig. 52: Comparison of conserved water positions in the HK620 TSP E372A oligosaccharide binding site from crystal structure and MD simulations
The binding site protein surface (white) is shown with crystal water positions as small spheres (blue). Hydrogen bond based analysis led to water positions as small spheres for TIP3P (yellow), TIP4P (orange) and TIP5P (red). End point analysis positions are indicated as small spheres for TIP3P (light gray), TIP4P (middle gray) and TIP5P (dark gray). Large transparent spheres indicate positions of conserved water molecule positions which are occupied by at least two water molecules when analyzed with hydrogen bond or end point based methods from simulations or with crystal structure based information. **(A)** E372A-apo has marked water positions with a yellow frame for a position replaced by the hexasaccharide ligand and blue framed positions, which are replaced by both ligands. **(B)** E372A-pent with black framed conserved water positions which are immobilized upon binding of the pentasaccharide (purples sticks) (not present in apo). **(C)** E372A-hex with the ligand (purple sticks) and yellow framed conserved water positions, which are not present in the apo form.

3.4.3 Free energy calculations of water molecule displacement from the glucose binding pocket

Using the last two methodologies water networks and conserved water positions could be assigned, which is a useful tool to evaluate the influence of the water molecules in the hydrophobic glucose binding groove on the thermodynamic contribution of penta- and hexasaccharide binding to HK620 TSP. The number of observed crystal water molecules in the glucose binding groove of apo, penta- and hexasaccharide were numbered depending on their positions from 0 to 4 (Fig. 53). As already described, upon hexamer binding, water molecules 1-4 are removed and only water 0 remains in the binding groove in every HK620 TSP variant (Fig. 53C). However, not every apo or pentasaccharide crystal structure includes the same number of water molecules. From the apo structures only WT included the described five water molecules, while the high affinity mutants only contain three (EQ position 0, 1 and 3) or two water molecules (EA position 0 and 1) (Fig. 53A). For pentasaccharide structures, E372A has an additional sixth water molecule in the glucose binding groove (Fig. 53B indicated by the label “5”).

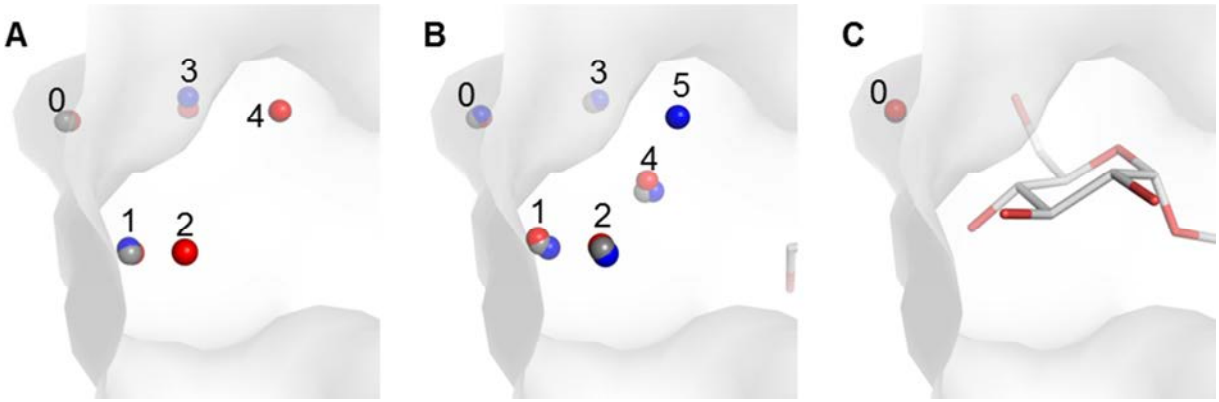

Fig. 53: Comparison of number of water molecules in the glucose binding groove.
Crystallographic water molecules in the glucose binding pocket of HK620 TSP WT (red), E372A (blue) and E372Q (grey) from apo **(A)**, penta- **(B)** and hexasaccharide structures **(C)**. Water molecules are numbered according to their position. The respective ligand is shown as white sticks.

The determination of conserved water molecules by simulation derived two water positions for E372A-apo, as in the crystal structure. However, in the pentasaccharide form only three of six crystal positions are identified in simulation. Water molecules derived by the hydrogen bond based method were quantified with the Q-factor, with small Q-factors indicating more stably bound water molecules. The higher the Q-factor, the more exchanges at this position occurred. The Q-factors for the water molecules at position 0 from E372A-hexasaccharide complex simulations resulted in a value of 0.1 for all water models, while in complex with the pentasaccharide the Q-factor increased to 0.7 (TIP5P) and 3.92 (TIP4P) and without ligand to 0.81 (TIP5P) and 6.92 (TIP3P). Position 1 in the glucose binding groove is occupied in apo with 53.8 (TIP3P) and in pentamer with 18.2 (TIP4P). For position 3 in the pentasaccharide complex simulation, a water molecule with a Q-factor of 13.7 (TIP3P) and 20.3 (TIP4P) was found. These quantified positions emphasize that the stability and conservation of water molecules in the glucose binding groove of E372A-complex simulations decreases with decreasing size of the ligand. Furthermore, the deeply buried position 0 in the groove was more stably occupied than the more accessible positions 1 and 3.

In contrast, in the mobility based analysis all five positions known from the crystal structure were correctly predicted. Thus the number of water molecules in the glucose binding groove strongly varies and the impact in the complex binding thermodynamics of each single water molecule cannot be concluded so far. To evaluate the thermodynamic influence and the probable existence of the glucose binding groove water molecules, free energy (FE) calculations as a statistical mechanics approach, which calculates chemical potentials in solution via partition functions, was used on the described and numbered crystal water molecules (Fig. 53) for comparison with experimental ITC data (Fig. 54A).

Tab. 14: Free energy calculation data and water root mean square displacements (RMSD) during 1 ns simulation. The depicted crystal structures with crystallographic glucose groove water molecules were simulated for 1 ns for RMSD calculations. Free energy calculations were performed individually for each water molecule. Water molecules from simulation in the glucose binding groove are in italic. Position 5 of E372A-pentasaccharide is shown in blue.

Crystal structure	Water molecule residue[1]	RMSD in Å	$\Delta\Delta G_{FE}$ in kJ/mol	$\Delta\Delta G_{FE}$ sum in kJ/mol
WT-apo (2VJI)	0 = 1245	3.706	2.44±1.1	-28.65±3.59
	1 = 1196	12.05	-3.33±0.67	
	2 = 1230	34.92	-12.17±0.59	
	3 = 1199	19.21	-8.14±0.73	
	4 = 1233	19.58	-7.44±0.5	
WT-pent (4PAT)	0 = 677	0.430	13.8±1.29	60.46±6.01
	1 = 800	2.241	6.89±1.12	
	2 = 927	1.724	10.17±0.71	
	3 = 787	0.614	16.79±1.78	
	4 = 1023	0.879	12.82±1.12	
WT-hex (2X85)	0 = 1010	0.570	16.3±3.56	16.3±1.0
E372Q-apo (4AVZ)	0 = 705[2]	2.085	8.04±2.6	-0.60±3.11
	1 = 945	23.16	-8.64±0.51	*-1.99±5.39*
	2 = 5825	*2.373*	*7.36±1.08*	
	3 = 5809	*21.64*	*-7.38±0.66*	
	4 = 5795	*9.724*	*-1.37±0.54*	
E372Q-pent (4XOT)	0 = 691	0.497	12.95±1.87	21.08±6.01
	1 = 820	7.371	2.10±0.97	
	2 = 1002	1.795	8.41±1.24	
	3 = 759	28.56	-9.63±0.8	
	4 = 1167	2.593	7.25±1.14	
E372Q-hex (2X6W)	0 = 1296	0.548	5.43±0.63	5.43±0.63
E372A-apo (4XN0)	0 = 677	14.94	-3.70±0.74	15.89±2.08
	1 = 833	1.571	10.16±0.61	*9.56±3.1*
	2 = 5805	*12.14*	*-2.13*±0.51	
	3 = 812	1.054	9.43±0.74	
	4 = 5832	*10.59*	*-4.21*±0.51	
E372A-pent (4XM3)	0 = 775	8.215	-3.92±0.86	11.84±4.99
	1 = 720	1.722	8.72±1.07	4.21±5.59
	2 = 737	14.06	-4.15±0.55	
	3 = 723	1.174	12.11±1.5	
	4 = 1028	20.51	-0.93±1.02	
	5 = 805	10.12	-7.62±0.6	
E372A-hex (4XN3)	0 = 628	0.325	5.50±0.69	5.50±0.69

1 Crystal water molecules are named by their residue number and simulated water molecules are named by their resID.
2 The value for this water molecule was averaged from five FE calculations each for the test of convergence.

The free energy difference upon displacement of an individual water molecule from the binding site was calculated by the double decoupling method (Tab. 14) (Hamelberg and McCammon, 2004) with no restrictions on any other water molecules. Then, for each complex type all $\Delta\Delta G_{FE}$ were summed up (Fig. 54B). The water molecule positions which were not occupied in all of the three protein variants with apo, penta- and hexasaccharide were added in the simulations and additional individual FE calculations were performed on these water molecules (Tab. 14 and Fig. 54B). The derived sums of free energy difference did not change within the error of the overall result. The error of the individual $\Delta\Delta G_{FE}$ on each water molecule was estimated by five repetitions of the calculations on water position 0 of E372Q-apo, which resulted in an error of ±3 kJ/mol for a set-up with restrained backbone and flexible glucose binding groove loop (Fig. S 14). In conclusion, the crystal structures include all relevant water molecules for the binding thermodynamics in the glucose binding groove.

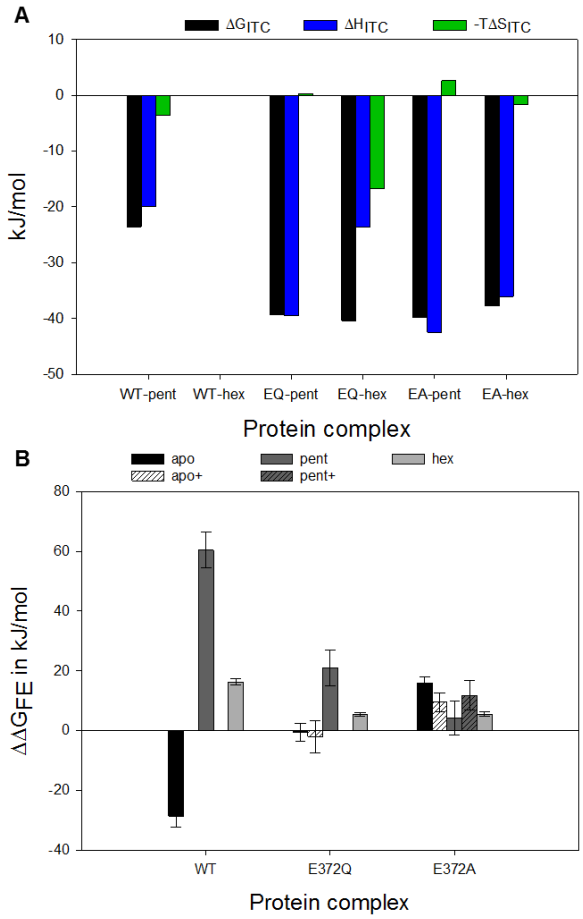

Fig. 54: Free energy calculations on crystallographic water molecules in the glucose binding groove
(A) Results of ITC measurements as shown in Fig. 37. **(B)** FE calculation data with the summed up free energy differences of all crystallographic water molecules found in the glucose binding groove for apo (black), penta- (dark gray) and hexasaccharide (light gray). The individual result for each water molecule is listed in Tab. 14. Calculated free energy differences with additional water molecules with the corrected $\Delta\Delta G_{FE}$ sums are included in bars for HK620 TSP complexes without ligand (apo+: white striped) and with pentasaccharide (pent+: gray striped).

For the interpretation of the calculated $\Delta\Delta G_{FE}$ sums, it has to be considered that the values are not directly transferable on $\Delta\Delta G_{ITC}$ values and more likely depict a qualitative estimation. A $\Delta\Delta G_{FE}$ value stands for the difference between the energy needed to remove a water molecule from the bulk versus the energy needed to remove the water molecule from the glucose binding groove. Therefore, a $\Delta\Delta G_{FE}<0$ indicates that more energy is needed to remove these water molecules from the protein surface than would be needed to remove it from the bulk. This results in favorable water positions. In contrast, a $\Delta\Delta G_{FE}>0$ depicts unfavorable water positions, as these water molecules would easily diffuse from the protein surface into the bulk and have to be held to stay on the surface. Thus, a harmonic potential is applied on the water molecule to enable a stable occupied position on the protein surface and this force contributes as a positive value to $\Delta\Delta G_{FE}$ (Chapter 2.7) (Hamelberg and McCammon, 2004).

The calculated free energy differences for water displacement from the glucose binding groove strongly vary between WT and high affinity mutants (Fig. 54B). While WT-apo has a negative free energy, EQ has a value around zero or EA is strongly positive in the apo form. For the pentasaccharide bound form $\Delta\Delta G_{FE}$ is positive for the three protein variants with the highest value for WT, decreased for EQ and the lowest for EA. In the hexasaccharide bound form the free energy differences are positive for WT with 16 kJ/mol and 5.5 kJ/mol for the high affinity mutants. The comparison of the calculated free energy differences (Fig. 54B) and the simulated water network structures (Fig. 50) with experimentally measured thermodynamics (Fig. 54A) can therefore be interpreted with the following conclusions:

For the apo-structures of HK620 TSP, the crystal structure and *MobyWat* analysis of WT-apo have the most extended water network from all the protein variants. Furthermore, the water molecules in the glucose binding groove are relatively tightly bound. Thus it is likely to cost enthalpy to break these bonds and the free binding enthalpy cannot be determined in ITC experiments due to a lack of heat release upon the binding event (Freire et al., 1990). In EQ-apo the glucose binding groove water is almost "bulk-like"; therefore the entropy is favorable to release this water. However, the enthalpy in EQ-apo is slightly unfavorable. In EA-apo the water in the glucose binding groove is highly dynamic, thus no enthalpic cost occurs upon ligand binding but also no entropic gain.

For the pentasaccharide complexes of HK620 TSP, *MobyWat* analysis depicts the largest gap in front of the glucose binding groove for WT-pent compared to the high-affinity mutants, which show a connected water chain in this area. For WT-pent, it requires a lot of energy to disconnect the glucose binding groove water from the residual water binding site network, which might be highly influenced by the charges on E372. This separation presumably also leads to unfavorable entropy. In EQ-pent, the separation of the water from the residual network by the pentasaccharide ligand costs less energy as the water is not as fixed, but still an entropic cost is connected with the separation. In EA-pent, the bulk-like water might still exchange freely with the binding site water network, as a connecting water chain can be seen in crystal structures and *MobyWat* analysis and therefore neither enthalpy nor entropy are affected by these water molecules.

For the hexasaccharide complexes of HK620 TSP, the same environment is given for the residual water molecule in the glucose binding groove in WT, as in the high affinity mutants EQ and EA. Thus, the $\Delta\Delta G_{FE}$ also results in the same value for EQ and EA, while it is minimally higher for WT. This may be due to the additional charge in the WT binding site.

In conclusion the free energy calculations delivered a better understanding and interpretation of the experimental data together with the water network distributions obtained for MD simulations.

4 Discussion

4.1 Sf6 tailspike protein as a specific binder of *Shigella flexneri* O-antigen

The complex formation of the inactive Sf6 TSP_{EADA} to *S. flexneri* Y polysaccharide was shown by fluorescence spectroscopy. The affinity was predominantly generated by the formation of direct hydrogen bonds between protein and *RU*1 and *RU*3 of the ligand and water replacement as suggested by crystal structure and MD simulations. These findings nicely agreed between the simulated Sf6 TSP*wt* and Sf6 TSP_{EADA}-octasaccharide complexes. However, the dodecasaccharide in complex with Sf6 TSP showed different conformers in simulations with either the wild type or the EADA mutant. Whereas it was straight and elongated in the binding groove in Sf6 TSP*wt,* a bended conformation was found in complex with Sf6 TSP_{EADA}.

In the wild type complex, to achieve the straight dodecasaccharide conformation, the protein backbone loops had to be kept flexible and the third repeating unit of the dodecasaccharide was rotated into a high-energy starting conformation to avoid clashes within the protein (Kang et al., 2016). However, exactly this high-energy conformer from Sf6 TSP*wt* did not result in a stably bound position in Sf6 TSP_{EADA}. The alanine residues at the mutated position 366 and 399 open up space in the binding groove and it would be worth a try to fit the non-rotated low-energy conformer as an alternative starting structure into the Sf6 TSP_{EADA} binding site prior to simulation. SPR interaction experiments were performed on the same surfaces as described in this work with Sf6 TSP single point mutants like E366A, E366Q, D399N, D399A and the double mutant EQDN (data not shown). All these mutants showed strongly decreased binding signals compared to Sf6 TSP_{EADA}. This would suggest that the binding affinity of Sf6 TSP_{EADA} is already increased compared to Sf6 TSP*wt* and the single point mutations and EQDN. Apparently, in EADA this rotation might not be necessary and due to that the binding affinity is increased as the ligand can bind as a low energy conformation (Ahmad et al., 2016).

The high-energy conformer of SfY dodecasaccharide might be necessary for the glycosidase activity of Sf6 TSP*wt*. Glycosidase activity is classified into configuration retaining and inverting mechanisms (Koshland, 1953). Both mechanisms require an oxacarbenium-ion-like transition state and two carboxylic acids. For an inverting mechanism, these residues must have a distance around 10 Å as a water molecule is included in the transition state. In the

retaining mechanism, the carboxylic acids are closer to each other with around 5.5 Å and the water molecule is only involved after the formation of the oxacarbenium adduct (Rye and Withers, 2000). In Sf6 TSP*wt* the distance between the two active site carboxylate residues with 7.8 Å lies between these two distances. Nevertheless a recent paper showed that the distance of inverting enzymes is lower at around 8±2 Å (Mhlongo et al., 2014), which would classify Sf6 TSP as inverting α-glycosidase, which is generally known for α-L-rhamnosidases (Zverlov et al., 2000). Glycosidase mechanisms use distortions of the substrate (Miyazaki et al., 2016) or hydrogen bonding to the sugar-2-hydroxyl group to stabilize the transition state with a contribution of 8-10 kcal/mol (Rye and Withers, 2000). This would exactly reflect the energy needed for the high-energy conformer, which was used to fit the dodecasaccharide in the Sf6 TSP binding groove. However, this is based on speculation and studies with reaction intermediates like fluorinated or deoxy-oligosaccharides (Rye and Withers, 2000) and QM-MM simulations (Mhlongo et al., 2014; Pan et al., 2013) would be needed.

The binding affinity of the Sf6 TSP_{EADA}-polysaccharide complex was quantified by SPR experiments as a bivalent heterogeneous binding process on immobilized polysaccharide with two dissociation constants of ~0.01 and ~10 µM. A curved Scatchard plot was derived which has been described for multivalent binding interactions (Häyrinen et al., 2002; MacKenzie et al., 1996). The bivalent and the heterogeneous binding model, which are provided by SPR evaluation software (BIAevaluation3.1 or Tracedrawer1.7), describe a binding process in solution. However, the binding process on a surface involves geometrical parameters and is influenced by the ligand density of the surface (Müller et al., 1998). In this model, the reaction volume V_R and a probability factor for the multivalent binding event $multifac_i$ are included, which results in four differential equations describing the monovalent and the bivalent binding state for the first ligand and the bivalent binding state of the second ligand. The model has been successfully used for the trivalent S-layer protein SbsB, which binds to a carbohydrate coated surface with the secondary cell wall polymer (SCWP) of *Geobacillus stearothermophilus* (Mader et al., 2004). Furthermore it has been extended for a heterogeneous ligand (van Steeg et al., 2016). Therefore, using the Müller model to describe Sf6 TSP binding to polysaccharide coated surfaces could improve the here presented interaction curves to the described bivalent and heterogeneous fits in solution.

Nevertheless, the use of polysaccharide activated by oxidation prior to chip coating is not the optimal solution, as the alteration of the structure and the oxidiation level is hard to control so far. Enzymatic cleavage experiments using Sf6 TSP*wt* and mass spectrometry characterization of oligosaccharides obtained from periodate oxidized SfY polysaccharide are needed for a

structural analysis. For a better control of the number and length of immobilized polysaccharide chains, other immobilization strategies would be considerable for example via biotinylated polysaccharides on streptavidin surfaces (Linman et al., 2008; Nilsson et al., 1995). The more stable amine coupling could be used for example by increasing the PEtN modifications of Kdo using *Shigella flexneri* bacteria cultivated with 5-50 mM Ca^{2+} in the medium for polysaccharide preparations (Kanipes et al., 2001). Furthermore, the immobilization of defined chain lengths could be achieved with synthetic *Shigella flexneri* O-antigen fragments. For example, in this work the Sf2a-icosasaccharide with conjugated primary amino groups was used (Bélot et al., 2004) which are suitable for immobilization by amide coupling (Gedig, 2008).

However, the multivalent binding behavior observed for Sf6 TSP on the activated SfY polysaccharide surface probably resembles the natural Sf6 bacteriophage binding on a LPS covered *Shigella flexneri* surface. Multivalent interactions are known to govern protein-carbohydrate interactions and a lot of equivalent studies in SPR with carbohydrate-binding proteins can be found (Brogioni and Berti, 2014; Häyrinen et al., 2002; Houser et al., 2015; Nakajima, 2001). The bimodal chain length distribution of *Shigella flexneri* LPS further suggests that a heterogeneous ligand surface is presented, which is evolutionary optimized for bacterial virulence (Morona, 2003). Thereby it was observed that Sf6 TSP primarily cleaves the very long-type O-antigen chains (Morona and Bosch, 2003), what would suggest that different binding affinities are exhibited depending on the chain length due to the cluster glycoside or chelate effect (Lundquist and Toone, 2002; Wittmann and Pieters, 2013).

Kinetic analyses with fluorescence spectroscopy of Sf6 TSP_{EADA} determined a dissocation rate of SfY polysaccharide of 0.033/s in solution with a weak indication for a conformational selection binding process. This binding process would fit to the suggestion that Sf6 TSP has a progressive conformational binding model (Kang et al., 2016). Glycosidases have been categorized into three general topologies: pockets or craters, clefts or grooves and tunnels (Davies and Henrissat, 1995). The first are considered to cleave single monosaccharide branches, which can be excluded for an endorhamnosidase like Sf6 TSP. However, Sf6 TSP exhibits a mixture of the two latter topologies. The binding cleft of Sf6 TSP certainly has an open structure like the other tailspike proteins HK620 TSP and P22 TSP (Barbirz et al., 2008). However, in the intersubunit cleft, which distinguishes Sf6 TSP from the other tailspike binding sites, the loops enclose the cleft partly forming a tunnel. Tunnel topologies are known for releasing the product, while they stay firmly bound to the polysaccharide chain for a processive cleavage action. Thereby the loops are the driving force of the enzyme

motion (Davies and Henrissat, 1995). Due to the high flexibility of high-rhamnose containing polysaccharides (Galochkina et al., 2016) the intersubunit cleft with the flexible loops might be necessary to govern the flexibility of the ligand in contrast to the branched O-antigen ligands in HK620 and P22 TSP, which are more easily bound and processed on a β-helix sheet surface (Andres et al., 2013; Zaccheus et al., 2012).

The disconnection of one side of the carbohydrate ligand in carbohydrate-processing enzymes has been observed before (Brás et al., 2008) and was also seen for the dodecasaccharide non-reducing end in simulation with Sf6 TSP_{EADA}. Tyrosines are known to assist in reorientation of carbohydrate hydroxyl-groups in glycosidases and thereby facilitate the sliding of the substrate in the active site cleft (Brás et al., 2008). Tyr400 adjacent to the active site residues might assist in the sliding or contribute to the glycosidase activity (Jongkees and Withers, 2014). The mutation Y400C has a decreased binding affinity eventhough no CH,π-interactions are found in the crystal structure, what might support this hypothesis.

The enzyme processivity could be a prerequisite for the infection process of *S. flexneri* Y by Sf6 bacteriophages. Although six tailspike proteins with three binding sites each resulting in 18 binding sites provide multivalent affinity for the attachment of the bacteriophage on the bacterial surface, the affinity to the dodecasaccharide is still low. Thus the detachment of the phage could occur especially if only short polysaccharide chains are present. In contrast to P22, Sf6 needs the outer membrane proteins A and C (Omp) in the *Shigella flexneri* membrane as a second receptor for the infection process (Parent et al., 2014). After the polysaccharide chains have been shortened by the TSP, subsequently, the secondary receptors take over the role of phage attachment and initiate the further steps in the infection cycle.

Sf6 TSP has proven to be applicable as a *Shigella flexneri* sensor in an absorption based microtiter plate assay and as a fluorescent sensor. The signal observed in ELITA did not vary between the four different *Shigella* strains, in contrast to ELITA tests with a *Salmonella* library, where different strains showed different signal intensities (Schmidt et al., 2016). However, only four strains were tested in this work. Therefore the sample size is not sufficient for a conclusion. Other strategies with whole phages have also been developed for the detection of *Shigella,* for example by a T4-like *Shigella flexneri* phage with an integrated bacterial luciferase gene that transduces a bioluminescent phenotype of infected cells (Schofield et al., 2014). Moreover, it has been shown that *Shigella flexneri* O-antigen structures are involved in high-affinity glycan-glycan interactions with blood group A glycans. Here, the *Shigella flexneri* O-antigen primarily interacts with fucose to provide the contact for bacterial infection of the host (Day et al., 2015). Thus Sf6 TSP could also be used

as active enzymes to cleave *Shigella flexneri* LPS or as enzymatically inactive multivalent anti-adhesive agents that cover the surface of bacteria *in vivo* to prevent pathogen-host interactions as a specific antibody-like compound (Chowers et al., 2007).

The data in previous works as well as in this work showed that Sf6 TSP can bind and cleave *Shigella flexneri* 2a O-antigen with an additional branched RAMI β(1-4)-linked glucose (Simmons and Romanowska, 1987). Sf6 bacteriophage is known to infect *Shigella flexneri* Y and X (Casjens et al., 2004) but not *Shigella flexneri* 2a. In the crystal structure of Sf6 TSP-octasaccharide, the glucose branch on a Sf2a-decasaccharide would point away from the protein binding site and towards the solvent when bound to O4 of RAM1 and RAM5. Thus the glucose branch does not interfere with the binding mode found for octa- and dodecasaccharides. Due to the pucker of RAM5 of the SfY-octasaccharide, the glucose branch on RAM5-O4 could even be buried in an underlying groove (Fig. 55). This might probably add to the affinity, although the example of HK620 TSP showed, that binding of an additional glucose branch did not contribute (see Chapter 4.2).

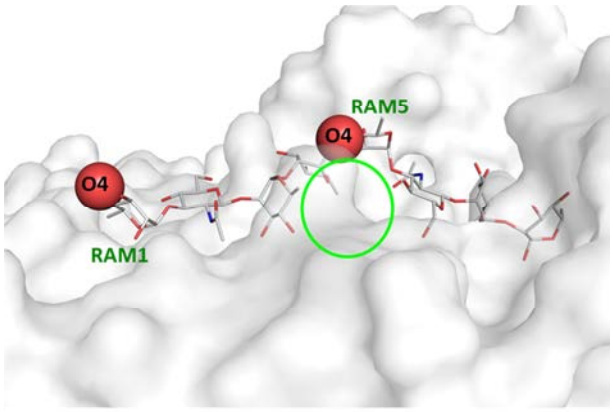

Fig. 55: Sf6 TSP binding site with potential glucose branch binding sites of Sf2a-oligosaccharides
Sf6 TSP_{EADA}-octasaccharide binding site (white surface) with octasaccharide ligand (white sticks) are depicted with the relevant monosaccharide names in green. The hydroxyl oxygens O4 of RAM1 and RAM5 are depicted as red spheres as attachment points of the glucose branches in Sf2a O-antigen repeating units. The green circle indicates the space in the binding site around RAM5.

MALDI-TOF analysis on one of the *S. flexneri* Y isolates showed about 14 % non-stoichiometric acetylation of the O-antigen. O-acetyl transferases for O3,4 O-acetlyation on RAMIII and O-acetlyation on GlcNAc-O6 have been identified, which modify the *S. flexneri* Y O-antigen in a non-stoichiometric manner (Perepelov et al., 2012; Sun et al., 2014; Wang et al., 2014). Sf6 bacteriophage is a serogroup converting phage, which adds an O-acetyl-group to the RAMI-O2 in the repeating unit and changes the serogroup into 3b (Allison and Verma, 2000; Clark et al., 1991; Verma et al., 1991). The linkage of the O-acetyl groups in the *S. flexneri* Y 99-2001 O-antigen could be elucidated by NMR experiments (Perepelov et al., 2012).

The characterization of the inactive Sf6 TSP_{EADA} and the use of the protein for detection of *Shigella flexneri* isolates have brought more insights into the Sf6 tailspike binding mechanism. This knowledge will hopefully be useful for further modifications and application of Sf6 bacteriophages and tailspikes.

4.2 The thermodynamic contributions of water molecules in carbohydrate-protein interactions

The release of water molecules during protein and carbohydrate ligand complex formation is one of the main driving forces of this binding event. However, it is not well understood. Furthermore, crystallographic water positions are under strong discussion as they are highly dependent on resolution of the X-ray structure (Finney and Eley, 1977), crystallization conditions and refinement techniques (Afonine et al., 2013; Badger, 1997). Furthermore most crystal structures of proteins are obtained at 90-120 K in a nitrogen gas cooled beam line (Halle, 2004). Therefore, crystal structures deliver static averaged models which do not include exchange and flexibility of water molecules. Crystallographic water molecules were found favorable or unfavorable for the binding event and their pure appearance in the binding site does not give an impact about their thermodynamic contribution (Beuming et al., 2012).

Thus, 27 crystal structures of HK620 TSP were analyzed based on structural differences of the two binding partners penta- and hexasaccharide and the water molecules involved in the binding. Thereby the high-affinity of the two mutants E372Q and E372A derived from the removal of the glutamate from the binding site. The elimination of the charge increases the affinity as the desolvation is less expensive (Biela et al., 2013). However, the other half derived from either the favorable entropy of the removal of one water upon ligand binding in E372Q (Broeker et al., 2013) or the favorable enthalpic fixation of two water molecules in E372A. Thus, the water molecules resolved in the crystal structure were able to depict the relevant water molecules for the binding event. Moreover these positions were confirmed by MD.

Evaluation of the MD results were obtained for water molecules in the HK620 TSP-complex with three different approaches based on water mobility, hydrogen bonding or water residence time (Jeszenői et al., 2015; Kang et al., 2016).

The MD simulation based tool *MobyWat* assigned a mobile mutation dependent water network to the HK620 TSP binding site, which included a high number of water molecules and a complex hydrogen bond connection in and around the glucose binding groove. Hereby the network complexity was decreased with the increasing size of the ligand. This means that the mobile network is extended over the binding site in the ligand free form. However with the ligand binding event, the mobility of water is decreased as their conservation increases if the mobile and conserved MD-based networks are considered together. Nevertheless the derived networks are complex and the analysis by visual observation and calculation of mean

connectivity and continuous net size is not sufficient to extract the differences of the networks easily. In their more recent paper the developers of *MobyWat* introduced an improved method for the exploration of water networks at target-ligand interfaces which resulted in a high reproducibility of the networks (Jeszenői et al., 2016). Additionally a NetDraw mode is available now that results in a network graph with the mobility of the water assigned to it. The derived networks can be divided into a static and dynamic subnet. Moreover a network can be described by the shortest path length, which together with the mean of connectivity, would maybe characterize water networks as a small world network. With this small world concept the global effects of the binding event could be evaluated by respective algorithms (Taylor, 2013). In case of HK620 TSP-oligosaccharide complexes, the *MobyWat* tool was first used for evaluating calculated positions of protein-carbohydrate interfacial binding water. The tool was successfully applied with a 20-25 % top cut-off with all-inclusive based IDa list to predict a mobile water network on tailspike proteins.

Two additional methods were used to determine conserved water molecules in the binding site of E372A. The endpoint method only identified a low number of highly conserved water positions in the binding site with TIP4P and -5P water. A β-sheet as in the HK620 TSP binding site is a hydrophobic binding surface. This hydrophobicity is even increased upon the mutated alanine (Beuming et al., 2009) and therefore the β-sheet dominated binding site of HK620 TSP is unfavorable for water compared to the loop-based binding site of Sf6 TSP for example (Beuming et al., 2012; Kang et al., 2016).

The usage of hydrogen bond direction for the reproduction of crystallographic water positions is an old idea (Vedani and Huhta, 1991) and has been used in the AcquaAlta tool for docking (Rossato et al., 2011). Furthermore there are more tools like GRID (Goodford, 1985), WaterMap (Abel et al., 2008) or WatSite (Hu and Lill, 2014), which even may give estimates on the thermodynamic contributions of hydration sites (Lazaridis, 1998). This would make the combined use of several tools to determine water networks and free energy calculations redundant. However, in most hydration site determination studies protein-protein and protein-ligand interactions are analyzed in a restrained manner, which does not include conformational changes of receptors or ligands upon binding (Yang and Lill, 2016). Carbohydrate ligands on the contrary are highly flexible and therefore treating receptor and ligand with only protein backbone restraints is more meaningful.

Different numbers of crystallographic and simulated water positions were found enclosed in the glucose binding groove. Free energy calculations of selected water molecules were performed with a double-decoupling approach (Hamelberg and McCammon, 2004). These

resulted in a qualitative agreement of ITC and FE data and helped in developing a model that explains the thermodynamic driving forces, except for a probable overestimation of positive $\Delta\Delta G_{FE}$ values (Barillari et al., 2007) (Fig. 56). HK620 TSP WT-apo had the most crystal water molecules in the glucose binding groove and also the free energy of this water resulted to be negative. This suggested a strongly connected water network in the binding site, which would also be supported by the charged E372 in the central part of the binding site. Upon pentasaccharide binding the glucose binding groove network is interrupted from the residual water in the binding site, which leads to an enthalpic and entropic loss. This enthalpic loss is even increased for the hexasaccharide binding, but the entropic contribution is more favorable. This is in contrast with the ligand flexibility as the pentasaccharide is more flexibly bound than the hexasaccharide in the WT binding site. The same is valid for binding site residues, but not for the glucose binding groove lid. The lid is more open in the pentasaccharide bound form but low in flexibility compared to the hexasaccharide bound form which is closed but as flexible as in the apo form. On one hand, this suggests that the relatively tight water network in WT-apo imposes an enthalpic penalty to desolvation, explaining the overall lower binding free enthalpy compared to the high affinity mutants. On the other hand the conformational space sampled by the glucose binding groove lid and the water in the groove is reduced. This speaks for a relatively high entropic cost compensating the entropy of released water from the overall water network. In the hexasaccharide complex, the enthalpic gain obtained from ligand binding is almost fully compensated for by the release of the water network. Therefore the binding is probably dominantly entropy driven as water is released to a greater extent than for the pentasaccharide and the glucose binding groove lid stays flexible (Fig. 56A).

By the mutation from glutamate to glutamine in E372Q the same side chain geometry without charge is achieved. The water network in and around the glucose binding groove appears to be more loosely bound resulting in a $\Delta\Delta G_{FE} \approx 0$ kJ/mol for the removal of water molecules from the apolar groove. With the pentasaccharide bound, the enclosure and separation of the groove water molecules lead to a low enthalpy loss but potentially the same entropic prize as in the wild type. With the hexasaccharide bound, the enthalpy cost again is less and the same entropic gain is expectable as in the wild type. The flexibilities of the side chains are the same for both ligands including increased and decreased flexibilities upon binding. However, for ligand flexibility a strong difference between penta- and hexasaccharide is apparent with a higher flexibility for the pentasaccharide. The glucose binding groove lid as well is the most flexible for the pentasaccharide bound case. Additionally, the relatively closed apo form

opens up in both cases. In conclusion, the water molecules in the glucose binding groove in E372Q act like in the wild type for penta- and hexasaccharide with the difference that the enthalpic cost is lower due to the missing charge in the central part of the binding site.

The pentasaccharide binding, mostly enthalpically driven, is achieved by replacing a flexible water network with a flexible ligand and almost no changes in the positions of the amino acid side chains occur. Additionally, the glucose binding groove lid is opened up gaining more degrees of freedom compared to the apo state. The entropic gain counterparts the entropic cost of the enclosed water molecules in the glucose groove (Fig. 56B) (Broeker et al., 2013).

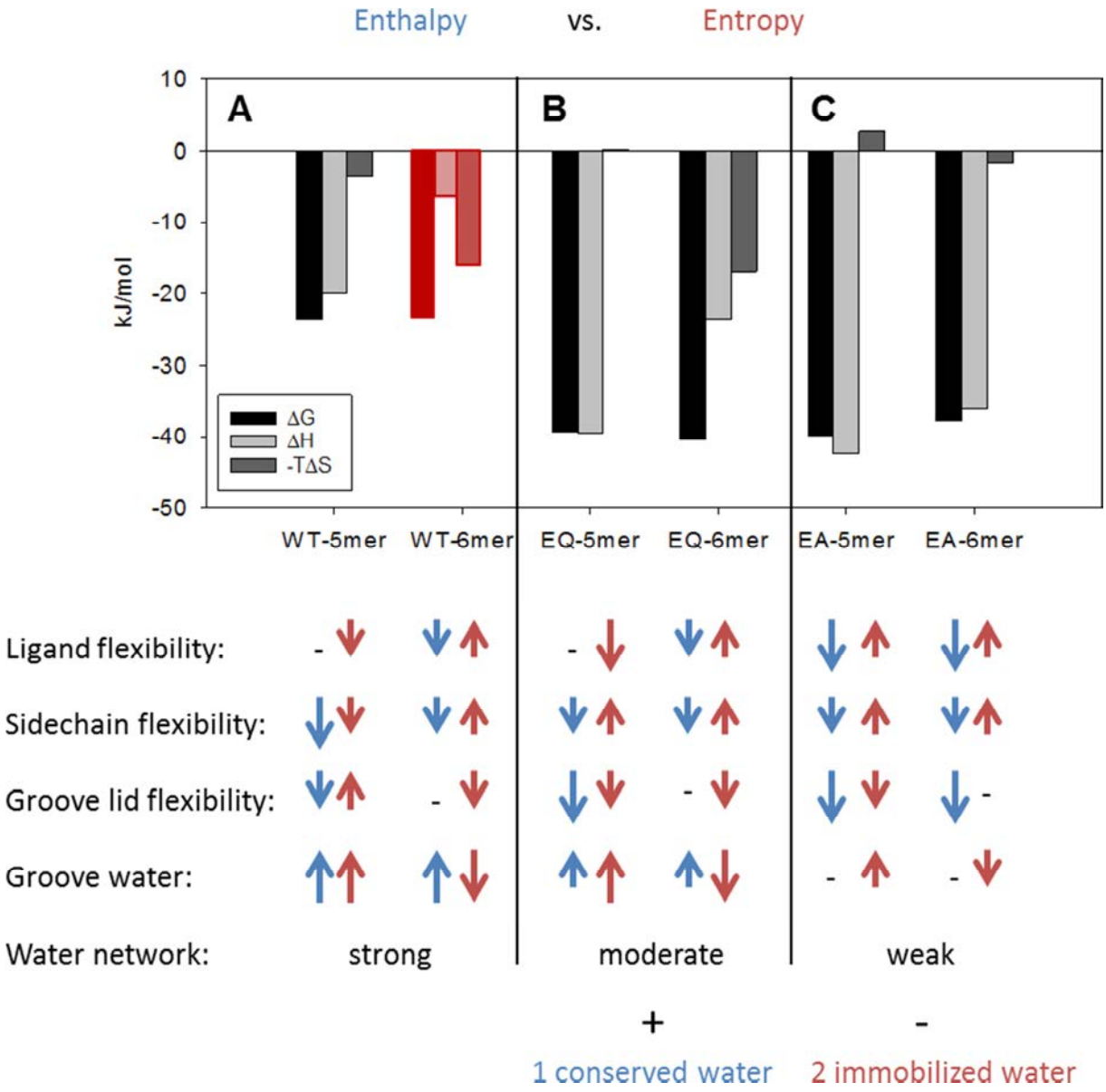

Fig. 56: Scheme for the entropic and enthalpic binding contributions
ITC data are depicted as bar diagrams with a suggested distribution of enthalpy and entropy for WT-hexasaccharide (6mer) in dark red for ΔG, rose for ΔH and red for –TΔS. Arrows explain the contribution of enthalpy (blue) and entropy (red) to the binding event by the named binding partner influences. Arrow up = enthalpic/entropic favorable, Arrow down, enthalpic/entropic unfavorable.

In the E372A mutant charge and geometry of the side chain are altered. The water network in this high-affinity mutant seems to be the most exchangeable. In the penta- and hexasaccharide complex enthalpy and entropy of water molecules in the groove are almost unchanged, as there is probably still an exchange between binding site and glucose binding groove. In the derived *MobyWat* network an additional water molecule can be found in the glucose binding groove next to water 0 in the hexasaccharide complex. Therefore we can conclude that E372A binding affinity is the least dependent on the glucose binding groove water molecules. Nevertheless the water molecules are present in crystal structures, what would emphasize the pre-formation of the glucose binding groove by the water with a high occupancy exchange (Jeszenői et al., 2016; Saraboji et al., 2012). Also the flexibility of the two ligands and the side chains is the same for both ligands. The only difference is observed in an increased flexibility for the glucose binding lid in the pentasaccharide compared to an unchanged flexibility in the hexasaccharide complex compared to apo. The increased flexibility of the lid in the pentasaccharide complex hereby derives from the closing of the lid, which is only observed in simulations in this variant. Thus, the enthalpy of E372A binding comes from replacing a highly flexible water network and of binding two ligands, which still allow the exchange of water molecules around it (Fig. 56C).

In summary, the same distribution of driving forces contribute to the binding of penta- and hexasaccharide in WT and E372Q mutant, while the high affinity for the latter derived from the replacement of a highly conserved water molecule from the β-sheet (Beuming et al., 2009) and a looser bound water network in apo (Chen et al., 2008). In E372A, in contrast, the driving forces for the binding derive from a relatively stiffly bound ligand in the most populated solution conformer (Zaccheus et al., 2012), what would emphasize that no conformational changes disfavor the binding process (Ahmad et al., 2016). Furthermore, a loose water network (Clarke et al., 2001) drives the binding, even though an entropic loss is observed for the immobilization of water upon the binding event. To prove our interpretations further kinetic evaluations by MD simulation of relaxation time of water molecules (Tiwary et al., 2015), entropic contributions or side chain residues (Schauperl et al., 2016) would be necessary. The computational based finding would need to be confirmed by NMR relaxation dispersion experiments to evaluate side chain and water dynamics (Saraboji et al., 2012; Tzeng and Kalodimos, 2012).

4.3 Technical challenges in the design of high-affinity carbohydrate binding proteins

The design of a high-affinity carbohydrate binding protein was probed in this work on the base of Sf6 TSP cysteine mutations with computational prediction and experimental validation. The computational prediction was based on 100 ns simulations of the mutants and the analysis of ligand flexibility, hydrogen bonds and water network distribution. Thereby two candidates were identified as binders with an increased affinity due to decreased ligand flexibility and increased total hydrogen bond occupancy. From the two candidates only one could be identified unambigouosly as a binder with an increased affinity: S246C. This residue participates in the hydrogen bond contacts to the carbohydrate ligand via the backbone oxygen and the hydrogen bond occupancy increased for cysteine compared to serine. Cysteines primarily form hydrogen bonds with carbonyl groups and intraresidue hydrogen bonds of C=O···H-S have been observed to lead to constraints in the backbone and reduce flexibility (Pal and Chakrabarti, 1998), which might influence the ligand binding. Furthermore one could claim the hypothesis that the electronegativity of sulfur (χ=2.4) is reduced in comparison to oxygen (χ=3.5) and the higher the electronegativity, the stronger the hydrogen bonds. The backbone oxygen in a cysteine residue would therefore have a higher negative partial charge than in a serine as the counter action of the thiol group is lower than of the hydroxyl group, what would increase the hydrogen bond strength of the backbone oxygen in the cysteine case. However, the effect is probably weak as the two functional groups are separated by two carbons.

Beside the chemical nature of the hydrogen bonds formed by cysteines, the residue also tends to from disulfide bonds leading to aggregation of several Sf6 TSP trimers (data not shown). Although the formation of a disulfide bridge between two C246 residues would probably block the two respective binding sites for ligand binding, the number of binding sites on a dimeric Sf6 TSP-trimer would still increase to four. With further aggregation the number of sites would result in (n+2)-binding sites with n as the number of tailspike trimers. This multimerization would therefore increase the binding due to the avidity (Lis and Sharon, 1998; Shinohara et al., 1997).

Avidity could also be achieved in a more controlled way by engineering a dimer of Sf6 TSP trimers. This strategy has been shown to increase carbohydrate binding affinity in CBM by six to eleven fold, when two CBM were linked by a twelve amino acid linker (Ribeiro et al., 2016). However, tailspike would require a certain architecture, where the two trimers are

parallel with their binding sites to each other as in the phage tail, so the two trimers could bind neighbouring polysaccharide strains on the cell surface. Therefore N-to-N fusion would be appealing by a sortase-click chemistry mechanism (Witte et al., 2012).

The mutant design approach in this work was computationally based. However the alchemical free-energy calculations could not be developed to an agreeable level of reproducibility of experimental data so far. Alchemical free energy calculations have been shown to be reliable in peptides of 20 amino acids for protein stability analysis with a web server for the alchemical set-up now also available (Gapsys and de Groot, 2017). However protein-protein interaction energies involving amino acid binding site alchemical changes have shown to be more challenging and especially the transformation of amino acids from uncharged to charged states like in the transformation of glutamate to glutamine or alanine are linked with a higher error (Gapsys et al., 2016). As computational power improved, alchemical free energy calculations are more affordable and the undersampling problem connected with this method can be overcome more easily (Kutzner et al., 2015). Additionally docking algorithms could be improved and the application of docking on fragments up to undecasaccharides was successful as well as the docking of the SfY pentasaccharide on a FAB fragment (Nivedha et al., 2016). Rosetta has been extended to include carbohydrate moieties on glycoproteins and glycan ligand docking (Labonte et al., 2017). However for a reasonable carbohydrate binding protein design it would be necessary to combine it with an explicit water approach (Li et al., 2015).

Despite the computational design approaches, the combination of the latter with an experimental approach to generate a higher number of potential high-affinity binders by error-prone PCR or directed evolution would probably increase the success rate in the finding of a high-affinity Sf6 TSP (Rosenfeld et al., 2016; Sun et al., 2016). For that, also the ELITA approach can be used as a form of display assay of the derived mutants, as the *Strep*-tag®II tailspikes can already be detected in crude cell extracts (data not shown).

The here presented approach for the design of a high-affinity carbohydrate binding protein resulted in a bacterial sensor for the detection of *Shigella flexneri* cells. Furthermore the obtained insights of carbohydrate-protein interactions might help to improve carbohydrate binding drugs in the future.

5 References

Abel, R., Young, T., Farid, R., Berne, B.J., Friesner, R.A., (2008). Role of the active-site solvent in the thermodynamics of factor Xa ligand binding. *J. Am. Chem. Soc. 130*, 2817–2831.

Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B., Lindahl, E., (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX 1–2*, 19–25.

Ackermann, H.W., (2011). Bacteriophage taxonomy. *Microbiol. Aust. 32*, 90–94.

Afonine, P.V., Grosse-Kunstleve, R.W., Adams, P.D., Urzhumtsev, A., (2013). Bulk-solvent and overall scaling revisited: faster calculations, improved results. *Acta Crystallogr. D Biol. Crystallogr. 69*, 625–634.

Agaisse, H., (2016). Molecular and cellular mechanisms of *Shigella flexneri* dissemination. *Front. Cell. Infect. Microbiol. 6*.

Ahmad, M., Helms, V., Kalinina, O.V., Lengauer, T., (2016). The role of conformational changes in molecular recognition. *J. Phys. Chem. B 120*, 2138–2144.

Allison, G.E., Verma, N.K., (2000). Serotype-converting bacteriophages and O-antigen modification in *Shigella flexneri*. *TRENDS Microbiol. 17*.

Anderson, M., Sansonetti, P.J., Marteyn, B.S., (2016). *Shigella* diversity and changing landscape: Insights for the twenty-first century. *Front. Cell. Infect. Microbiol. 6*.

Andres, D., Gohlke, U., Broeker, N.K., Schulze, S., Rabsch, W., Heinemann, U., Barbirz, S., Seckler, R., (2013). An essential serotype recognition pocket on phage P22 tailspike protein forces *Salmonella enterica* serovar *Paratyphi* A O-antigen fragments to bind as nonsolution conformers. *Glycobiology 23*, 486–494.

Arnold, F.H., (1998). Design by directed evolution. *Acc. Chem. Res. 31*, 125–131.

Asensio, J.L., Ardá, A., Cañada, F.J., Jiménez-Barbero, J., (2013). Carbohydrate–aromatic interactions. *Acc. Chem. Res. 46*, 946–954.

Avigad, G., (1983). A simple spectrophotometric determination of formaldehyde and other aldehydes: Application to periodate-oxidized glycol systems. *Anal. Biochem. 134*, 499–504.

Badger, J., (1997). [17] Modeling and refinement of water molecules and disordered solvent, in: Enzymology, B.-M. in (Ed.), *Macromolecular Crystallography Part B. Academic Press*, pp. 344–352.

Barbirz, S., Becker, M., Freiberg, A., Seckler, R., (2009). Phage tailspike proteins with β-solenoid fold as thermostable carbohydrate binding materials. *Macromol. Biosci. 9*, 169–173.

Barbirz, S., Müller, J.J., Uetrecht, C., Clark, A.J., Heinemann, U., Seckler, R., (2008). Crystal structure of *Escherichia coli* phage HK620 tailspike: podoviral tailspike endoglycosidase modules are evolutionarily related. *Mol. Microbiol. 69*, 303–316.

Barillari, C., Taylor, J., Viner, R., Essex, J.W., (2007). Classification of water molecules in protein binding sites. *J. Am. Chem. Soc. 129*, 2577–2587.

Baron, R. (Ed.), (2012). Computational drug discovery and design, methods in molecular biology. *Springer New York*, New York, NY.

Baxa, U., Cooper, A., Weintraub, A., Pfeil, W., Seckler, R., (2001). Enthalpic barriers to the hydrophobic binding of oligosaccharides to phage P22 tailspike protein. *Biochemistry (Mosc.) 40*, 5144–5150.

Baxa, U., Steinbacher, S., Miller, S., Weintraub, A., Huber, R., Seckler, R., (1996). Interactions of phage P22 tails with their cellular receptor, *Salmonella* O-antigen polysaccharide. *Biophys. J. 71*, 2040–2048.

Bélot, F., Wright, K., Costachel, C., Phalipon, A., Mulard, L.A., (2004). Blockwise approach to fragments of the O-specific polysaccharide of *Shigella flexneri* serotype 2a: Convergent synthesis of a decasaccharide representative of a dimer of the branched repeating unit. *J. Org. Chem. 69*, 1060–1074.

Ben-Naim, A., Marcus, Y., (1984). Solvation thermodynamics of nonionic solutes. *J. Chem. Phys. 81*, 2016.

Bennett, C.H., (1976). Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys. 22*, 245–268.

Berman, H.M., Battistuz, T., Bhat, T.N., Bluhm, W.F., Bourne, P.E., Burkhardt, K., Feng, Z., Gilliland, G.L., Iype, L., Jain, S., others, (2002). The protein data bank. *Acta Crystallogr. D Biol. Crystallogr. 58*, 899–907.

Beuming, T., Che, Y., Abel, R., Kim, B., Shanmugasundaram, V., Sherman, W., (2012). Thermodynamic analysis of water molecules at the surface of proteins and applications to binding site prediction and characterization. *Proteins Struct. Funct. Bioinforma. 80*, 871–883.

Beuming, T., Farid, R., Sherman, W., (2009). High-energy water sites determine peptide binding affinity and specificity of PDZ domains. *Protein Sci. 18*, 1609–1619.

Biela, A., Nasief, N.N., Betz, M., Heine, A., Hangauer, D., Klebe, G., (2013). Dissecting the hydrophobic effect on the molecular level: The role of water, enthalpy, and entropy in ligand binding to Thermolysin. *Angew. Chem. Int. Ed. 52*, 1822–1828.

Bobone, S., van de Weert, M., Stella, L., (2014). A reassessment of synchronous fluorescence in the separation of Trp and Tyr contributions in protein emission and in the determination of conformational changes. *J. Mol. Struct. 1077*, 68–76.

Boraston, A.B., Bolam, D., Gilbert, H., Davies, G., (2004). Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem J 382*, 769–781.

Bradley, D.E., (1967). Ultrastructure of bacteriophages and bacteriocins. *Bacteriol. Rev. 31*, 230–314.

Brás, N.F., Cerqueira, N.M.F.S.A., Fernandes, P.A., Ramos, M.J., (2008). Carbohydrate-binding modules from family 11: Understanding the binding mode of polysaccharides. *Int. J. Quantum Chem. 108*, 2030–2040.

Broeker, N.K., Andres, D., Kang, Y., Gohlke, U., Schmidt, A., Kunstmann, S., Santer, M., Barbirz, S., (2017). Complex carbohydrate recognition by proteins: Fundamental insights from bacteriophage cell adhesion systems. *Perspect. Sci. 11,* 45-52.

Broeker, N.K., Gohlke, U., Müller, J.J., Uetrecht, C., Heinemann, U., Seckler, R., Barbirz, S., (2013). Single amino acid exchange in bacteriophage HK620 tailspike protein results in thousand-fold increase of its oligosaccharide affinity. *Glycobiology 23*, 59–68.

Brogioni, B., Berti, F., (2014). Surface plasmon resonance for the characterization of bacterial polysaccharide antigens: a review. *MedChemComm 5*, 1058.

Bullock, W.O., Fernandez, J.M., Short, J.M., (1987). XL1-Blue: a high efficiency plasmid transforming *recA Escherichia coli* strain with beta-galactosidase selection. *Biotechniques 5*, 376–378.

Bundle, D.R., Young, N.M., (1992). Carbohydrate-protein interactions in antibodies and lectins. *Curr. Opin. Struct. Biol. 2*, 666–673.

Calcuttawala, F., Hariharan, C., Pazhani, G.P., Ghosh, S., Ramamurthy, T., (2015). Activity spectrum of colicins produced by *Shigella sonnei* and genetic mechanism of colicin resistance in conspecific *S. sonnei* strains and *Escherichia coli*. *Antimicrob. Agents Chemother. 59*, 152–158.

Callis, P.R., (2014). Binding phenomena and fluorescence quenching. II: Photophysics of aromatic residues and dependence of fluorescence spectra on protein conformation. *J. Mol. Struct. 1077*, 22–29.

Carlin, N.I., Lindberg, A.A., (1983). Monoclonal antibodies specific for O-antigenic polysaccharides of *Shigella flexneri*: clones binding to II, II: 3, 4, and 7, 8 epitopes. *J. Clin. Microbiol. 18*, 1183–1189.

Carlin, N.I., Wehler, T., Lindberg, A.A., (1986). *Shigella flexneri* O-antigen epitopes: Chemical and immunochemical analyses reveal that epitopes of type III and group 6 antigens are identical. *Infect. Immun. 53*, 110–115.

Case, D.A., Darden, T.A., Cheatham, T.E.I., Simmerling, C.L., Wang, J., Duke, R.E., Luo, R., Walker, R.C., Zhang, W., Merz, K.M., Roberts, B., Hayik, S., Roitberg, A., Seabra, G., Swails, J., Götz, A.W., Kolossváry, I., Wong, K.F., Paesani, J., Vanicek, J., R.M. Wolf, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, P.A. Kollman, (2013). AmberTools 13. *University of California*, San Francisco. (ambermd.org/doc12/AmberTools13.pdf, 28.04.2017)

Case, D.A., Darden, T.A., Cheatham, T.E.I., Simmerling, C.L., Wang, J., Duke, R.E., Luo, R., Walker, R.C., Zhang, W., Merz, K.M., Roberts, B., Hayik, S., Roitberg, A., Seabra, G., Swails, J., Götz, A.W., Kolossváry, I., Wong, K.F., Paesani, J., Vanicek, J., R.M. Wolf, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, P.A. Kollman, (2012). Amber 12. *University of California*, San Francisco. (ambermd.org/doc12/Amber12.pdf, 28.04.2017)

Casjens, S., Winn-Stapley, D.A., Gilcrease, E.B., Morona, R., Kühlewein, C., Chua, J.E.H., Manning, P.A., Inwood, W., Clark, A.J., (2004). The chromosome of *Shigella flexneri* bacteriophage Sf6: Complete nucleotide sequence, genetic mosaicism, and DNA packaging. *J. Mol. Biol. 339*, 379–394.

Chang, J., Weigele, P., King, J., Chiu, W., Jiang, W., (2006). Cryo-EM asymmetric reconstruction of bacteriophage P22 reveals organization of its DNA packaging and infecting machinery. *Structure 14*, 1073–1082.

Chávez, M.I., Andreu, C., Vidal, P., Aboitiz, N., Freire, F., Groves, P., Asensio, J.L., Asensio, G., Muraki, M., Cañada, F.J., Jiménez-Barbero, J., (2005). On the importance of carbohydrate–aromatic interactions for the molecular recognition of oligosaccharides by proteins: NMR studies of the structure and binding affinity of AcAMP2-like peptides with non-natural naphthyl and fluoroaromatic residues. *Chem. – Eur. J. 11*, 7060–7074.

Chen, X., Weber, I., Harrison, R.W., (2008). Hydration water and bulk water in proteins have distinct properties in radial distributions calculated from 105 atomic resolution crystal structures. *J. Phys. Chem. B 112*, 12073–12080.

Chen, Y., Liu, B., Yu, H.-T., Barkley, M.D., (1996). The peptide bond quenches indole fluorescence. *J. Am. Chem. Soc. 118*, 9271–9278.

Chen, Y., Zychlinsky, A., (1994). Apoptosis induced by bacterial pathogens. *Microb. Pathog. 17*, 203–212.

Chervenak, M.C., Toone, E.J., (1994). A direct measure of the contribution of solvent reorganization to the enthalpy of binding. *J. Am. Chem. Soc. 116*, 10533–10539.

Chowers, Y., Kirschner, J., Keller, N., Barshack, I., Bar-Meir, S., Ashkenazi, S., Schneerson, R., Robbins, J., Passwell, J.H., (2007). O-specific polysaccharide conjugate vaccine-induced IgG antibodies prevent invasion of *Shigella* into Caco-2 cells and may be curative. *Proc. Natl. Acad. Sci. 104*, 2396–2401.

Chua, (1999). The *Shigella flexneri* bacteriophage Sf6 tailspike protein (TSP)/ endorhamnosidase is related to the bacteriophage P22 TSP and has a motif common to exo- and endoglycanases, and C-5 epimerases. *Microbiology. 145*, 1649-59.

Cirino, P., Mayer, K., Umeno, D., (2003). Generating mutant libraries using error-prone PCR, in: Arnold, F., Georgiou, G. (Eds.), Directed Evolution Library Creation, Methods in Molecular Biology™. *Humana Press*, pp. 3–9.

Clark, C.A., Beltrame, J., Manning, P.A., (1991). The O~C gene encoding a lipopolysaccharide O-antigen acetylase maps adjacent to the integrase-encoding gene on the genome of *Shigella flexneri* bacteriophage Sf6. *Gene 107*, 43–52.

Clarke, C., Woods, R.J., Gluska, J., Cooper, A., Nutley, M.A., Boons, G.-J., (2001). Involvement of water in carbohydrate−protein binding. *J. Am. Chem. Soc. 123*, 12238–12247.

Cooper, M.A., (2004). Advances in membrane receptor screening and analysis. *J. Mol. Recognit. 17*, 286–315.

Cooper, M.A., Hansson, A., Löfås, S., Williams, D.H., (2000). A vesicle capture sensor chip for kinetic analysis of interactions with membrane-bound receptors. *Anal. Biochem. 277*, 196–205.

Cooper, M.A., Williams, D.H., (1999). Kinetic analysis of antibody-antigen interactions at a supported lipid monolayer. *Anal. Biochem. 276*, 36–47.

Copeland, R.A., (2000). Protein-ligand binding equilibria. *Enzym. Pract. Introd. Struct. Mech. Data Anal.*, 76–108.

Coutinho, P.M., Deleury, E., Davies, G.J., Henrissat, B., (2003). An evolving hierarchical family classification for glycosyltransferases. *J. Mol. Biol. 328*, 307–317.

Creighton, T.E. (Ed.), (1997). Protein structure: A practical approach, Second Edition. ed, Practical Approach Series. *Oxford University Press*, Oxford, New York.

Darveau, R.P., Hancock, R.E., (1983). Procedure for isolation of bacterial lipopolysaccharides from both smooth and rough *Pseudomonas aeruginosa* and *Salmonella typhimurium* strains. *J. Bacteriol. 155*, 831–838.

Das, R., Baker, D., (2008). Macromolecular modeling with Rosetta. *Annu. Rev. Biochem. 77*, 363–382.

Davies, G., Henrissat, B., (1995). Structures and mechanisms of glycosyl hydrolases. *Structure 3*, 853–859.

Davis, A.P., Wareham, R.S., (1999). Carbohydrate recognition through noncovalent interactions: A challenge for biomimetic and supramolecular Chemistry. *Angew. Chem. Int. Ed. 38*, 2978–2996.

Day, C.J., Tran, E.N., Semchenko, E.A., Tram, G., Hartley-Tassell, L.E., Ng, P.S.K., King, R.M., Ulanovsky, R., McAtamney, S., Apicella, M.A., Tiralongo, J., Morona, R., Korolik, V., Jennings, M.P., (2015). Glycan:glycan interactions: High affinity biomolecular interactions that can mediate binding of pathogenic bacteria to host cells. *Proc. Natl. Acad. Sci. 112*, E7266–E7275.

de Mol, N.J., Fischer, M.J., (2008). Kinetic and thermodynamic analysis of ligand-receptor interactions: SPR applications in drug development. *Handb. Surf. Plasmon Reson.* 123–172.

DeMarco, M.L., Woods, R.J., (2008). Structural glycobiology: A game of snakes and ladders. *Glycobiology 18*, 426–440.

D'Errico, G., Silipo, A., Mangiapia, G., Vitiello, G., Radulescu, A., Molinaro, A., Lanzetta, R., Paduano, L., (2010). Characterization of liposomes formed by lipopolysaccharides from *Burkholderia cenocepacia*, *Burkholderia multivorans* and *Agrobacterium tumefaciens*: from the molecular structure to the aggregate architecture. *Phys. Chem. Chem. Phys. 12*, 13574.

Deslaurier, R., Ian C. P., S., Roderic, W., (1974). Conformational mobility of the pyrrolidine ring of proline in peptides and peptide hormones as manifest in carbon 13 spin-lattice relaxation times. *J. Biochem. (Tokyo) 249*, 7006–7010.

Dhillon, T.S., Poon, A.P.W., Chan, D., Clark, A.J., (1998). General transducing phages like *Salmonella* phage P22 isolated using a smooth strain of *Escherichia coli* as host. *FEMS Microbiol. Lett. 161*, 129–133.

Dröge, W., Lehmann, V., Lüderitz, O., Westphal, O., (1970). Structural investigations on the 2-keto-3-deoxyoctonate region of lipopolysaccharides. *Eur. J. Biochem. 14*, 175–184.

Duan, Y., Wu, C., Chowdhury, S., Lee, M.C., Xiong, G., Zhang, W., Yang, R., Cieplak, P., Luo, R., Lee, T., others, (2003). A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem. 24*, 1999–2012.

Dubendorf, J.W., Studier, F.W., (1991). Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with *lac* repressor. *J. Mol. Biol. 219*, 45–59.

Dunitz, J.D., (1995). Win some, lose some: enthalpy-entropy compensation in weak intermolecular interactions. *Chem. Biol. 2*, 709–712.

Dunitz, J.D., (1994). The entropic cost of bound water in crystals and biomolecules. *Science 264*, 670–670.

DuPont, H.L., Levine, M.M., Hornick, R.B., Formal, S.B., (1989). Inoculum size in shigellosis and implications for expected mode of transmission. *J. Infect. Dis. 159*, 1126–1128.

Eargle, J., Wright, D., Luthey-Schulten, Z., (2006). Multiple alignment of protein structures and sequences for VMD. *Bioinformatics 22*, 504–506.

ENA European Nucleotide Archive, (2017). AF128887.2 *Shigella flexneri* bacteriophage VI tailspike protein gene, complete cds. (www.ebi.ac.uk/ena/data/view/AF128887, 1.11.17).

Erb, E.-M., Chen, X., Allen, S., Roberts, C.J., Tendler, S.J.B., Davies, M.C., Forsén, S., (2000). Characterization of the surfaces generated by liposome binding to the modified dextran matrix of a surface plasmon resonance sensor chip. *Anal. Biochem. 280*, 29–35.

Evans, S.V., Sigurskjold, B.W., Jennings, H.J., Brisson, J.R., To, R., Tse, W.C., Altman, E., Frosch, M., Weisgerber, C., Kratzin, H.D., (1995). Evidence for the extended helical nature of polysaccharide epitopes. The 2.8 Å resolution structure and thermodynamics of ligand binding of an antigen binding fragment specific for alpha-(2→8)-polysialic acid. *Biochemistry (Mosc.) 34*, 6737–6744.

Fadda, E., Woods, R.J., (2011). On the role of water models in quantifying the binding free energy of highly conserved water molecules in proteins: The case of Concanavalin A. *J. Chem. Theory Comput. 7*, 3391–3398.

Fernández-Alonso, M. del C., Cañada, F.J., Jiménez-Barbero, J., Cuevas, G., (2005). Molecular recognition of saccharides by proteins. Insights on the origin of the carbohydrate-aromatic interactions. *J. Am. Chem. Soc. 127*, 7379–7386.

Finney, J.L., Eley, D.D., (1977). The organization and function of water in protein crystals [and discussion]. *Philos. Trans. R. Soc. B Biol. Sci. 278*, 3–32.

Francis, C.L., Ryan, T.A., Jones, B.D., Smith, S.J., Falkow, S., (1993). Ruffles induced by *Salmonella* and other stimuli direct macropinocytosis of bacteria. *Nature 364*, 639–642.

Frankel, G., Giron, J.A., Valmassoi, J., Schoolnik, G.K., (1989). Multi-gene amplification: simultaneous detection of three virulence genes in diarrhoeal stool. *Mol. Microbiol. 3*, 1729–1734.

Freiberg, A., Morona, R., van den Bosch, L., Jung, C., Behlke, J., Carlin, N., Seckler, R., Baxa, U., (2003). The tailspike protein of *Shigella* phage Sf6: A structural homolog of *Salmonella* phage P22 tailspike protein without sequence similarity in the β-helix domain. *J. Biol. Chem. 278*, 1542–1548.

Freire, E., Mayorga, O.L., Straume, M., (1990). Isothermal titration calorimetry. *Anal. Chem. 62*, 950A–959A.

Gabius, H.-J., (2008). Glycans: bioactive signals decoded by lectins. *Biochem. Soc. Trans. 36*, 1491–1496.

Galloway, S.M., Raetz, C.R., (1990). A mutant of *Escherichia coli* defective in the first step of endotoxin biosynthesis. *J. Biol. Chem. 265*, 6394–6402.

Galochkina, T., Zlenko, D., Nesterenko, A., Kovalenko, I., Strakhovskaya, M., Averyanov, A., Rubin, A., (2016). Conformational dynamics of the single lipopolysaccharide O-antigen in solution. *ChemPhysChem 17*, 2839–2853.

Gapsys, V., de Groot, B.L., (2017). pmx webserver: A user friendly interface for alchemistry. *J. Chem. Inf. Model. 57*, 109–114.

Gapsys, V., Michielssens, S., Seeliger, D., de Groot, B.L., (2016). Accurate and rigorous prediction of the changes in protein free energies in a large-scale mutation scan. *Angew. Chem. 128*, 7490–7494.

Gedig, E.T., (2008). Surface chemistry in SPR technology. *Handb. Surf. Plasmon Reson.* 173–220.

Gemski, P., Koeltzow, D.E., Formal, S.B., (1975). Phage conversion of *Shigella flexneri* group antigens. *Infect. Immun. 11*, 685–691.

Ghisaidoobe, A., Chung, S., (2014). Intrinsic tryptophan fluorescence in the detection and analysis of proteins: A focus on Förster resonance energy transfer techniques. *Int. J. Mol. Sci. 15*, 22518–22538.

Gilson, M.K., Given, J.A., Bush, B.L., McCammon, J.A., (1997). The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophys. J. 72*, 1047.

Goodford, P.J., (1985). A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem. 28*, 849–857.

Gouda, H., Kuntz, I.D., Case, D.A., Kollman, P.A., (2003). Free energy calculations for theophylline binding to an RNA aptamer: Comparison of MM-PBSA and thermodynamic integration methods. *Biopolymers 68*, 16–34.

Guex, N., Peitsch, M.C., (1997). SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling. *Electrophoresis 18*, 2714–2723.

Halle, B., (2004). Biomolecular cryocrystallography: structural changes during flash-cooling. *Proc. Natl. Acad. Sci. U. S. A. 101*, 4793–4798.

Hamelberg, D., McCammon, J.A., (2004). Standard free energy of releasing a localized water molecule from the binding pockets of proteins: Double-decoupling method. *J. Am. Chem. Soc. 126*, 7683–7689.

Häyrinen, J., Haseley, S., Talaga, P., Mühlenhoff, M., Finne, J., Vliegenthart, J.F., (2002). High affinity binding of long-chain polysialic acid to antibody, and modulation by divalent cations and polyamines. *Mol. Immunol. 39*, 399–411.

Henderson, I.R., Czeczulin, J., Eslava, C., Noriega, F., Nataro, J.P., (1999). Characterization of Pic, a secreted protease of *Shigella flexneri* and enteroaggregative *Escherichia coli*. *Infect. Immun. 67*, 5587–5596.

Hermanson, G.T., (2013). Bioconjugate techniques, 3rd Edition. *Academic Press.*

Hess, B., Bekker, H., Berendsen, H.J.C., Fraaije, J.G.E.M., (1997). LINCS: a linear constraint solver for molecular simulations. *J Comp Chem 18*, 1463–1472.

Hess, B., Kutzner, C., van der Spoel, D., Lindahl, E., (2008). GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput. 4*, 435–447.

Heukeshoven, J., Dernick, R., (1988). Improved silver staining procedure for fast staining in PhastSystem Development Unit. I. Staining of sodium dodecyl sulfate gels. *Electrophoresis 9*, 28–32.

Höcker, B., Beismann-Driemeyer, S., Hettwer, S., Lustig, A., Sterner, R., (2001). Dissection of a (βα)8-barrel enzyme into two folded halves. *Nat. Struct. Mol. Biol. 8*, 32–36.

Hodnik, V., Anderluh, G., (2010). Capture of intact liposomes on Biacore sensor chips for protein–membrane interaction studies, in: Mol, N.J., Fischer, M.J.E. (Eds.), Surface Plasmon Resonance. *Humana Press, Totowa, NJ*, pp. 201–211.

Houser, J., Komarek, J., Cioci, G., Varrot, A., Imberty, A., Wimmerova, M., (2015). Structural insights into *Aspergillus fumigatus* lectin specificity: AFL binding sites are functionally non-equivalent. *Acta Crystallogr. D Biol. Crystallogr. 71*, 442–453.

Hu, B., Lill, M.A., (2014). WATsite: Hydration site prediction program with PyMOL interface. *J. Comput. Chem. 35*, 1255–1260.

Humphrey, W., Dalke, A., Schulten, K., (1996). VMD – Visual molecular dynamics. *J. Mol. Graph. 14*, 33–38.

Israel, J.V., Anderson, T.F., Levine, M., (1967). *In vitro* morphogenesis of phage P22 from heads and base-plate parts. *Proc. Natl. Acad. Sci. 57*, 284–291.

Jeszenői, N., Bálint, M., Horváth, I., van der Spoel, D., Hetényi, C., (2016). Exploration of interfacial hydration networks of target–ligand complexes. *J. Chem. Inf. Model. 56*, 148–158.

Jeszenői, N., Horváth, I., Bálint, M., van der Spoel, D., Hetényi, C., (2015). Mobility-based prediction of hydration structures of protein surfaces. *Bioinformatics 31*, 1959–1965.

Jongkees, S.A.K., Withers, S.G., (2014). Unusual enzymatic glycoside cleavage mechanisms. *Acc. Chem. Res. 47*, 226–235.

Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., Klein, M.L., (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys. 79*, 926.

Jung, M., Philpott, M., Muller, S., Schulze, J., Badock, V., Eberspacher, U., Moosmayer, D., Bader, B., Schmees, N., Fernandez-Montalvan, A., Haendler, B., (2014). Affinity map of bromodomain protein 4 (BRD4) interactions with the histone H4 tail and the small molecule inhibitor JQ1. *J. Biol. Chem. 289*, 9304–9319.

Kadirvelraj, R., Foley, B.L., Dyekjær, J.D., Woods, R.J., (2008). Involvement of water in carbohydrate−protein binding: Concanavalin A revisited. *J. Am. Chem. Soc. 130*, 16933–16942.

Kajava, A.V., Steven, A.C., (2006). β-rolls, β-helices, and other β-solenoid proteins, in: Chemistry, B.-A. in P. (Ed.), Fibrous Proteins: Amyloids, Prions and Beta Proteins. *Academic Press*, pp. 55–96.

Kämpf, M.M., Braun, M., Sirena, D., Ihssen, J., Thöny-Meyer, L., Ren, Q., (2015). *In vivo* production of a novel glycoconjugate vaccine against *Shigella flexneri* 2a in recombinant *Escherichia coli*: identification of stimulating factors for *in vivo* glycosylation. *Microb. Cell Factories 14*.

Kang, Y., Barbirz, S., Lipowsky, R., Santer, M., (2014). Conformational diversity of O-antigen polysaccharides of the Gram-negative bacterium *Shigella flexneri* serotype Y. *J. Phys. Chem. B 118*, 2523–2534.

Kang, Y., Gohlke, U., Engström, O., Hamark, C., Scheidt, T., Kunstmann, S., Heinemann, U., Widmalm, G., Santer, M., Barbirz, S., (2016). Bacteriophage tailspikes and bacterial O-antigens as a model system to study weak-affinity protein–polysaccharide interactions. *J. Am. Chem. Soc. 138*, 9109–9118.

Kanipes, M.I., Lin, S., Cotter, R.J., Raetz, C.R.H., (2001). Ca^{2+}-induced phosphoethanolamine transfer to the outer 3-deoxy-D-manno-octulosonic acid moiety of *Escherichia coli* lipopolysaccharide: A novel membrane enzyme dependent upon phosphatidylethanolamine. *J. Biol. Chem. 276*, 1156–1163.

Kenne, L., Lindberg, B., Petersson, K., Katzenellenbogen, E., Romanowska, E., (1978). Structural studies of *Shigella flexneri* O-antigens. *Eur. J. Biochem. 91*, 279–284.

Kerzmann, A., Fuhrmann, J., Kohlbacher, O., Neumann, D., (2008). BALLDock/SLICK: A new method for protein-carbohydrate docking. *J. Chem. Inf. Model. 48*, 1616–1625.

Khalifa, M.B., Choulier, L., Lortat-Jacob, H., Altschuh, D., Vernet, T., (2001). Biacore data processing: An evaluation of the global fitting procedure. *Anal. Biochem. 293*, 194–203.

Kilambi, K.P., Pacella, M.S., Xu, J., Labonte, J.W., Porter, J.R., Muthu, P., Drew, K., Kuroda, D., Schueler-Furman, O., Bonneau, R., Gray, J.J., (2013). Extending RosettaDock with water, sugar, and pH for prediction of complex structures and affinities for CAPRI rounds 20-27: RosettaDock for CAPRI rounds 20-27. *Proteins Struct. Funct. Bioinforma. 81*, 2201–2209.

Kirschner, K.N., Yongye, A.B., Tschampel, S.M., González-Outeiriño, J., Daniels, C.R., Foley, B.L., Woods, R.J., (2008). GLYCAM06: A generalizable biomolecular force field. *Carbohydrates. J. Comput. Chem. 29*, 622–655.

Klein, G., Müller-Loennies, S., Lindner, B., Kobylak, N., Brade, H., Raina, S., (2013). Molecular and structural basis of inner core lipopolysaccharide alterations in *Escherichia coli*: Incorporation of glucuronic acid and phosphoethanolamine in the heptose region. *J. Biol. Chem. 288*, 8111–8127.

Kline, K.A., Fälker, S., Dahlberg, S., Normark, S., Henriques-Normark, B., (2009). Bacterial adhesins in host-microbe interactions. *Cell Host Microbe 5*, 580–592.

Knirel, Y.A., Kondakova, A.N., Vinogradov, E., Lindner, B., Perepelov, A.V., Shashkov, A.S., (2011). Lipopolysaccharide core structures and their correlation with genetic groupings of *Shigella* strains. A novel core variant in *Shigella boydii* type 16. *Glycobiology 21*, 1362–1372.

Knirel, Y.A., Lan, R., Senchenkova, S.N., Wang, J., Shashkov, A.S., Wang, Y., Perepelov, A.V., Xiong, Y., Xu, J., Sun, Q., (2013). O-antigen structure of *Shigella flexneri* serotype Yv and effect of the *lpt-O* gene variation on phosphoethanolamine modification of *S. flexneri* O-antigens. *Glycobiology 23*, 475–485.

Koshland, D.E., (1953). Stereochemistry and the mechanism of enzymatic reactions. *Biol. Rev. 28*, 416–436.

Kosma, P., Müller-Loennies, S., (2011). Anticarbohydrate antibodies: From molecular basis to clinical application. *Springer Science & Business Media.*

Kotloff, K.L., Nataro, J.P., Blackwelder, W.C., Nasrin, D., Farag, T.H., Panchalingam, S., Wu, Y., Sow, S.O., Sur, D., Breiman, R.F., Faruque, A.S., Zaidi, A.K., Saha, D., Alonso, P.L., Tamboura, B., Sanogo, D., Onwuchekwa, U., Manna, B., Ramamurthy, T., Kanungo, S., Ochieng, J.B., Omore, R., Oundo, J.O., Hossain, A., Das, S.K., Ahmed, S., Qureshi, S., Quadri, F., Adegbola, R.A., Antonio, M., Hossain, M.J., Akinsola, A., Mandomando, I., Nhampossa, T., Acácio, S., Biswas, K., O'Reilly, C.E., Mintz, E.D., Berkeley, L.Y., Muhsen, K., Sommerfelt, H., Robins-Browne, R.M., Levine, M.M., (2013). Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. *The Lancet 382*, 209–222.

Kulharia, M., Bridgett, S.J., Goody, R.S., Jackson, R.M., (2009). InCa-SiteFinder: A method for structure-based prediction of inositol and carbohydrate binding sites on proteins. *J. Mol. Graph. Model. 28*, 297–303.

Kutzner, C., Páll, S., Fechner, M., Esztermann, A., de Groot, B.L., Grubmüller, H., (2015). Best bang for your buck: GPU nodes for GROMACS biomolecular simulations. *J. Comput. Chem. 36*, 1990–2008.

Labonte, J.W., Adolf-Bryfogle, J., Schief, W.R., Gray, J.J., (2017). Residue-centric modeling and design of saccharide and glycoconjugate structures. *J. Comput. Chem. 38*, 276–287.

Ladburry, J.E., (1996). Just add water! The effect of water on the specificity of protein ligand binding sites and its potential application to drug design. *Chem. Biol.*, 973–980.

Laemmli, U.K., (1970). Cleavage of structural proteins during the assembly of head of bacteriophage T4. *Nature 227*, 680–685.

Lan, R., Reeves, P.R., (2002). *Escherichia coli* in disguise: molecular origins of *Shigella*. *Microbes Infect. 4*, 1125–1132.

Lasky, L.A., (1992). Selectins: interpreters of cell-specific carbohydrate information during inflammation. *Science 258*, 964–969.

Lawrenz, M., Baron, R., Wang, Y., McCammon, J.A., (2012). Independent-trajectory thermodynamic integration: A practical guide to protein-drug binding free energy calculations using distributed computing, in: Baron, R. (Ed.), Computational Drug Discovery and Design. *Springer New York, New York, NY*, pp. 469–486.

Lazaridis, T., (1998). Inhomogeneous fluid approach to solvation thermodynamics. 1. Theory. *J. Phys. Chem. B 102*, 3531–3541.

Lee, Y.C., Lee, R.T., (1995). Carbohydrate-protein interactions: basis of glycobiology. *Acc. Chem. Res. 28*, 321–327.

Lehmann, V., Lüderitz, O., Westphal, O., (1971). The linkage of pyrophosphorylethanolamine to heptose in the core of *Salmonella minnesota* lipopolysaccharides. *Eur. J. Biochem. 21*, 339–347.

Li, L., Xu, W., Lü, Q., (2015). Improving protein-ligand docking with flexible interfacial water molecules using SWRosettaLigand. *J. Mol. Model. 21*.

Li, Y., Cao, B., Liu, B., Liu, D., Gao, Q., Peng, X., Wu, J., Bastin, D.A., Feng, L., Wang, L., (2009). Molecular detection of all 34 distinct O-antigen forms of *Shigella*. *J. Med. Microbiol. 58*, 69–81.

Lindberg, A.A., (1973). Bacteriophage receptors. *Annu. Rev. Microbiol. 27*, 205–241.

Lindberg, A.A., Pál, T., (1993). Strategies for development of potential candidate *Shigella* vaccines. *Vaccine 11*, 168–179.

Lindberg, A.A., Wollin, R., Gemski, P., Wohlhieter, J.A., (1978). Interaction between bacteriophage Sf6 and *Shigella flexneri*. *J. Virol. 27*, 38–44.

Linman, M.J., Taylor, J.D., Yu, H., Chen, X., Cheng, Q., (2008). Surface plasmon resonance study of protein−carbohydrate interactions using biotinylated sialosides. *Anal. Chem. 80*, 4007–4013.

Lis, H., Sharon, N., (1998). Lectins: carbohydrate-specific proteins that mediate cellular recognition. *Chem. Rev. 98*, 637–674.

Löfas, S., Johnsson, B., (1990). A novel hydrogel matrix on gold surfaces in surface plasmon resonance sensors for fast and efficient covalent immobilization of ligands. *J. Chem. Soc. Chem. Commun.*, 1526–1528.

Loferer-Krößbacher, M., Klima, J., Psenner, R., (1998). Determination of bacterial cell dry mass by transmission electron microscopy and densitometric image analysis. *Appl. Environ. Microbiol. 64*, 688–694.

Lundquist, J.J., Toone, E.J., (2002). The cluster glycoside effect. *Chem. Rev. 102*, 555–578.

Lütteke, T., von der Lieth, C.W., (2009). Data mining the PDB for glyco-related data. *Methods Mol. Biol. Clifton NJ 534*, 293–310.

MacDonald, R.C., MacDonald, R.I., Menco, B.P.M., Takeshita, K., Subbarao, N.K., Hu, L., (1991). Small-volume extrusion apparatus for preparation of large, unilamellar vesicles. *Biochim. Biophys. Acta BBA-Biomembr. 1061*, 297–303.

MacKenzie, C.R., Hirama, T., Deng, S., Bundle, D.R., Narang, S.A., Young, N.M., (1996). Analysis by surface plasmon resonance of the influence of valence on the ligand binding affinity and kinetics of an anti-carbohydrate antibody. *J. Biol. Chem. 271*, 1527–1533.

Mader, C., Huber, C., Moll, D., Sleytr, U.B., Sara, M., (2004). Interaction of the crystalline bacterial cell surface layer protein SbsB and the secondary cell wall polymer of *Geobacillus stearothermophilus* PV72 assessed by real-time surface plasmon resonance biosensor technology. *J. Bacteriol. 186*, 1758–1768.

Mahoney, M.W., Jorgensen, W.L., (2000). A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J. Chem. Phys. 112*, 8910.

Maia, J., Carvalho, R.A., Coelho, J.F.J., Simões, P.N., Gil, M.H., (2011). Insight on the periodate oxidation of dextran and its structural vicissitudes. *Polymer 52*, 258–265.

Malik, A., Firoz, A., Jha, V., Ahmad, S., (2010). PROCARB: A database of known and modelled carbohydrate-binding protein structures with sequence-based prediction tools. *Adv. Bioinforma. 2010*.

Maniloff, J., Ackermann, H.-W., (1998). Taxonomy of bacterial viruses: establishment of tailed virus genera and the other *Caudovirales*. *Arch. Virol. 143*, 2051–2063.

Mao, S., Gao, C., Lo, C.-H.L., Wirsching, P., Wong, C.-H., Janda, K.D., (1999). Phage-display library selection of high-affinity human single-chain antibodies to tumor-associated carbohydrate antigens sialyl LewisX and LewisX. *Proc. Natl. Acad. Sci. 96*, 6953–6958.

Marshall, S.A., Lazar, G.A., Chirino, A.J., Desjarlais, J.R., (2003). Rational design and engineering of therapeutic proteins. *Drug Discov. Today 8*, 212–221.

Meiler, J., Baker, D., (2006). ROSETTALIGAND: Protein–small molecule docking with full side-chain flexibility. *Proteins Struct. Funct. Bioinforma. 65*, 538–548.

Mhlongo, N.N., Skelton, A.A., Kruger, G., Soliman, M.E.S., Williams, I.H., (2014). A critical survey of average distances between catalytic carboxyl groups in glycoside hydrolases: Average GH inter-carboxyl distances. *Proteins Struct. Funct. Bioinforma. 82*, 1747–1755.

Miller, S., Schuler, B., Seckler, R., (1998). A reversibly unfolding fragment of P22 tailspike protein with native structure: The isolated β-helix domain. *Biochemistry (Mosc.) 37*, 9160–9168.

Mishra, S.K., Calabró, G., Loeffler, H.H., Michel, J., Koča, J., (2015). Evaluation of selected classical force fields for alchemical binding free energy calculations of protein-carbohydrate complexes. *J. Chem. Theory Comput. 11*, 3333–3345.

Miyazaki, T., Nishikawa, A., Tonozuka, T., (2016). Crystal structure of the enzyme-product complex reveals sugar ring distortion during catalysis by family 63 inverting α-glycosidase. *J. Struct. Biol. 196*, 479–486.

Modenutti, C., Gauto, D., Radusky, L., Blanco, J., Turjanski, A., Hajos, S., Marti, M., (2015). Using crystallographic water properties for the analysis and prediction of lectin-carbohydrate complex structures. *Glycobiology 25*, 181–196.

Morin, A., Kaufmann, K.W., Fortenberry, C., Harp, J.M., Mizoue, L.S., Meiler, J., (2011). Computational design of an endo-1,4-β-xylanase ligand binding site. *Protein Eng. Des. Sel. 24*, 503–516.

Morona, R., (2003). Genetic modulation of *Shigella flexneri* 2a lipopolysaccharide O-antigen modal chain length reveals that it has been optimized for virulence. *Microbiology 149*, 925–939.

Morona, R., Bosch, L., (2003). Lipopolysaccharide O-antigen chains mask IcsA (VirG) in *Shigella flexneri*. *FEMS Microbiol. Lett. 221*, 173–180.

Morona, R., van den Bosch, L., Manning, P.A., (1995). Molecular, genetic, and topological characterization of O-antigen chain length regulation in *Shigella flexneri*. *J. Bacteriol. 177*, 1059–1068.

Müller, J.J., Barbirz, S., Heinle, K., Freiberg, A., Seckler, R., Heinemann, U., (2008). An intersubunit active site between supercoiled parallel β-helices in the trimeric tailspike endorhamnosidase of *Shigella flexneri* phage Sf6. *Structure 16*, 766–775.

Müller, K.M., Arndt, K.M., Plückthun, A., (1998). Model and simulation of multivalent binding to fixed ligands. *Anal. Biochem. 261*, 149–158.

Müller-Loennies, S., Lindner, B., Brade, H., (2002). Structural analysis of deacylated lipopolysaccharide of *Escherichia coli* strains 2513 (R4 core-type) and F653 (R3 core-type): Chemical structure of *E. coli* R3 and R4 LPS. *Eur. J. Biochem. 269*, 5982–5991.

Muraki, M., Ishimura, M., Harata, K., (2002). Interactions of wheat-germ agglutinin with GlcNAcβ1,6Gal sequence. *Biochim. Biophys. Acta BBA - Gen. Subj. 1569*, 10–20.

Murray, B.E., (1986). Resistance of *Shigella*, *Salmonella*, and other selected enteric pathogens to antimicrobial agents. *Rev. Infect. Dis. 8 Suppl 2*, S172-181.

Nakajima, H., (2001). Kinetic analysis of binding between Shiga toxin and receptor glycolipid Gb3Cer by surface plasmon resonance. *J. Biol. Chem. 276*, 42915–42922.

Nilsson, P., Persson, B., Uhlen, M., Nygren, P.-A., (1995). Real-time monitoring of DNA manipulations using biosensor technology. *Anal. Biochem. 224*, 400–408.

Nivedha, A.K., Makeneni, S., Foley, B.L., Tessier, M.B., Woods, R.J., (2014). Importance of ligand conformational energies in carbohydrate docking: Sorting the wheat from the chaff. *J. Comput. Chem. 35*, 526–539.

Nivedha, A.K., Thieker, D.F., Makeneni, S., Hu, H., Woods, R.J., (2016). Vina-Carb: Improving glycosidic angles during carbohydrate docking. *J. Chem. Theory Comput. 12*, 892–901.

Notenboom, V., Boraston, A.B., Chiu, P., Freelove, A.C.J., Kilburn, D.G., Rose, D.R., (2001). Recognition of cello-oligosaccharides by a family 17 carbohydrate-binding module: an X-ray crystallographic, thermodynamic and mutagenic study. *J. Mol. Biol. 314*, 797–806.

Osborn, M.J., Rosen, S.M., Rothfield, L., Zeleznick, L.D., Horecker, B.L., (1964). Lipopolysaccharide of the Gram-negative cell wall. *Science 145*, 783–789.

Padlan, E.A., Kabat, E.A., (1988). Model-building study of the combining sites of two antibodies to α(1→6)dextran. *Proc. Natl. Acad. Sci. U. S. A. 85*, 6885–6889.

Pal, D., Chakrabarti, P., (1998). Different types of interactions involving cysteine sulfhydryl group in proteins. *J. Biomol. Struct. Dyn. 15*, 1059–1072.

Pan, X.-L., Liu, W., Liu, J.-Y., (2013). Mechanism of the glycosylation step catalyzed by human α-galactosidase: A QM/MM metadynamics study. *J. Phys. Chem. B 117*, 484–489.

Pande, J., Szewczyk, M.M., Grover, A.K., (2010). Phage display: Concept, innovations, applications and future. *Biotechnol. Adv. 28*, 849–858.

Parent, K.N., Erb, M.L., Cardone, G., Nguyen, K., Gilcrease, E.B., Porcek, N.B., Pogliano, J., Baker, T.S., Casjens, S.R., (2014). OmpA and OmpC are critical host factors for bacteriophage Sf6 entry in *Shigella*. *Mol. Microbiol. 92*, 47–60.

Parent, K.N., Gilcrease, E.B., Casjens, S.R., Baker, T.S., (2012). Structural evolution of the P22-like phages: Comparison of Sf6 and P22 procapsid and virion architectures. *Virology 427*, 177–188.

Parrinello, M., (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys. 52*, 7182.

Parsot, C., (2005). *Shigella spp.* and enteroinvasive *Escherichia coli* pathogenicity factors. *FEMS Microbiol. Lett. 252*, 11–18.

Passwell, J.H., Ashkenzi, S., Banet-Levi, Y., Ramon-Saraf, R., Farzam, N., Lerner-Geva, L., Even-Nir, H., Yerushalmi, B., Chu, C., Shiloach, J., Robbins, J.B., Schneerson, R., (2010). Age-related efficacy of *Shigella* O-specific polysaccharide conjugates in 1–4-year-old Israeli children. *Vaccine 28*, 2231–2235.

Paul, F., Weikl, T.R., (2016). How to distinguish conformational selection and induced fit based on chemical relaxation rates. *PLOS Comput. Biol. 12*, e1005067.

Perepelov, A.V., Shekht, M.E., Liu, B., Shevelev, S.D., Ledov, V.A., Senchenkova, S.N., L'vov, V.L., Shashkov, A.S., Feng, L., Aparin, P.G., Wang, L., Knirel, Y.A., (2012). *Shigella flexneri* O-antigens revisited: Final elucidation of the O-acetylation profiles and a survey of the O-antigen structure diversity. *FEMS Immunol. Med. Microbiol. 66*, 201–210.

Pokala, N., Handel, T.M., (2001). Review: Protein design—Where we were, where we are, where we're going. *J. Struct. Biol. 134*, 269–281.

Popescu, A., Doyle, R.J., (1996). The Gram stain after more than a century. *Biotech. Histochem. 71*, 145–151.

Pronk, S., Pall, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M.R., Smith, J.C., Kasson, P.M., van der Spoel, D., Hess, B., Lindahl, E., (2013). GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics 29*, 845–854.

Raetz, C.R., (1990). Biochemistry of endotoxins. *Annu. Rev. Biochem. 59*, 129–170.

Raetz, C.R., Whitfield, C., (2002). Lipopolysaccharide endotoxins. *Annu. Rev. Biochem. 71*, 635–700.

Reeke, G.N., Becker, J.W., (1986). Three-dimensional structure of favin: Saccharide binding-cyclic permutation in leguminous lectins. *Science 234*, 1108–1111.

Remmel, N., Locatelli-Hoops, S., Breiden, B., Schwarzmann, G., Sandhoff, K., (2007). Saposin B mobilizes lipids from cholesterol-poor and bis(monoacylglycero)phosphate-rich membranes at acidic pH: Unglycosylated patient variant saposin B lacks lipid-extraction capacity. *FEBS J. 274*, 3405–3420.

Rhiel, L., Krah, S., Günther, R., Becker, S., Kolmar, H., Hock, B., (2014). REAL-Select: Full-length antibody display and library screening by surface capture on yeast cells. *PLoS ONE 9*.

Ribeiro, J.P., Pau, W., Pifferi, C., Renaudet, O., Varrot, A., Mahal, L.K., Imberty, A., (2016). Characterization of a high-affinity sialic acid-specific CBM40 from *Clostridium perfringens* and engineering of a divalent form. *Biochem. J. 473*, 2109–2118.

Roe, D.R., Cheatham, T.E., (2013). *ptraj* and *cpptraj*: Software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput. 9*, 3084–3095.

Rohl, C.A., Strauss, C.E.M., Misura, K.M.S., Baker, D., (2004). Protein structure prediction using Rosetta, in: Enzymology, B.-M. in (Ed.), Numerical Computer Methods, *Part D. Academic Press*, pp. 66–93.

Rojas-Macias, M.A., Lütteke, T., (2015). Statistical analysis of amino acids in the vicinity of carbohydrate residues performed by GlyVicinity. *Methods Mol. Biol. Clifton NJ 1273*, 215–226.

Roldós, V., Cañada, F.J., Jiménez-Barbero, J., (2011). Carbohydrate–protein interactions: A 3D view by NMR. *ChemBioChem 12*, 990–1005.

Ropartz, D., Bodet, P.-E., Przybylski, C., Gonnet, F., Daniel, R., Fer, M., Helbert, W., Bertrand, D., Rogniaux, H., (2011). Performance evaluation on a wide set of matrix-assisted laser desorption ionization matrices for the detection of oligosaccharides in a high-throughput mass spectrometric screening of carbohydrate depolymerizing enzymes. *Rapid Commun. Mass Spectrom. 25*, 2059–2070.

Rosenfeld, L., Heyne, M., Shifman, J.M., Papo, N., (2016). Protein engineering by combined computational and *in vitro* evolution approaches. *Trends Biochem. Sci. 41*, 421–433.

Rossato, G., Ernst, B., Vedani, A., Smieško, M., (2011). AcquaAlta: A directional approach to the solvation of ligand–protein complexes. *J. Chem. Inf. Model. 51*, 1867–1881.

Rye, C.S., Withers, S.G., (2000). Glycosidase mechanisms. *Curr. Opin. Chem. Biol. 4*, 573–580.

Sacchettini, J.C., Baum, L.G., Brewer, C.F., (2001). Multivalent protein−carbohydrate interactions. A new paradigm for supermolecular assembly and signal transduction. *Biochemistry (Mosc.) 40*, 3009–3015.

Samsonov, S., Teyra, J., Pisabarro, M.T., (2008). A molecular dynamics approach to study the importance of solvent in protein interactions. *Proteins Struct. Funct. Bioinforma. 73*, 515–525.

Samsonov, S.A., Teyra, J., Pisabarro, M.T., (2011). Docking glycosaminoglycans to proteins: Analysis of solvent inclusion. *J. Comput. Aided Mol. Des. 25*, 477–489.

Sansonetti, P.J., Van Nhieu, G.T., Égile, C., (1999). Rupture of the intestinal epithelial barrier and mucosal invasion by *Shigella flexneri*. *Clin. Infect. Dis. 28*, 466–475.

Santos, N.C., Silva, A.C., Castanho, M.A., Martins-Silva, J., Saldanha, C., (2003). Evaluation of lipopolysaccharide aggregation by light scattering spectroscopy. *ChemBioChem 4*, 96–100.

Saraboji, K., Håkansson, M., Genheden, S., Diehl, C., Qvist, J., Weininger, U., Nilsson, U.J., Leffler, H., Ryde, U., Akke, M., Logan, D.T., (2012). The carbohydrate-binding site in galectin-3 is preorganized to recognize a sugarlike framework of oxygens: Ultra-high-resolution structures and water dynamics. *Biochemistry (Mosc.) 51*, 296–306.

Sasakawa, C., (1997). Early stages of *Shigella* interaction with host cells. *J. Infect. Chemother. 3*, 63–72.

Schauperl, M., Podewitz, M., Waldner, B.J., Liedl, K.R., (2016). Enthalpic and entropic contributions to hydrophobicity. *J. Chem. Theory Comput. 12*, 4600–4610.

Schmidt, A., Rabsch, W., Broeker, N.K., Barbirz, S., (2016). Bacteriophage tailspike protein based assay to monitor phase variable glucosylations in *Salmonella* O-antigens. *BMC Microbiol. 16*.

Schofield, D.A., Wray, D.J., Molineux, I.J., (2014). Isolation and development of bioluminescent reporter phages for bacterial dysentery. *Eur. J. Clin. Microbiol. Infect. Dis. 34*, 395–403.

Schoonbroodt, S., Steukers, M., Viswanathan, M., Frans, N., Timmermans, M., Wehnert, A., Nguyen, M., Ladner, R.C., Hoet, R.M., (2008). Engineering antibody heavy chain CDR3 to create a phage display Fab library rich in antibodies that bind charged carbohydrates. *J. Immunol. 181*, 6213–6221.

Schrodinger, (2015). The PyMOL molecular graphics system, *Version 1.8.*

Schuck, P., Minton, A.P., (1996). Analysis of mass transport-limited binding kinetics in evanescent wave biosensors. *Anal. Biochem. 240*, 262–272.

Seckler, R., (1998). Folding and function of repetitive structure in the homotrimeric phage P22 tailspike protein. *J. Struct. Biol. 122*, 216–222.

Seeliger, D., de Groot, B.L., (2010). Protein thermostability calculations using alchemical free energy simulations. *Biophys. J. 98*, 2309–2316.

Sharabi, O., Erijman, A., Shifman, J.M., (2013). Computational methods for controlling binding specificity, in: Methods in Enzymology. *Elsevier*, pp. 41–59.

Shinohara, Y., Hasegawa, Y., Kaku, H., Shibuya, N., (1997). Elucidation of the mechanism enhancing the avidity of lectin with oligosaccharides on the solid phase surface. *Glycobiology 7*, 1201–1208.

Shionyu-Mitsuyama, C., Shirai, T., Ishida, H., Yamane, T., (2003). An empirical approach for structure-based prediction of carbohydrate-binding sites on proteins. *Protein Eng. Des. Sel. 16*, 467–478.

Simmons, D.A.R., Romanowska, E., (1987). Structure and biology of *Shigella flexneri* O-antigens. *J. Med. Microbiol. 23*, 289–302.

Sinowatz, F., Plendl, J., Kölle, S., (1998). Protein-carbohydrate interactions during fertilization. *Cells Tissues Organs 161*, 196–205.

Skerman, V.B.D., McGowan, V., Sneath, P.H.A., (1980). Approved lists of bacterial names. *Int. J. Syst. Evol. Microbiol. 30*, 225–420.

Small, P., Blankenhorn, D., Welty, D., Zinser, E., Slonczewski, J.L., (1994). Acid and base resistance in *Escherichia coli* and *Shigella flexneri*: role of σ^S and growth pH. *J. Bacteriol. 176*, 1729–1737.

Steinbacher, S., Miller, S., Baxa, U., Budisa, N., Weintraub, A., Seckler, R., Huber, R., (1997). Phage P22 tailspike protein: Crystal structure of the head-binding domain at 2.3 Å, fully refined structure of the endorhamnosidase at 1.56 Å resolution, and the molecular basis of O-antigen recognition and cleavage. *J. Mol. Biol. 267*, 865–880.

Stevenson, G., Kessler, A., Reeves, P.R., (1995). A plasmid-borne O-antigen chain length determinant and its relationship to other chain length determinants. *FEMS Microbiol. Lett. 125*, 23–30.

Strohalm, M., Hassman, M., Košata, B., Kodíček, M., (2008). mMass data miner: An open source alternative for mass spectrometric data analysis. *Rapid Commun. Mass Spectrom. 22*, 905–908.

Studier, F.W., (2005). Protein production by auto-induction in high-density shaking cultures. *Protein Expr. Purif. 41*, 207–234.

Sun, M.G.F., Seo, M.-H., Nim, S., Corbi-Verge, C., Kim, P.M., (2016). Protein engineering by highly parallel screening of computationally designed variants. *Sci. Adv. 2*, e1600692–e1600692.

Sun, Q., Knirel, Y.A., Lan, R., Wang, J., Senchenkova, S.N., Jin, D., Shashkov, A.S., Xia, S., Perepelov, A.V., Chen, Q., Wang, Y., Wang, H., Xu, J., (2012). A novel plasmid-encoded serotype conversion mechanism through addition of phosphoethanolamine to the O-antigen of *Shigella flexneri. PLoS ONE 7*, e46095.

Sun, Q., Knirel, Y.A., Wang, J., Luo, X., Senchenkova, S.N., Lan, R., Shashkov, A.S., Xu, J., (2014). Serotype-converting bacteriophage SfII encodes an acyltransferase protein that mediates 6-O-acetylation of GlcNAc in *Shigella flexneri* O-antigens, conferring on the host a novel O-antigen epitope. *J. Bacteriol. 196*, 3656–3666.

Sun, Q., Lan, R., Wang, Yiting, Zhao, A., Zhang, Shaomin, Wang, J., Wang, Yan, Xia, S., Jin, D., Cui, Z., Zhao, H., Li, Z., Ye, C., Zhang, Shuxia, Jing, H., Xu, J., (2011). Development of a multiplex PCR Aassay targeting O-antigen modification genes for molecular serotyping of *Shigella flexneri*. *J. Clin. Microbiol. 49*, 3766–3770.

Tailspike protein - *Shigella* phage Sf6 (*Shigella flexneri* bacteriophage VI) uniprot.org/uniprot/Q9XJP3 (accessed 1.11.17).

Taroni, C., Jones, S., Thornton, J.M., (2000). Analysis and prediction of carbohydrate binding sites. *Protein Eng. 13*, 89–98.

Taylor, N.R., (2013). Small world network strategies for studying protein structures and binding. *Comput. Struct. Biotechnol. J. 5*, 1–7.

Tiwary, P., Mondal, J., Morrone, J.A., Berne, B.J., (2015). Role of water and steric constraints in the kinetics of cavity–ligand unbinding. *Proc. Natl. Acad. Sci. 112*, 12015–12019.

Tzeng, S.-R., Kalodimos, C.G., (2012). Protein activity regulation by conformational entropy. *Nature 488*, 236–240.

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E., Berendsen, H.J.C., (2005). GROMACS: Fast, flexible, and free. *J. Comput. Chem. 26*, 1701–1718.

Van Steeg, T.J., Bergmann, K.R., Dimasi, N., Sachsenmeier, K.F., Agoram, B., (2016). The application of mathematical modelling to the design of bispecific monoclonal antibodies. *mAbs 8*, 585–592.

Varki, A., Cummings, R.D., Esko, J.D., Freeze, H.H., Stanley, P., Bertozzi, C.R., Hart, G.W., Etzler, M.E. (Eds.), (2009). Essentials of glycobiology, 2nd ed. *Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY).*

Varki, A., Cummings, R.D., Esko, J.D., Stanley, P., Hart, G., Aebi, M., Darvill, A., Kinoshita, T., Packer, N.H., Prestegard, J.J., Schnaar, R.L., Seeberger, P.H., (2016). Symbol nomenclature for glycans (SNFG). *Cold Spring Harbor Laboratory Press.*

Vedani, A., Huhta, D.W., (1991). Algorithm for the systematic solvation of proteins based on the directionality of hydrogen bonds. *J. Am. Chem. Soc. 113*, 5860–5862.

Verma, N.K., Brandt, J.M., Verma, D.J., Lindberg, A.A., (1991). Molecular characterization of the O-acetyl transferase gene of converting bacteriophage Sf6 that adds group antigen 6 to *Shigella flexneri*. *Mol. Microbiol. 5*, 71–75.

Vulliez-Le Normand, B., Saul, F.A., Phalipon, A., Belot, F., Guerreiro, C., Mulard, L.A., Bentley, G.A., (2008). Structures of synthetic O-antigen fragments from serotype 2a *Shigella flexneri* in complex with a protective monoclonal antibody. *Proc. Natl. Acad. Sci. 105*, 9976–9981.

Vyas, N.K., Vyas, M.N., Chervenak, M.C., Johnson, M.A., Pinto, B.M., Bundle, D.R., Quiocho, F.A., (2002). Molecular recognition of oligosaccharide epitopes by a monoclonal Fab specific for *Shigella flexneri* Y lipopolysaccharide: X-ray structures and thermodynamics. *Biochemistry (Mosc.) 41*, 13575–13586.

Wagner, R., Matrosovich, M., Klenk, H.-D., (2002). Functional balance between haemagglutinin and neuraminidase in influenza virus infections. *Rev. Med. Virol. 12*, 159–166.

Wang, J., Knirel, Y.A., Lan, R., Senchenkova, S.N., Luo, X., Perepelov, A.V., Wang, Y., Shashkov, A.S., Xu, J., Sun, Q., (2014). Identification of an O-acyltransferase gene (*oacB*) that mediates 3- and 4-O-acetylation of rhamnose III in *Shigella flexneri* O-antigens. *J. Bacteriol. 196*, 1525–1531.

Watarai, M., Tobe, T., Yoshikawa, M., Sasakawa, C., (1995). Contact of *Shigella* with host cells triggers release of Ipa invasins and is an essential function of invasiveness. *EMBO J. 14*, 2461–2470.

Welch, M.D., Way, M., (2013). Arp2/3-mediated actin-based motility: A tail of pathogen abuse. *Cell Host Microbe 14*, 242–255.

Witte, M.D., Cragnolini, J.J., Dougan, S.K., Yoder, N.C., Popp, M.W., Ploegh, H.L., (2012). Preparation of unnatural N-to-N and C-to-C protein fusions. *Proc. Natl. Acad. Sci. 109*, 11993–11998.

Wittmann, V., Pieters, R.J., (2013). Bridging lectin binding sites by multivalent carbohydrates. *Chem. Soc. Rev. 42*, 4492.

Wolfe, C.A., Hage, D.S., (1995). Studies on the rate and control of antibody oxidation by periodate. *Anal. Biochem. 231*, 123–130.

Xu, D., Zhang, W., Zhang, B., Liao, C., Shao, Y., (2016). Characterization of a biofilm-forming *Shigella flexneri* phenotype due to deficiency in Hep biosynthesis. *PeerJ 4*, e2178.

Yabuuchi, E., (2002). *Bacillus dysentericus* (sic) 1897 was the first rather than *Bacillus dysenteriae* 1898. *Int. J. Syst. Evol. Microbiol. 52*, 1041–1041.

Yang, Y., Lill, M.A., (2016). Dissecting the influence of protein flexibility on the location and thermodynamic profile of explicit water molecules in protein–ligand binding. *J. Chem. Theory Comput. 12*, 4578–4592.

Yethon, J.A., Heinrichs, D.E., Monteiro, M.A., Perry, M.B., Whitfield, C., (1998). Involvement of waaY, waaQ, and waaP in the modification of *Escherichia coli* lipopolysaccharide and their role in the formation of a stable outer membrane. *J. Biol. Chem. 273*, 26310–26316.

Yoder, M.D., Keen, N.T., Jurnak, F., (1993). New domain motif: the structure of pectate lyase C, a secreted plant virulence factor. *Science 260*, 1503–1507.

Zaccheus, M.V., Broeker, N.K., Lundborg, M., Uetrecht, C., Barbirz, S., Widmalm, G., (2012). Structural studies of the O-antigen polysaccharide from *Escherichia coli* TD2158 having O18 serogroup specificity and aspects of its interaction with the tailspike endoglycosidase of the infecting bacteriophage HK620. *Carbohydr. Res. 357*, 118–125.

Zacharias, M., Straatsma, T.P., McCammon, J.A., Quiocho, F.A., (1993). Inversion of receptor binding preferences by mutagenesis: Free energy thermodynamic integration studies on sugar binding to L-arabinose binding proteins. *Biochemistry (Mosc.) 32*, 7428–7434.

Zeng, X., Andrade, C.A.S., Oliveira, M.D.L., Sun, X.-L., (2012). Carbohydrate–protein interactions and their biosensing applications. *Anal. Bioanal. Chem. 402*, 3161–3176.

Zhao, H., Yang, Y., von Itzstein, M., Zhou, Y., (2014). Carbohydrate-binding protein identification by coupling structural similarity searching with binding affinity prediction. *J. Comput. Chem. 35*, 2177–2183.

Zinder, N.D., Lederberg, J., (1952). Genetic exchange in *Salmonella*. *J. Bacteriol. 64*, 679.

Zverlov, V.V., Hertel, C., Bronnenmeier, K., Hroch, A., Kellermann, J., Schwarz, W.H., (2000). The thermostable α-L-rhamnosidase RamA of *Clostridium stercorarium*: biochemical characterization and primary structure of a bacterial α-L-rhamnoside hydrolase, a new type of inverting glycoside hydrolase. *Mol. Microbiol. 35*, 173–179.

Zychlinsky, A., Prevost, M.C., Sansonetti, P.J., (1992). *Shigella flexneri* induces apoptosis in infected macrophages. *Nature 358*, 167–169.

Appendix

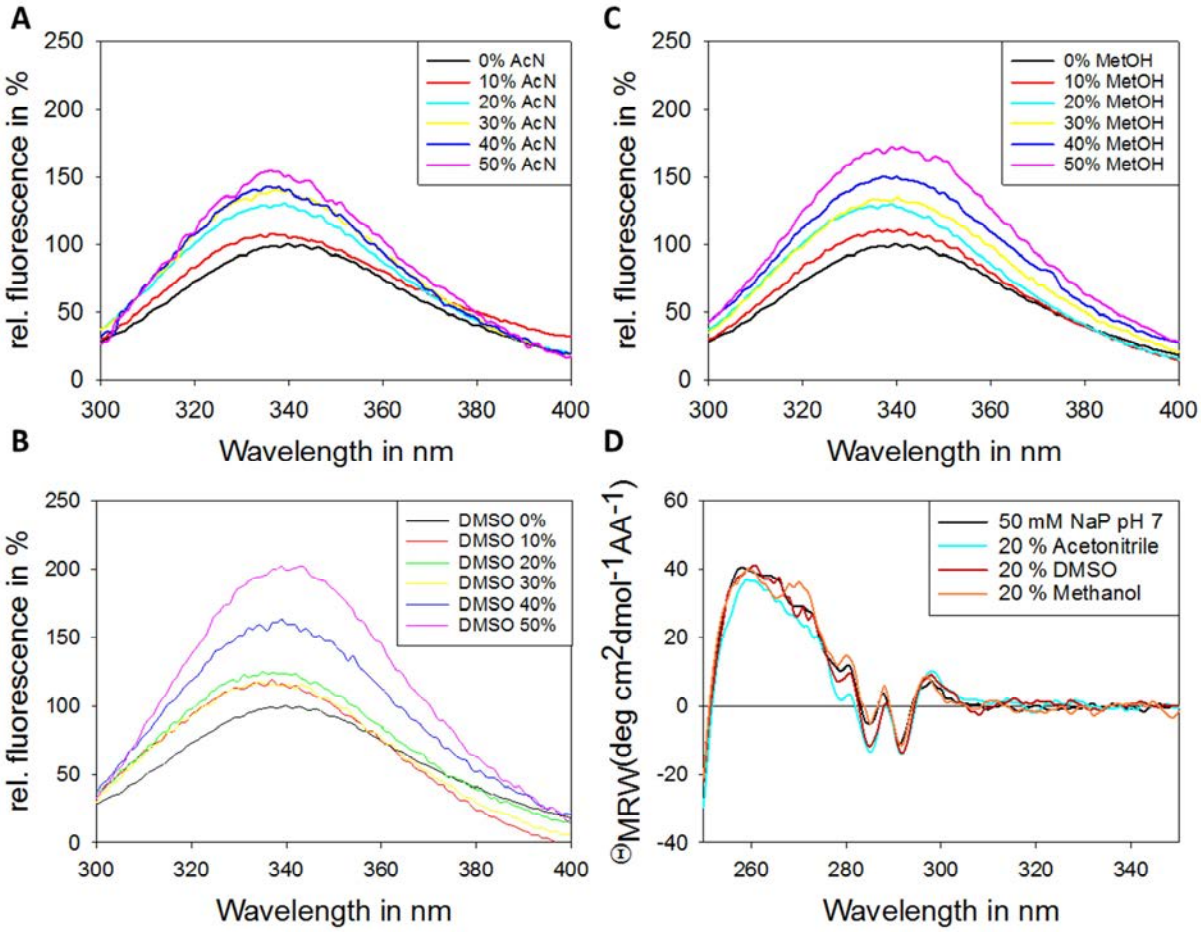

Fig. S 1: Fluorescence spectra of Sf6 TSP with organic solvent.
Tryptophan fluorescence of 184 nM Sf6 TSP_{EADA} excited at 280 nm and emission monitored between 300-400 nm with different amount of organic solvent in volume percent for acetonitrile (AcN) **(A)**, DMSO **(B)** and methanol (MetOH) **(C)**. **(D)** Near-UV CD spectra monitored between 350-250 nm of 36 µM Sf6 TSP_{EADA} in 50 mM sodium phosphate buffer pH 7 or with 20 % volume percent of organic solvent in the repesctive buffer.

Table S 1: Wavelength of spectra maxima of Sf6 TSP with organic solvent in the corresponding volume percent.

Solvent% (v/v)	AcNmax in nm	MetOHmax in nm	DMSOmax in nm
0	340	340	340
10	336	338	337
20	339	339	335
30	337	340	334
40	336	340	339
50	336	339	343

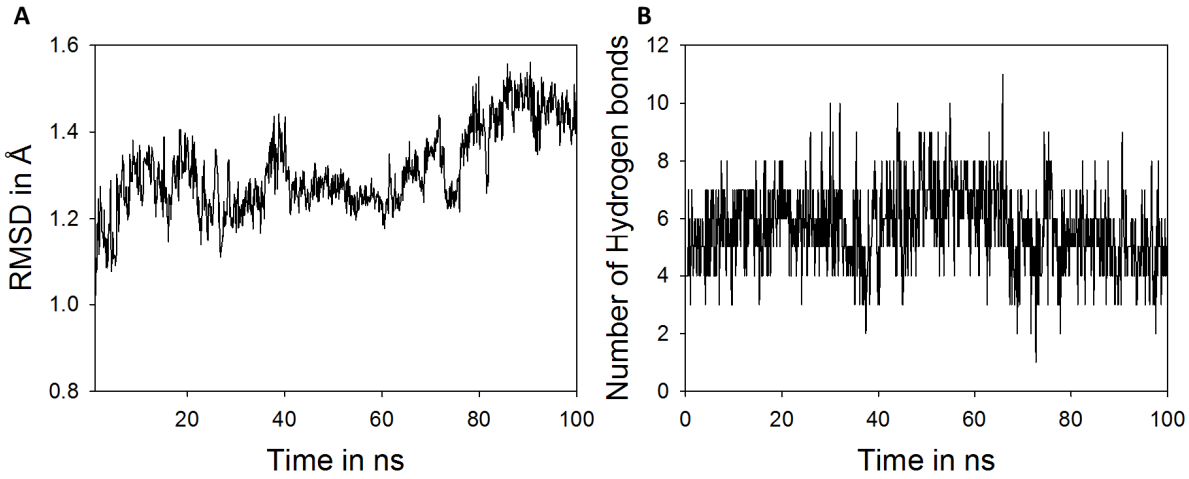

Fig. S 2: Trajectories of Sf6 TSP_{EADA} (4URR) simulation with octasaccharide ligand
(A) Root mean square deviation (RMSD) calculated for all protein atoms over 100 ns simulations. **(B)** Trajectory of the number of hydrogen bonds formed between protein and octasaccharide ligand within 100 ns simulation.

Table S 2: Contributing binding partners of Sf6 TSP_{EADA} with octasaccharide ligand in crystal structure 4URR

Hydrogen bonds	Protein	Ligand	Distance in Å
Upper site	Arg230-NH2	RAM1-O5	3.30
	Arg230-NH1	RAM1-O2	2.90
	Arg230-NH2	RAM1-O2	3.00
	Glu293-OE2	RAM1-O2	2.70
	Ser246-O	NAG2-O6	3.00
	Thr248-OG1	RAM3-O5	3.00
	Ser246-O	RAM4-O3	2.60
Lower site	Gln280-NE2	NAG6-O7	3.10
	Tyr400-OH	RAM8-O3	2.90
Nonpolar interactions	**Protein**	**Ligand**	**Distance in Å**
Upper site	Ala203-CB	RAM1-C6	3.70
	Thr315-CG2	NAG2-C8	4.80
	Thr248-CG2	RAM4-ring	4.30
Lower site	Tyr400-ring	NAG6-C6	4.10
	Val327-CG1	RAM7-C6	4.30
	Leu389-CD2	RAM8-C6	4.10
Replaced water	**Ligand**	**Water**	**Distance in Å**
Upper site	NAG7-O6	HOH2224	0.8
	RAM6-O3	HOH2645	1.1
	RAM5-O2	HOH2223	1.2
	RAM5-O5	HOH2135	1.0
Lower site	NAG3-O7	HOH2282	0.9
	NAG3-C8	HOH2283	0.6
	NAG3-O6	HOH2105	1.1
	NAG3-C6	HOH2187	0.9
	RAM2-O2	HOH2106	0.7
	RAM2-C1	HOH2203	1.3
	RAM2-O5	HOH2379	1.1
	RAM2-O4	HOH2182	0.8
	RAM1-O2	HOH2400	1.1
	RAM1-C3	HOH2399	0.4
	RAM1-C2	HOH2398	1.0
	RAM1-O5	HOH2649	0.8

Water bridges	Partner1	Partner2	Hbonds in Å	#water
Upper site	RAM1-O4	RAM1-O3	2.7; 3.4	1
	Asp247-OD1	NAG2-N2	2.7; 3.0; 2.8	2
	Asp247-OD2	NAG2-N2	2.8; 3.0; 2.8	2
	Arg257-NH2	NAG2-N2	3.4; 2.9; 2.8	2
	Glu293-OE2	NAG2-N2	2.6; 2.9; 2.8	2
	RAM3-O5	RAM3-O2	3.2; 3.1	1
	Asp245-O	RAM4-O4	2.8; 2.7	1
	Asp247-O	RAM4-O4	2.9; 2.7	1
	Ile278-O	RAM4-O4	3.3; 3.2	1
	Gln280-OE1	RAM4-O4	2.8; 3.2	1
	Asp245-O	RAM4-O4	2.9; 2.7; 3.2	2
	Gln280-N	RAM4-O4	3.0; 2.7; 3.2	2
	Gln280-O	RAM4-O4	3.4; 2.7; 3.2	2
Lower site	Asp397-OD1	RAM8-O1	3.2; 2.8	1
	Asp397-OD2	RAM8-O1	3.1; 2.8	1
Conserved water	**Water_apo**	**Water_8mer**	**Distance in Å**	
Upper site	HOH2183	HOH2110	0.2	
	HOH2120	HOH2193	0.6	
	HOH2235	HOH2138	0.4	
	HOH2180	HOH2348	0.7	
	HOH2219	HOH2084	0.2	
	HOH2341	HOH2179	0.3	
	HOH2087	HOH2230	0.4	
	HOH2229	HOH2006	0.4	
	HOH2280	HOH2007	0.9	
	HOH2164	HOH2005	0.7	
Lower site	HOH2184	HOH2004	1.3	
	HOH2185	HOH2010	0.7	
	HOH2375	HOH2129	0.6	
	HOH2417	HOH2205	0.4	
	HOH2423	HOH2224	0.6	
	HOH2122	HOH2319	0.3	

Table S 3: Conformational changes of binding site residues between X-ray and MD simulation

Binding site residues	Crystal vs. 8mer[a]		Crystal vs. 12mer[b]		8mer[a] vs. 12mer[b]	
	Distance in Å	Diff. Angle	Distance in Å	Diff. Angle	Distance in Å	Diff. Angle
Ala203	-	-	-	-	-	
Val204	-	-	-	-	-	
Arg230	-	-	-	-	-	
Arg257	-	-	-	-	NH1 2.0	χ4 22°
Thr259	-	-	-	-	-	-
Glu293	OE2 3.1	χ1 -111°	OE2 3.7	χ3 -156°	OE1 2.4	χ1 94°
Thr315	-	-	-	-	-	-
Asn340	ND2 3.3	χ1 -94°	ND2 4.4	χ1 -102°	-	-
Glu373	OE2 3.3	χ2 -105°	OE2 3.2	χ2 -112°	-	-
Ile377	-	-	-	-	-	-
Ala399	-	-	-	-	-	-
Tyr400	-	-	OH 5.7	χ1 -67°	OH 5.4	χ1 -72°
Trp421	-	-	-	-	-	-
Ala423	-	-	-	-	-	-
His453	-	-	-	-	-	-
Asn455	ND2 2.8	χ2 -83°	ND2 2.5	χ2 -67°	-	-
Ala457	-	-	-	-	CB 1.6	-
Thr483	-	-	-	-	-	-
Thr485	-	-	-	-	-	-
Ile506	-	-	-	-	-	-
Asn508	OD1 4.1	χ1 -85°	OD1 4.5	χ1 -95°	-	-
Tyr525	-	-	-	-	-	-
Asp245	OD1 3.6	χ1 104°	OD1 2.8	χ1 112°	-	-
Ser246	-	-	OG 2.3	χ1 105°	OG 2.1	χ1 108°
Asp247	-	-	-	-	-	-
Thr248	-	-	-	-	-	-
Gln280	NE2 1.5	χ2 66°	-	-	NE2 1.6	χ2 76°
Tyr282	-	-	OH 2.7	χ1 69°	OH 2.5	χ1 75°
Trp304	-	-	-	-	-	-
Gln325	-	-	-	-	-	-
Val327	-	-	-	-	-	-
Ala366	-	-	-	-	-	-
Leu389	-	-	-	-	CD1 1.7	ψ1 113°
Asn412	OD1 3.8	χ1 -93°	OD1 3.2	χ1 -110°	-	-
Gly413	O 3.2	ψ1 -143°	O 3.6	ψ1 -174°	-	-
Thr443	OG1 2.8	χ1 -27°	OG1 1.5	χ1 -22°	-	-
Gln444	OE1 2.9	χ2 -92°	-	-	OE1 3.1	χ2 -105°
Ser445	OG 1.7	χ1 -135°	OG 1.7	χ1 -132°	-	-
Phe476	-	-	-	-	-	-

a Frame considered by cluster analysis as the average conformation from a 100 ns simulation at time 82.86 ns
b Frame considered by cluster analysis as the average conformation from a 100 ns simulation at time 69.01 ns

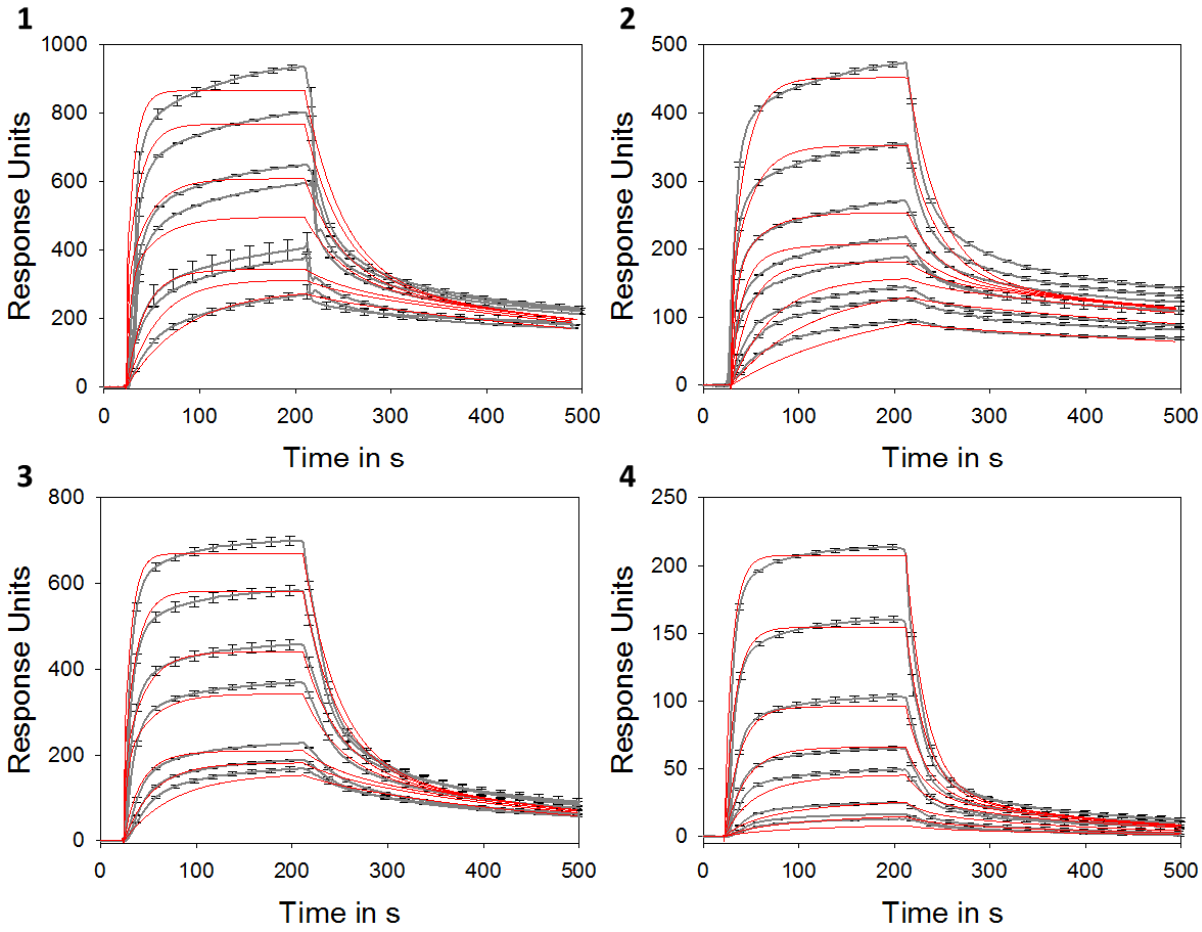

Fig. S 3: Sensograms of Sf6 TSP_{EADA} binding to immobilized SfY polysaccharide
The obtained sensograms (gray) were measured as triplicates and the average is depicted with standard deviation as error bars at every 10th point. The curves were monitored on two chips in four independent measurements (as numbers in the figure). The data were fitted by a heterogeneous ligand parallel binding model (red). Obtained parameters are listed in Tab. 6.

Bivalent binding model:

$$A + B \underset{k_{d1}}{\overset{k_{a1}}{\rightleftharpoons}} AB \qquad \frac{d[B]}{dt} = -(k_{a1}[A][B] - k_{d1}[AB]) - (k_{a2}[AB][B] - k_{d2}[ABB])$$

$$AB + B \underset{k_{d2}}{\overset{k_{a2}}{\rightleftharpoons}} ABB \qquad \frac{d[AB]}{dt} = (k_{a1}[A][B] - k_{d1}[AB]) - (k_{a2}[AB][B] - k_{d2}[ABB])$$

$$\frac{d[ABB]}{dt} = (k_{a2}[AB][B] - k_{d2}[ABB])$$

Heterogeneous binding:

$$A + B \underset{k_{d1}}{\overset{k_{a1}}{\rightleftharpoons}} AB \qquad \frac{d[AB]}{dt} = (k_{a1}[A] - k_{d1})[AB]$$

$$A + C \underset{k_{d2}}{\overset{k_{a2}}{\rightleftharpoons}} AC \qquad \frac{d[AC]}{dt} = (k_{a1}[A] - k_{d1})[AC]$$

$$Y = [AB] + [AC] \text{ (Recorded signal)}$$

Formulas are reproduced from TraceDrawer Handbook Vers1_6.

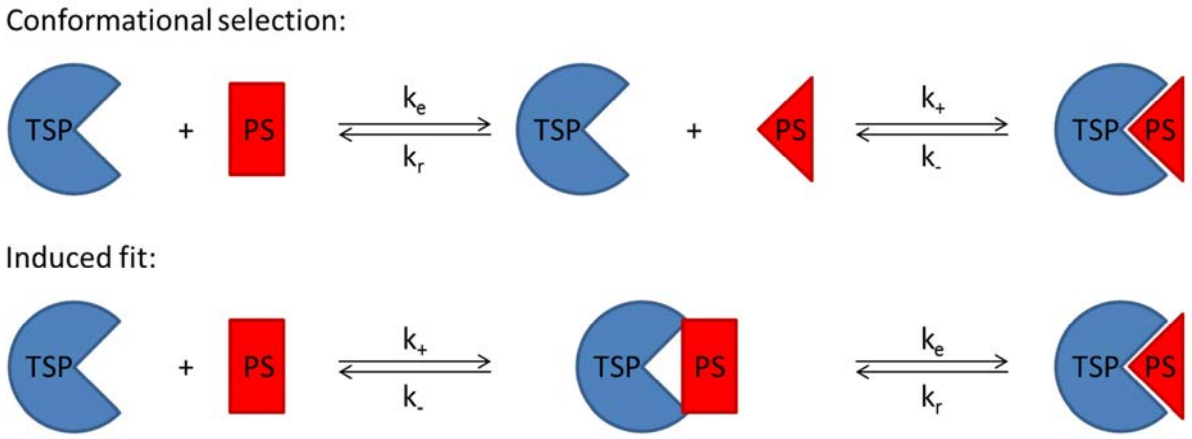

Fig. S 4: Schematic description of conformational selection and induced fit binding.
The presentation was adapted from Paul and Weikl, 2016.

Fitted parameters	k_e (s^{-1})	k_r (s^{-1})	k_- (s^{-1})	k_+ ($\mu M^{-1}s^{-1}$)	K_d (μM)	R^2
Conformational selection	0.14±0.056	11.03±1354	0.0038±0.0046	-	$6.71*10^{-3}$±$9.57*10^{-3}$	0.853
Induced fit	0.112±5241	0.0009±7602	0.1556±2361	0.577±1.78	$1.34*10^{-3}$±$3.19*10^{-2}$	0.775

Induced fit:

$$k_{obs} = k_e + k_r + \frac{1}{2}\gamma - \frac{1}{2}\sqrt{\gamma^2 + 4k_-k_r}$$

$$\gamma = -k_e - k_r + k_- + k_+(\delta - K_d)$$

$$\delta = \sqrt{([L]_0 - [P]_0 + K_d)^2 + 4[P]_0K_d}$$

$$K_d = \frac{k_-k_e}{k_+(k_e + k_r)}$$

Conformational selection:

$$k_{obs} = k_e + \frac{1}{2}\alpha - \frac{1}{2}\sqrt{\alpha^2 + \beta}$$

$$\alpha = k_r - k_e + \frac{k_-((2k_e + k_r)\delta + k_r([L]_0 - [P]_0 - K_d))}{2k_eK_d}$$

$$\beta = 2k_r(2k_e - k_- - \frac{k_-(\delta - [L]_0 + [P]_0)}{K_d})$$

$$K_d = \frac{k_-(k_e + k_r)}{k_+k_e}$$

With $[P]_0$=184 nM of Sf6 TSP_{EADA} and $[L]_0$ as the molar polysaccharide concentration estimated from an average molecular mass of 12 kDa. The model with figure and formulas was adapted from Paul and Weikl, 2016.

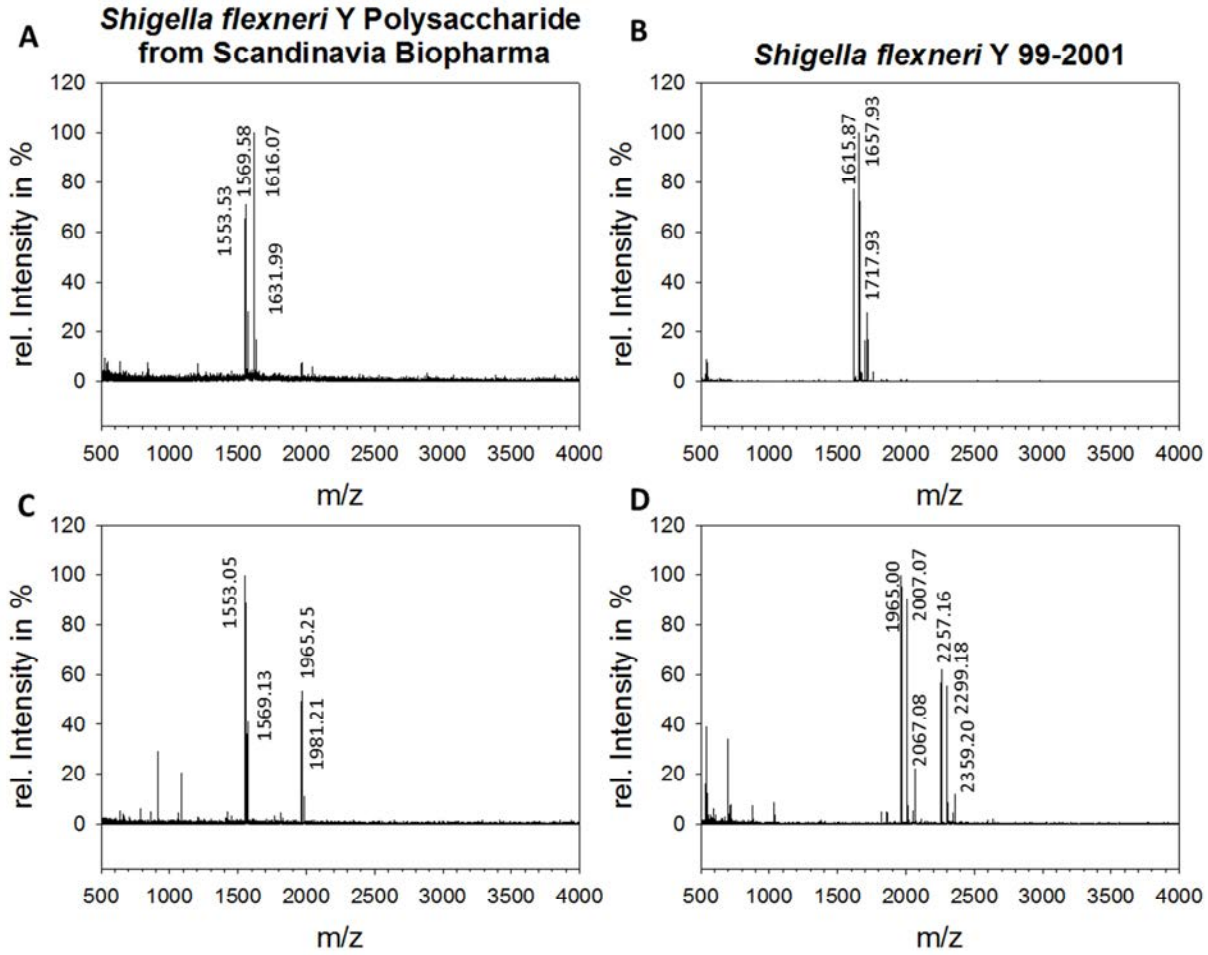

Fig. S 5: MALDI TOF mass spectra from Peak 1 and 2 of size exclusion seperation of SfY oligosaccharides.
Digestion products were separated in size exclusion chromatography resulting in three peaks as in Fig. 26A. Peak 2 was analyzed in MALDI-TOF MS, which resulted in the depicted peaks for SfY polysaccharide from Scandinavia Biopharma in **(A)** and from *S. flexneri* Y 99-2001 in **(B)**. The same was performed for Peak 1 with the corresponding origins in **C** and **D**. Masses which were calculated to correspond to the expected cleavage products are listed in Table S 4.

Table S 4: Molecular formula of SfY oligosaccharides for exact mass calculation

Oligosaccharide	Variation	Formula	Theoretical mass	m/z
Octasaccharide		C52O35H88N2	1300.52	
	+ Na+	C52O35H88N2Na	1323.51	1323.79
	+ K+	C52O35H88N2K	1339.48	1339.68
Octasaccharide-Ac (42)		C54O36H90N2	1342.53	
	+ Na+	C54O36H90N2Na	1365.52	1365.70
	+ K+	C54O36H90N2K	1381.49	
Decasaccharide (+2 Ram)		C64O43H108N2	1592.63	
	+ Na+	C64O43H108N2Na	1615.62	1615.83
	+ K+	C64O43H108N2K	1631.60	1631.99
Decasaccharide-Ac		C66O44H110N2	1634.64	
	+Na	C66O44H110N2Na	1657.63	1657.93
	+K	C66O44H110N2K	1673.61	
Decasaccharide-2Ac		C68O45H112N2	1676.65	
	+Na	C68O45H112N2Na	1699.64	
	+K	C68O45H112N2K	1715.62	
	+ACN	C68O45H112N2(CH3CN)	1717.68	1717.93
Dodecasaccharide		C78O52H131N3	1941.77	
	+ Na+	C78O52H131N3Na	1964.76	1965.25
	+ K+	C78O52H131N3K	1980.73	1981.21
Dodecasaccharide-Ac		C80O53H133N3	1983.78	
	+Na	C80O53H133N3Na	2006.77	2007.07
	+K	C80O53H133N3K	2022.74	
Dodecasaccharide-2Ac		C82O54H135N3	2025.79	
	+Na	C82O54H135N3Na	2048.78	
	+K	C82O54H135N3K	2064.76	
	+ACN	C82O54H135N3(CH3CN)	2066.82	2067.08
Tetrakaidecasaccharide		C90O60H151N3	2233.89	
	+Na	C90O60H151N3Na	2256.88	2257.16
	+K	C90O60H151N3K	2272.85	
Tetrakaidecasaccharide-Ac		C92O61H153N3	2275.90	
	+Na	C92O61H153N3Na	2298.89	2299.18
	+K	C92O61H153N3K	2314.86	
Tetrakaidecasaccharide-2Ac		C94O62H155N3	2317.91	
	+Na	C94O62H155N3Na	2340.90	
	+K	C94O62H155N3K	2356.88	
	+ACN	C94O62H155N3(CH3CN)	2358.93	2359.20

Masses are calculated by sisweb.com/referenc/tools/exactmass.html, 12.07.16.

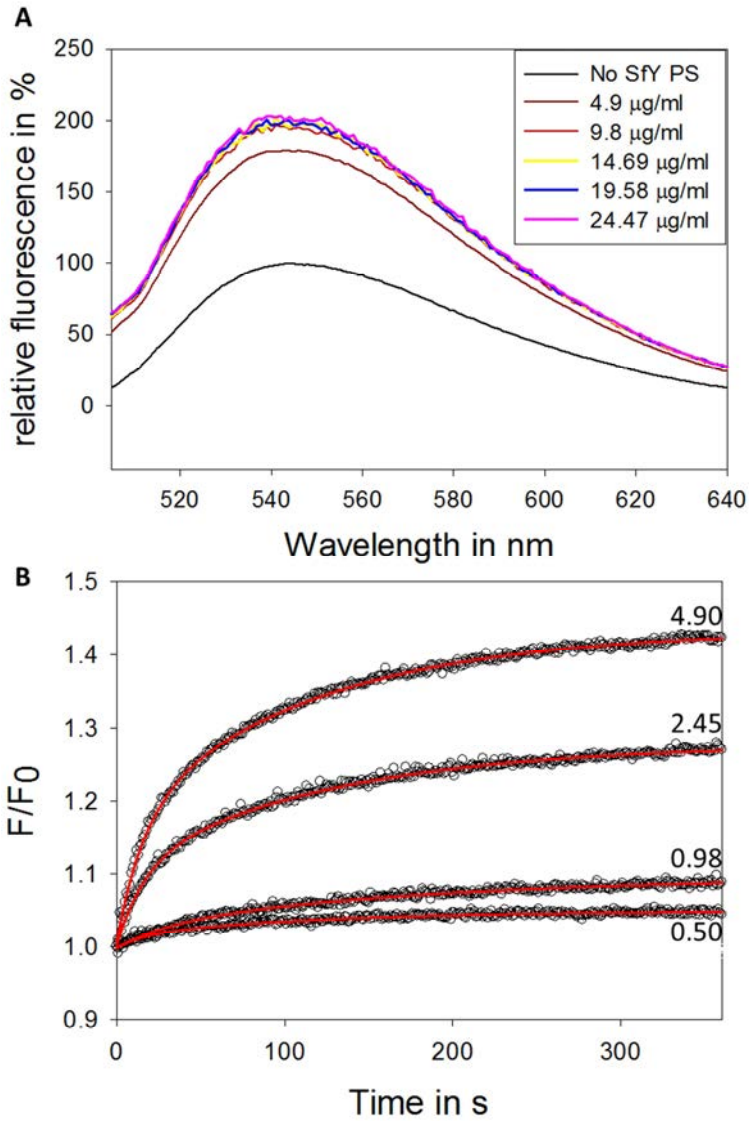

Fig. S 6: Fluorescence titration experiments with Sf6 TSP-Cys label
(A) Sf6 TSP-Cys-Label (181 µM) excited at 492 nm and emission messured between 500-650 nm with several SfY PS concentrations in the respective colors. **(B)** Kinetic traces of binding relaxations monitored at 540 nm between Sf6 TSP-Cys-Label with the respective SfY polysaccharide concentrations in µg/ml were fitted by a biexponential equation, which resulted in $k_{obs.}$

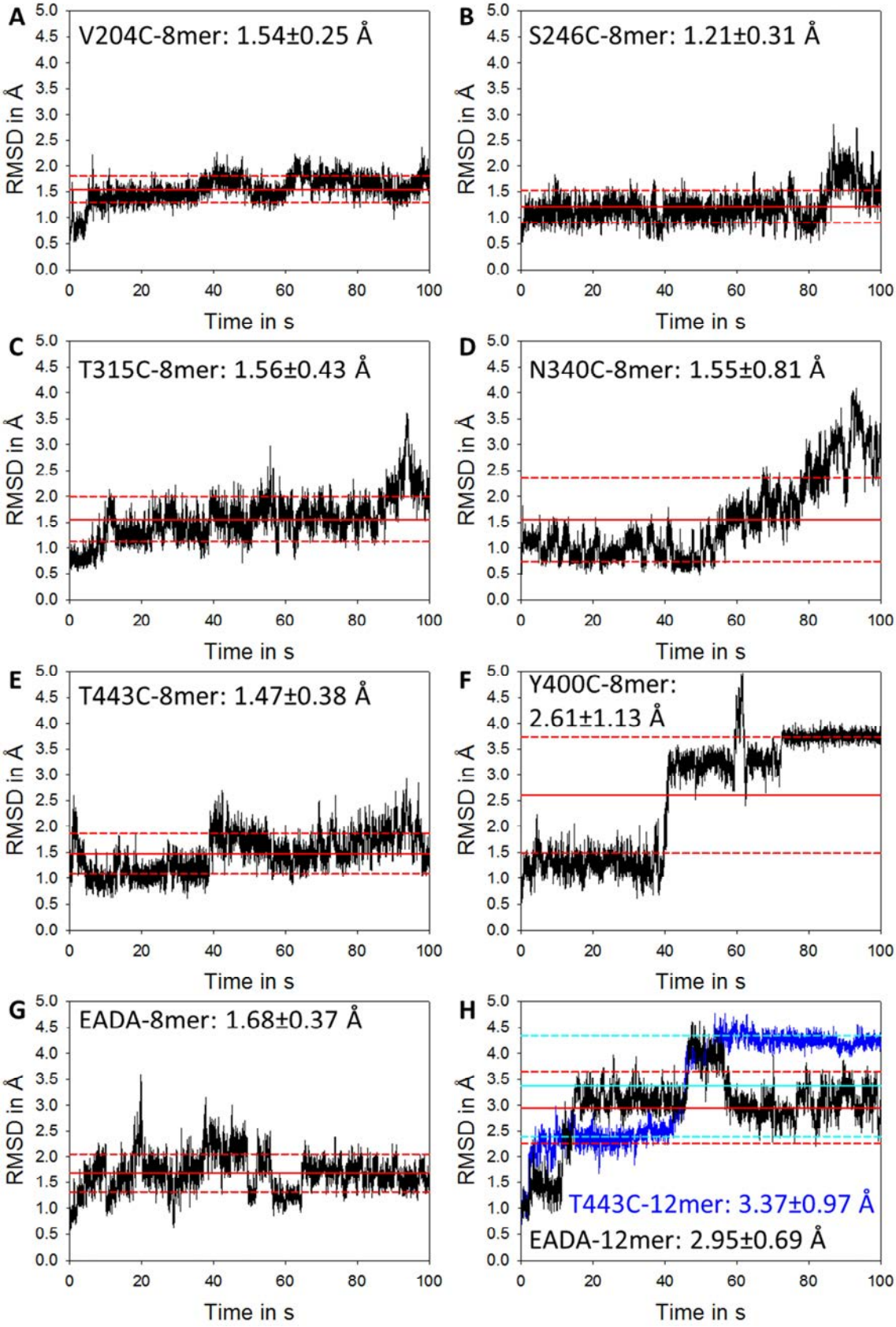

Fig. S 7: RMSD traces over 100 ns MD-simulation with octa- and dodecasaccharide ligand
Root mean square deviation of the ligand for octa- (8mer) and dodecasaccharide (12mer) for the various cysteine mutants of Sf6 TSP EADA with V204C (A), S246C (B), T315C (C), N340C (D), T443C (E), Y400C (F), EADA (G) with octasaccharide or EADA (black,H) and T443C (blue, H) with dodecasaccharide. The red solid line depicts the mean RMSD with the red dashed lines for one standard deviation (in cyan for T443C-12mer). Numbers depict the mean of the RMSD with one standard deviation.

Table S 5: Hydrogen bond occupancies of dodecasaccharide 100 ns simulations with Sf6 TSP_{EADA} and T443C

Protein	Ligand	EADA-12mer	T443C-12mer
ARG230-NH1	RAM1-O2	13.87 %	43.11 %
ARG230-NH2	RAM1-O2	10.98 %	19.76 %
GLU293-OE1	RAM1-O2	23.05 %	39.62 %
GLU293-OE2	RAM1-O2	24.85 %	52.89 %
ARG230-NH1	RAM1-O3	14.77 %	
GLU293-OE1	RAM1-O3	16.37 %	
GLU293-OE2	RAM1-O3	28.64 %	
ARG230-NH2	RAM1-O5		19.76 %
ASN340-ND2	NAG2-O7	11.88 %	
ASP247-OD1	RAM3-O3		30.74 %
SER246-O	RAM4-O3		77.15 %
ASN340-OD1	RAM4-O3	11.48 %	
SER246-O	RAM4-O4	13.77 %	
GLY339-O	RAM5-O4	35.03 %	
TYR400-OH	NAG6-N2	46.71 %	
GLN280-NE2	NAG6-O7		45.81 %
ASN455-OD1	NAG10-N2		21.76 %
ARG364-NH2	NAG10-O4	27.54 %	
ARG364-NH2	NAG10-O6	11.78 %	
ASN412-ND2	RAM11-O3		11.28 %
GLN444-O	RAM11-O3	10.38 %	
ASN455-OD1	RAM11-O3	41.12 %	
GLY413-N	RAM11-O4		10.68 %
CYS443-O	RAM11-O4		10.28 %
SER445-OG	RAM11-O4	11.78 %	
ASN455-OD1	RAM11-O4	63.87 %	
ASN508-OD1	RAM12-O1	19.36 %	10.38 %
GLN444-O	RAM12-O3	64.57 %	36.83 %
GLN444-O	RAM12-O4	65.17 %	29.94 %
ASN508-ND2	RAM12-O5	16.37 %	
Relative occupancy		**1.00**	**0.79**
Total number of hydrogen bonds		**22**	**15**

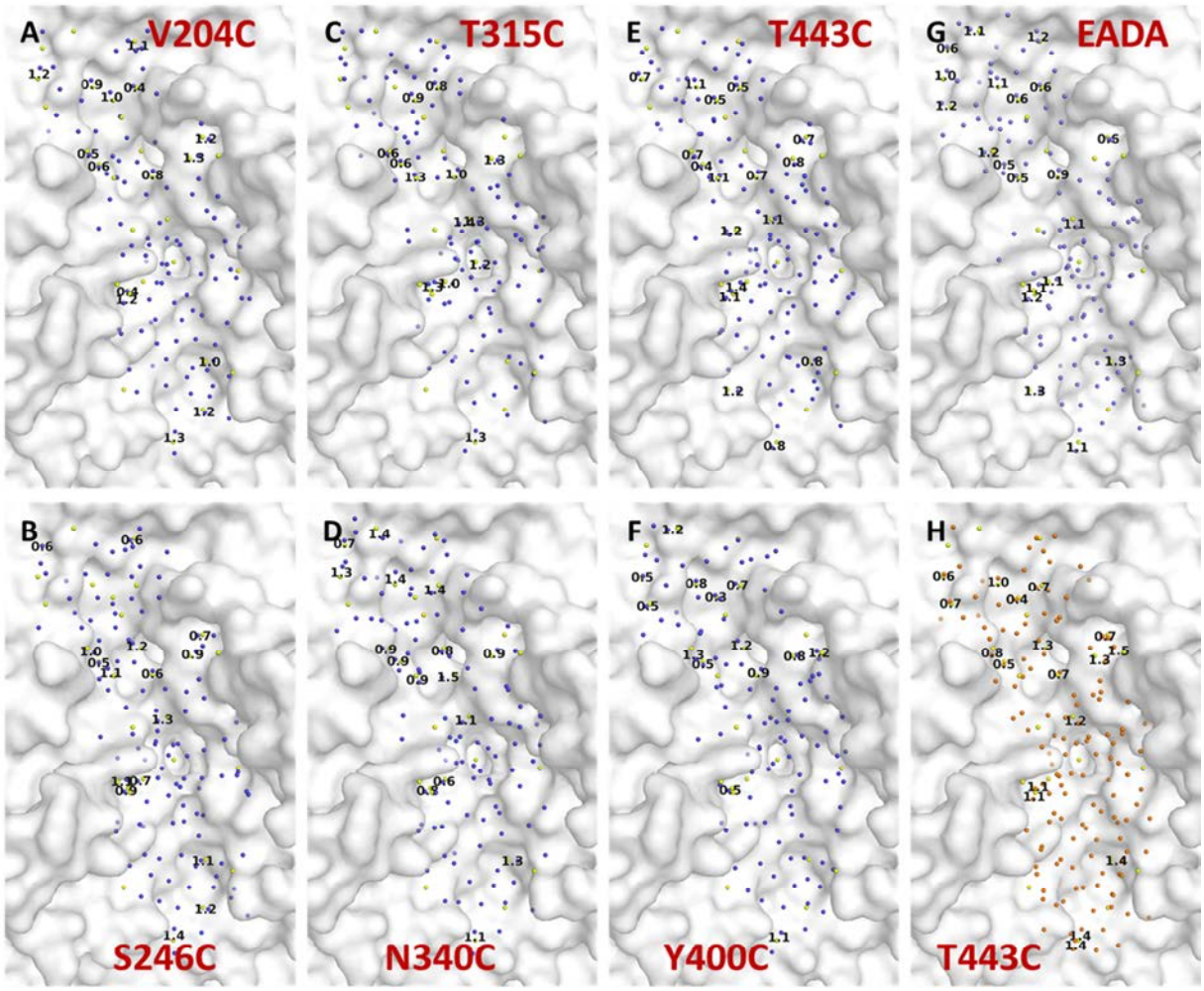

Fig. S 8: Water molecule positions derived by Mobywat for cysteine mutants of Sf6 TSP EADA
Simulated water positions from 100 ns simulations with octa- (blue balls) or dodecasaccharide ligand (orange balls) in comparison with crystallographic water positions (yellow balls) depicted with the given distances of water molecules, which could be reproduced by the simulation. The cut-off for superimposed water molecules was chosen at 1.5 Å.

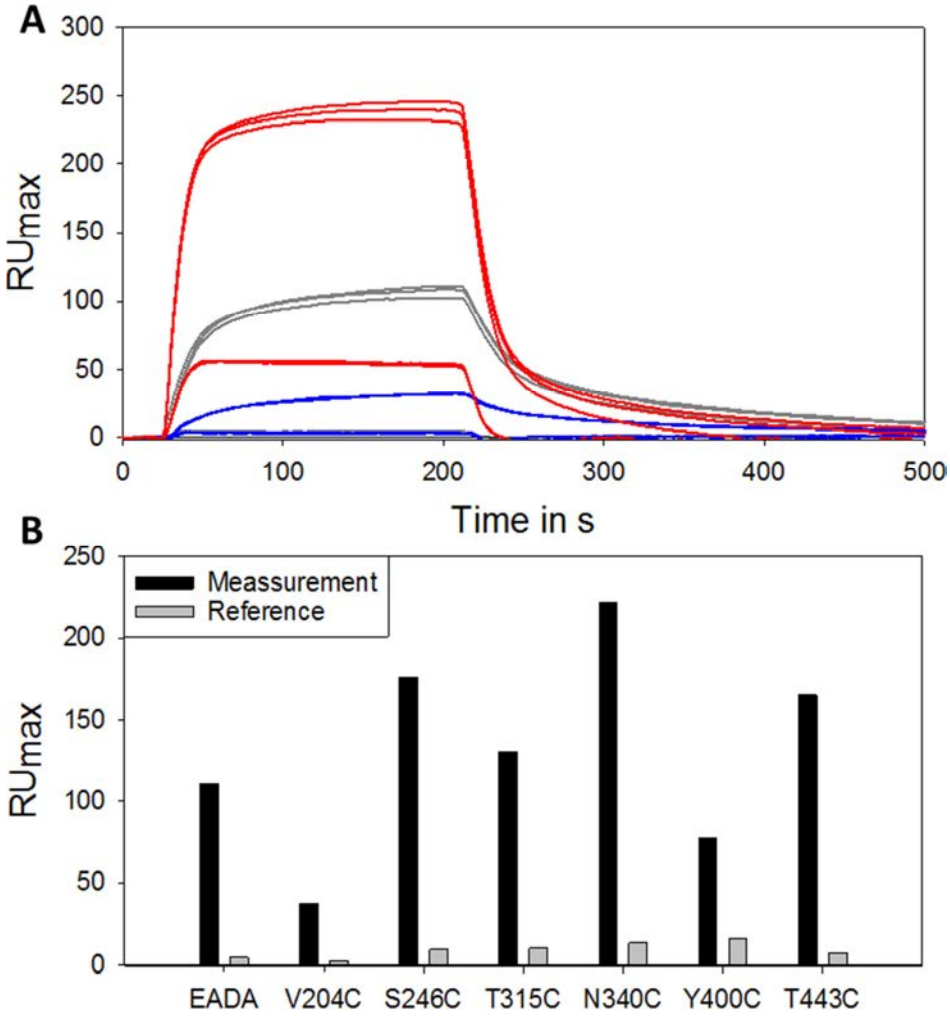

Fig. S 9: SPR sensogram of measurement and reference channel
(A) Sensograms are depicted of Sf6 TSP_{EADA} injections in three concentrations: 0.08 µM (blue), 0.8 µM (gray) and 8 µM (red) measured as triplicates on the second chip prepared with oxidized SfY polysaccharide. The upper curves correspond to the measurement channel and the lower curves to the reference channel, where no immobilization was performed. The percent of unspecific signal ranges for 0.08 µM concentrations around 12 %, for 0.8 µM at 4 % and for 8 µM at 24 % for all Sf6 TSP_{EADA} cysteine mutants except for Y400C, where the percent is higher. **(B)** The maximal responses measured at the end of 3 min. injections on the second chip are depicted for all Sf6 TSP_{EADA} cysteine mutants on the measurement channel (black) and the reference channel (gray). The unspecific signal on the reference channel ranges around 4 % for all protein variants except for Y400C, where the reference is increased to 20 %. However, the total signal on the measurement channel is lower for Y400C compared to most of the other cysteine mutants.

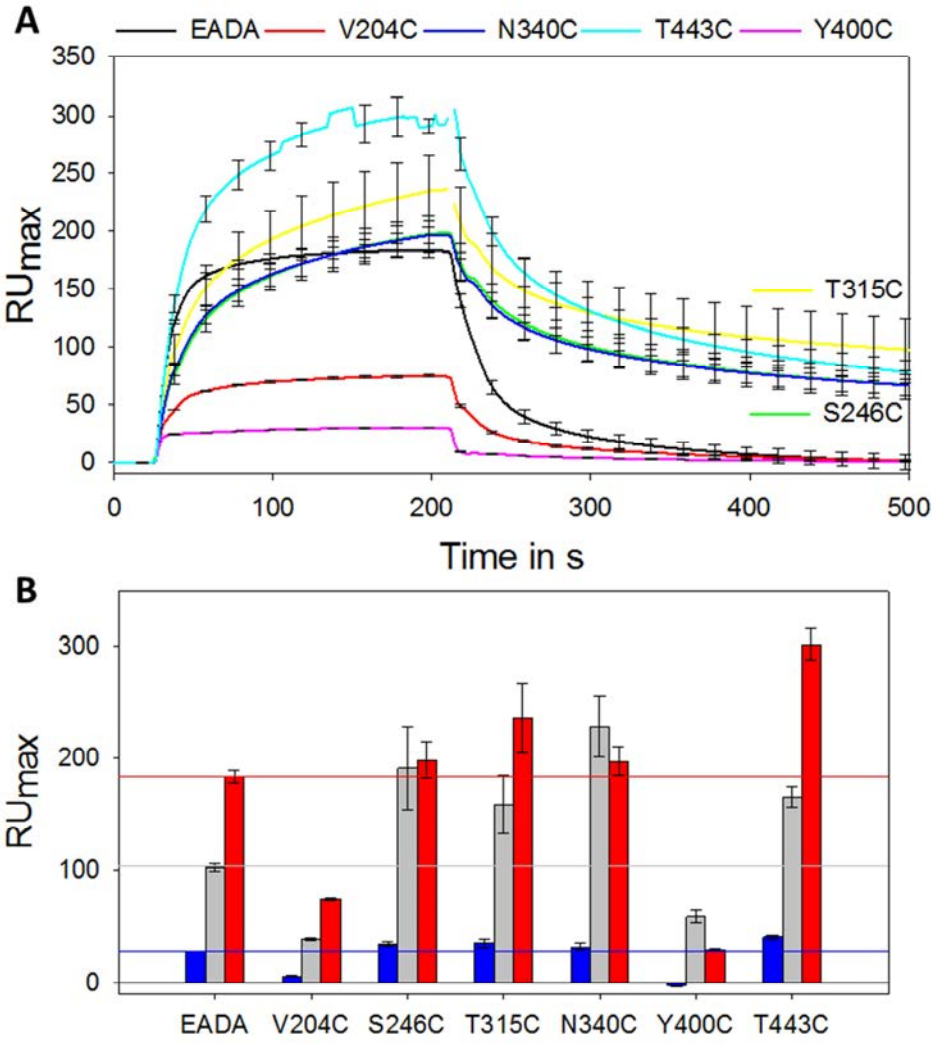

Fig. S 10: Cysteine mutant interaction curves and response unit maxima on Chip2
(A) Sensograms of 8 µM protein injections over the surface with immobilized SfY polysaccharide of Sf6 TSP_{EADA} (black) and each of the cysteine mutants: V204C (red), S246C (yellow), T315C (green), N340C (blue), Y400C (purple) and T443C (cyan). **(B)** Maximal response units at the end of 3 min injection runs with Sf6 TSP_{EADA} cysteine mutants injected in triplicates for each protein concentration with 0.08 (blue), 0.8 (gray) and 8 µM (red). The horizontal lines mark the maximal response of Sf6 TSP_{EADA} with the respective concentrations.

Table S 6: Binding site residue side chain conformations analyzed by superimposed crystal structures (Part 1)

Residues	#conf.	Bulk[a]	2nd	WT	D339N	D339A	D339N E372A	D339A E372A	E372A	D339N E372Q	D339A E372Q	E372Q
His212	3	Bound	Unbound χ1/χ2 ±7°		Apo NE2 0.5A					Apo NE2 0.5A		
Gln242	3	Bound		Apo χ3 -96° NE2 2.0 Å		Apo χ3 45° NE2 0.9 Å						
Leu282	3	Bound			Apo χ2 -100° CD2 2.6 Å					Apo χ2 15° CD2 0.5 Å		
Asn284	2	Bound	Unbound χ2 -15° NE2 0.5 Å									
Thr311	2	Bound		Apo χ1 -121° OG1 2.2 Å							Apo χ1 -121° OG1 2.2 Å	Apo χ1 -121° OG1 2.2 Å
Asn315	4	Bound					Apo, χ1 -87° OD1 3.3 Å	Apo χ2 20° OD1 0.5 Å	Apo, χ1 -87° OD1 3.3 Å			
							Apo χ2 12° OD1 0.7 Å		Apo χ2 12° OD1 0.7 Å			
339	5	Bound Asp	Bound Asn χ2 -9.6°	Apo χ2 25° OD2 1.0 Å	Apo/6mer χ2 55.5° OD2/ND2 1.6 Å		Apo χ2 55.5° OD2/ND2 1.6 Å		Apo χ2 25° OD2 1.0 Å	Apo χ2 62.4° OD2/ND2 1.8 Å		Apo χ2 25° OD2 1.0 Å
Ser341	2	Bound					5/6mer χ1 -107° OG 2.4 Å	5/6mer χ1 -107° OG 2.4 Å	5mer χ1 -107° OG 2.4 Å			
Tyr344	3	Bound	Unbound χ2 -10° CE1 0.6 Å							Apo χ2 61° CE2 2.7 Å	Apo χ2 61° CE2 2.7 Å	Apo χ2 61° CE2 2.7 Å
Asn346	2	Bound				Apo χ2 8°	Apo χ2 8°	Apo χ2 8°	Apo χ2 8°	Apo χ2 8°	Apo χ2 8°	Apo χ2 8°
372	5	Bound Glu	Bound Gln χ3 127° OE1 2.5 Å	Apo χ3 15° OE2 0.8 Å	Apo/6mer χ3 15°; OE2 0.8 Å	Apo χ3 15° OE2 0.8 Å				Apo χ1 102° OE2/NE2 3.6 Å	Apo χ1 102° OE2/NE2 3.6 Å	Apo χ1 102° OE2/NE2 3.6 Å
												5mer χ3 180° to bound Gln
His374	7	Bound		Apo χ2 18° CD2 1.2 Å	Apo χ2 -37° CD2 1.1 Å	Apo χ2 -37° CD2 1.1 Å	Apo χ2 72.5° CD2 1.5 Å	Apo χ2 72.5° CD2 1.5 Å	Apo χ2 72.5° CD2 1.5 Å	Apo χ2 8.5° CE 1.1 Å	Apo χ2 8.5° CE 1.1 Å	Apo χ2 8.5° CE 1.1 Å
				5mer χ2 54° CD2 1.3 Å	5/6mer χ2 54° CD2 1.3 Å	5mer χ2 54° CD2 1.3 Å						
				6mer χ2 17° CD2 0.8 Å		6mer χ2 17° CD2 0.8 Å						
Tyr393	3	Bound						Apo χ2 -161° OH 1.4 Å		5mer χ2 -10°	5mer χ2 -10°	5mer χ2 -10°
Val395	2	Bound		6mer χ1 142° CG1 2.7 Å		Apo/6 χ1 142° CG1 2.7 Å						

a The conformational changes are measured from the bulk state. The highest torsion angle change is listed together with the longest distance of corresponding atoms.

Table S 7: Binding site residue side chain conformations analyzed by superimposed crystal structures (Part 2)

Residues	# conf.	Bulk[a]	2nd	WT	D339N	D339A	D339N E372A	D339A E372A	E372A	D339N E372Q	D339A E372Q	E372Q
His397	6	Bound		Apo χ2 -19° CD2 0.6 Å	Apo χ1 8.8° CE1 0.7 Å	Apo χ1 8.8° CE1 0.7 Å	Apo χ2 16° CD2 0.7 Å	Apo χ2 16° CD2 0.7 Å	Apo χ2 16° CD2 0.7 Å	Apo χ2 -16.8° CD2 0.9 Å	Apo χ2 -16.8° CD2 0.9 Å	Apo χ2 -16.8° CD2 0.9 Å
				5mer χ2 -8.6°	5mer χ2 -8.6°	5mer χ2 -8.6°						
					6mer χ2 -19° CD2 0.6 Å							
Glu400	2	Bound								Apo χ3 42° OE1 1.6 Å		
Ser429	2	Bound			Apo χ1 85.6° CB 1.1 Å							
Asp430	4	Bound		Apo χ2 -9.7°	Apo χ2 40° OD2 1.3 Å	Apo χ2 -9.7°	Apo χ2 -9.7°	Apo χ2 -9.7°	Apo χ2 -9.7°	Apo χ2 -9.7°	Apo χ2 -9.7°	Apo χ2 -9.7°
					Apo OD1 0.6 Å							
Asn471	7	6mer	5mer χ2 21° OD1 1.1 Å		6mer χ2 21° OD1 1.1 Å	Apo χ2 -65.5° ND2 1.9 Å	Apo χ1 -96.7° ND2 3.5 Å	Apo χ2 -38.5° OD1 1.9 Å	Apo χ1 -96.7° ND2 3.5 Å		Apo χ2 58.5°; OD1 1.4 Å	

a The conformational changes are measured from the bulk state. The highest torsion angle change is listed together with the longest distance of corresponding atoms.

Table S 8: Potential hydrogen bonds derived from crystal structure comparison between binding site side chains and the carbohydrate ligand with distances in Å.

Hydrogen bonds	WT-6	WT-5	DA-6	DA-5	DN-6	DN-5	DNEA-6	DNEA-5	DAEA-6	DAEA-5	EA-6	EA-5	DNEQ-6	DNEQ-5	DAEQ-6	DAEQ-5	EQ-6	EQ-5
H212ND1-RAMO4	2.7	2.7	2.7	2.7	2.7	2.7	2.7	2.7	2.7	2.7	2.7	2.7	2.7	2.7	2.7	2.7	2.7	2.7
Q242Nd2-RAMO4	3.0	3.1	3.1	3.1	3.0	3.1	3.1	3.1	3.1	3.1	3.2	3.0	3.1	3.1	3.1	3.1	3.1	3.0
W308NE1-NAG1O7	2.9	3.0	2.9	2.9		3.0	3.0	2.9	2.9	3.1	3.0	3.0	3.1	3.0	3.1	2.9	3.0	3.0
N315OD1-NAG2O3	2.6	2.6	2.7	2.6	2.7	2.6	2.7	2.7	2.6	2.6	2.6	2.7	2.6	2.6	2.7	2.6	2.7	2.7
N339OD1-NAG1O1	3.5					3.5		3.4					3.5					3.0
N339OD2-NAG1N2	3.0	2.9									2.9	2.9					2.9	2.9
N339ND2-NAG1O1							3.0	3.4										
S341OG-NAG1O1	3.4	3.4							3.2		2.8	2.6	3.2	3.0	3.2			
S341OG-NAG1O7											3.3					3.0		
Y344OH-GALO2	3.1	3.0	3.0	3.0	3.1	3.0	2.9	2.8	2.9	2.9	3.0	2.9	2.9	2.9	2.9	2.9	2.9	2.9
N346ND2-NAG2O4	2.8	2.8	2.8	2.8	2.9	2.8	2.9	2.7	2.8	2.9	2.9	2.8	2.9	2.9	2.9	2.9	2.9	2.9
E372OE2-NAG1O1	2.6	2.6	2.5	2.4		2.6							3.1	2.6	2.7	2.8	2.7	
E372OE1-NAG1O6		2.8		2.8		2.8												2.8
E372NE2-NAG1O5													3.3	3.2	3.4	3.4		3.2
E372NE2-NAG1O6														2.4		2.6		
H374NE2-NAG1O6				3.3				2.9						3.0		3.0		3.0
H397Ne2-NAG1O4	2.8	2.6	2.7	2.6		2.5	2.7	2.6	2.8	2.5	2.8	2.7	2.6	2.7	2.7	2.7	2.7	2.7
H397NE2-NAG1O6										3.3								
E400OE2-NAG2N2	2.9	2.8	3.0	2.9	2.9	2.9	2.9	2.9	2.8	2.9	2.8	2.9	2.8	2.8	2.8	2.9	2.9	2.9
D430OD2-GLC2O6	2.8		2.8				2.7		2.8		2.7		2.7		2.8		2.8	
G468O-GLC2O4	3.0		3.0				2.8		2.8		2.9		2.9		2.8		2.7	
P469O-GLC2O3	2.7		2.7				2.9		2.9		2.8		2.8		2.9		2.9	
N471ND2-NAG1O4	2.9	2.8	3.0	2.8		2.8	3.0	2.8	2.9	3.0	2.9	2.8	3.1	2.9	2.9	2.8	3.2	2.8
Total number	**16**	**13**	**13**	**12**	**6**	**12**	**13**	**12**	**13**	**10**	**15**	**11**	**16**	**14**	**15**	**14**	**14**	**14**

Table S 9: Potential hydrogen bonds within the binding site residues with distances in Å

Hydrogen bonds	WT-apo	WT-6	WT-5	DA-apo	DA-6	DA-5	DN-apo	DN-6	DN-5	EA-apo	EA-6	EA-5	EQ-apo	EQ-6	EQ-5
H212O-Q242NE2		3.1	3.1	3.1	3.1	3	2.7	3.1	3.1	2.8	3.1	3.1	2.8	3.1	3.1
Q242OE1-N284ND2		3.2	3.2	3.1	3.2	3.2	3.2	3.2	3.2	3.3	3.2	3.1		3.1	3.1
N284ND2-T311OG1	2.9	3.1	3.3	2.9	3.2	3.2	2.9	3.2	3.3	3.1	3.2	3.3	3.1	3.3	3.3
T311O-T311OG1	2.9			2.9			2.9								
T307O-S341N	2.9	2.9	2.8	3.0	2.9	2.8	2.9	2.9	2.8	2.9	2.9	2.9	2.8	2.9	2.9
N315N-N315OD1				2.9			2.9			2.8			2.9		
N315ND2-A399O				3.2			3.5			3.5					
N315OD1-N346ND2	3.5	3.4	3.3	3.5	3.3	3.2	3.5	3.3	3.3	3.3	3.3	3.3	3.3	3.2	3.3
N346ND2-E400OE2	3.2			3.4									3.4		
E400OE2-Y344OH	2.5	2.6	2.6	2.4	2.6	2.6	2.6	2.5	2.6	2.6	2.5	2.6	2.3	2.6	2.6
Y344OH-H374ND1	3.1	3.1			3.1						3.2	3.1		3.1	3.2
H374NE2-E372OE1	2.6	2.6	2.6	2.5	2.6	2.7	2.6	2.6	2.7						2.8
E372OE2-S341OG	2.6	2.5	2.5	2.5	2.6	2.6	2.5	2.5	2.6				2.6	2.8	2.7
S341OG-H340O	3.3	3.3	3.2	3.3	3.3		3.3		3.2		3.4	3.3	3.3		
H340N-D339OD1	2.8	3.0	3.0					3.1	3.1	3.0	2.9	2.9	2.9	2.9	2.9
H397O-E400N	3.0	2.8	2.8	2.8	2.9	2.9	2.7	2.9	2.9	2.9	2.9	2.9	2.9	2.9	2.9
L427O-G468N	3.2	3.0	3.2	3.2	3.0	3.2	3.2	3.2	3.2	3.2	3.0	3.1	3.1	3.0	3.2
N429O-Q478NE2				3.5			3.2								
Q478OE1-Q478N	2.7	2.8	2.8	2.7	2.8	2.7	2.6	2.8	2.7	2.7	2.8	2.8	2.7	2.8	2.8
Total number	**14**	**14**	**13**	**17**	**13**	**11**	**16**	**13**	**13**	**12**	**12**	**12**	**15**	**12**	**13**

Hydrogen bonds	DNEA-apo	DNEA-6	DNEA-5	DAEA-apo	DAEA-6	DAEA-5	DNEQ-apo	DNEQ-6	DNEQ-5	DAEQ-apo	DAEQ-6	DAEQ-5
H212O-Q242NE2	2.8	3.1	3.0	2.9	3.0	3.1	3.5	3.1	3.1	2.9	3.1	3.1
Q242OE1-N284ND2	3.3	3.1	3.1	3.3	3.1	3.1		3.2	3.1	3.5	3.2	3.1
N284ND2-T311OG1	3.0	3.3	3.3	3.1	3.3	3.3	3.5	3.3	3.2	3.1	3.3	3.2
T311O-T311OG1	2.9			2.8						2.8		
T307O-S341N	2.9	2.9	2.9	3.0	2.9	2.9	2.9	2.8	2.8	2.8	2.8	2.8
N315N-N315OD1							2.9					
N315ND2-A399O	3.4			3.5			3.4					
N315OD1-N346ND2	3.5	3.3	3.2	3.4	3.3	3.2	3.5	3.3	3.3	3.5	3.3	3.3
N346ND2-Y344OH										3.3		
E400OE2-Y344OH	2.4	2.5	2.5	2.4	2.5	2.6	2.8	2.5	2.6	2.1	2.5	2.6
Y344OH-H374ND1		3.1	3.1		3.3	3.2		3.1	3.2		3.1	3.2
E372OE2-S341OG								2.5	2.6	2.7	2.6	2.7
S341OG-H340O	3.5	3.2		3.5	3.4		3.3			3.4		
H340N-D339OD1	2.9	3.0	2.9				2.9	2.9	3.0			
H397O-E400N	3.0	2.9	2.8	2.9	2.9	2.9	3.0	2.9	2.8	2.8	2.9	2.8
L427O-G468N	3.3	3.1	3.2	3.2	3.0	3.1	3.1	3.0	3.2	3.2	3.0	3.2
N429O-Q478NE2							3.2					
Q478OE1-Q478N	2.6	2.8	2.6	2.6	2.7	2.7	3	2.8	2.7	2.7	2.7	2.7
Total number	**14**	**12**	**11**	**12**	**11**	**10**	**13**	**12**	**12**	**15**	**11**	**11**

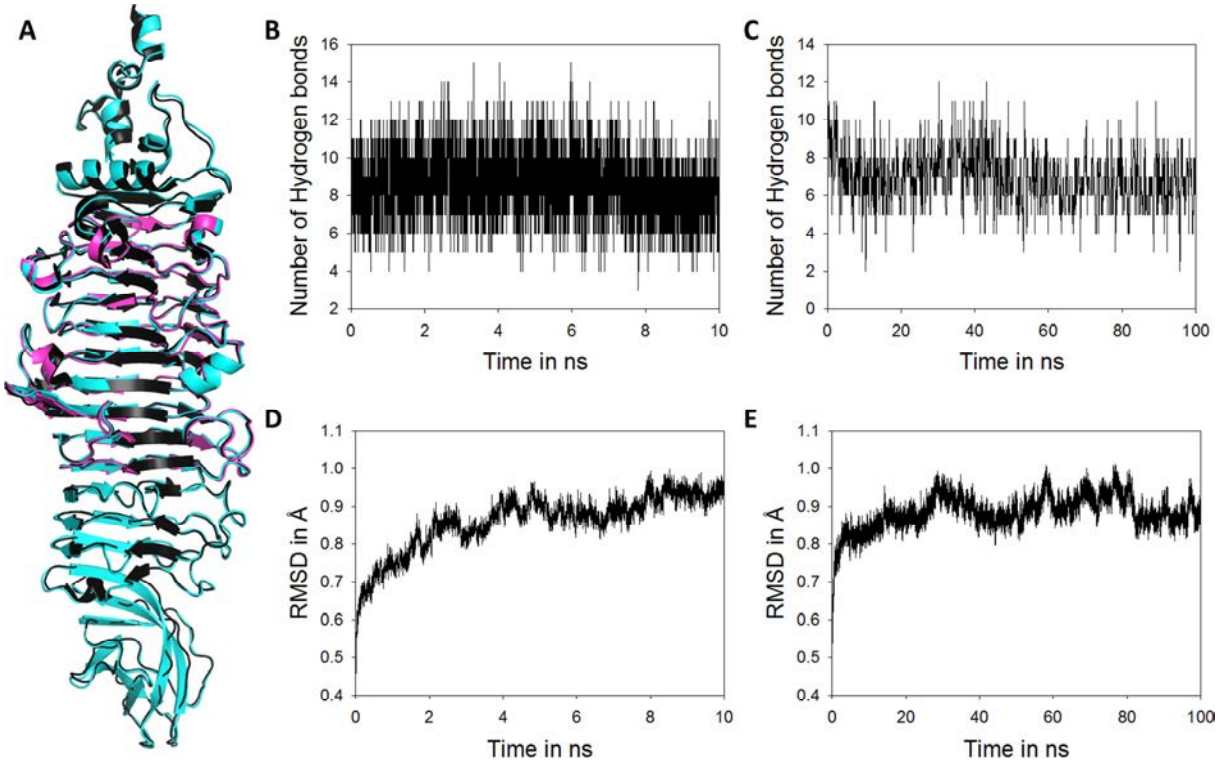

Fig. S 11: Trajectories of HK620 TSP from 10 and 100 ns simulations
(A) Overlay of HK620 TSP wild type hexasaccharide complex crystal structure (black backbone) with the end frames of a 100 ns simulation (cyan) and the truncated structure from the end frame of a 10 ns simulation (pink). **(B)** Hydrogen bonds between protein and ligand during 10 ns simulation and 100 ns simulation **(C)**. Root mean square deviation is shown for all protein atoms without hydrogens during 10 ns **(D)** and 100 ns simulation **(E)**.

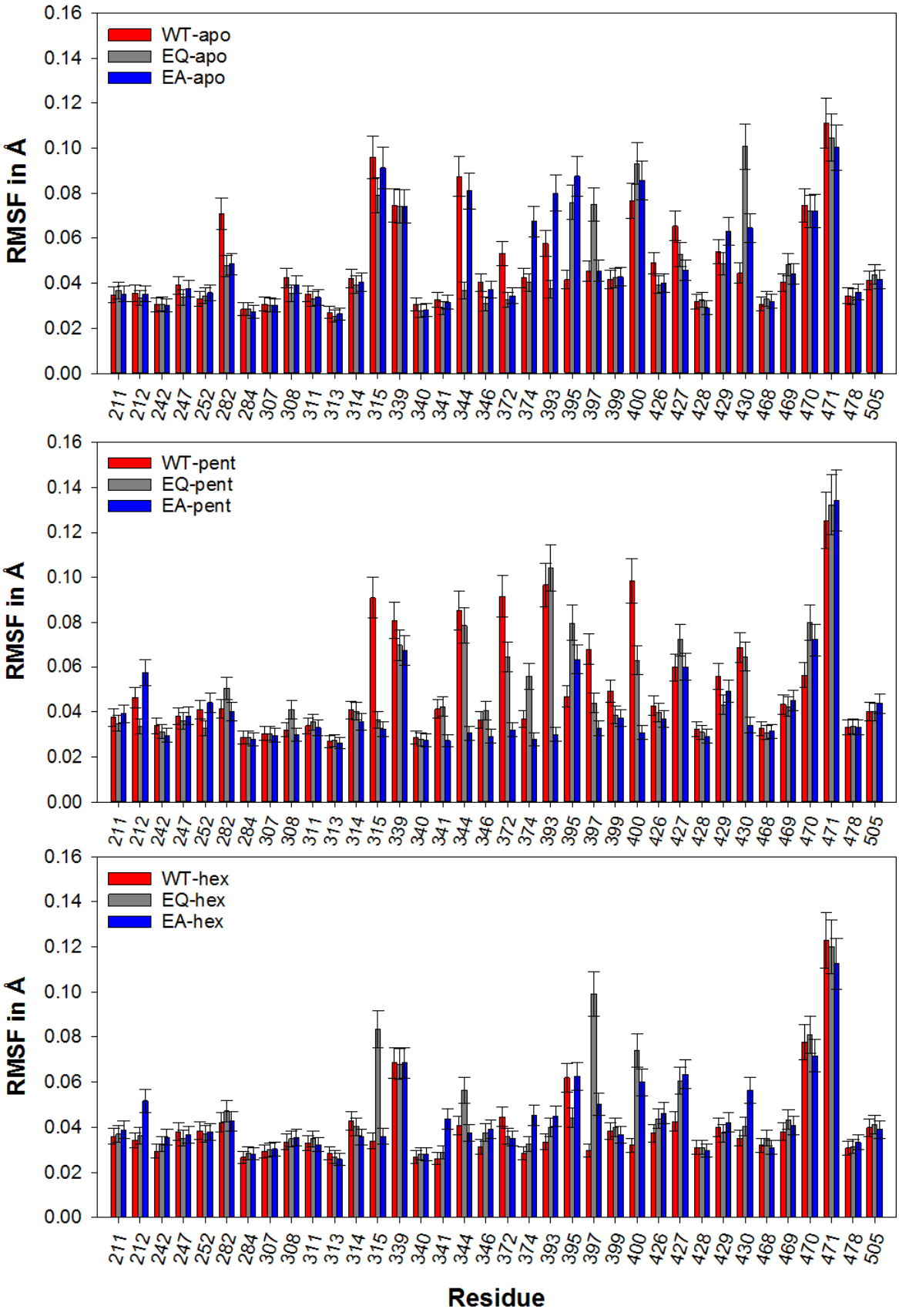

Fig. S 12: RMSF of all binding site amino acid residues comparing the mutations.
RMSF were derived from 10 ns simulations for whole residues with the *g_rmsf* of Gromacs/4.5.5 within 5 Å of the wild type hexamer ligand in the referred crystal structure. Error bars correspond to 10 % of the RMSF.

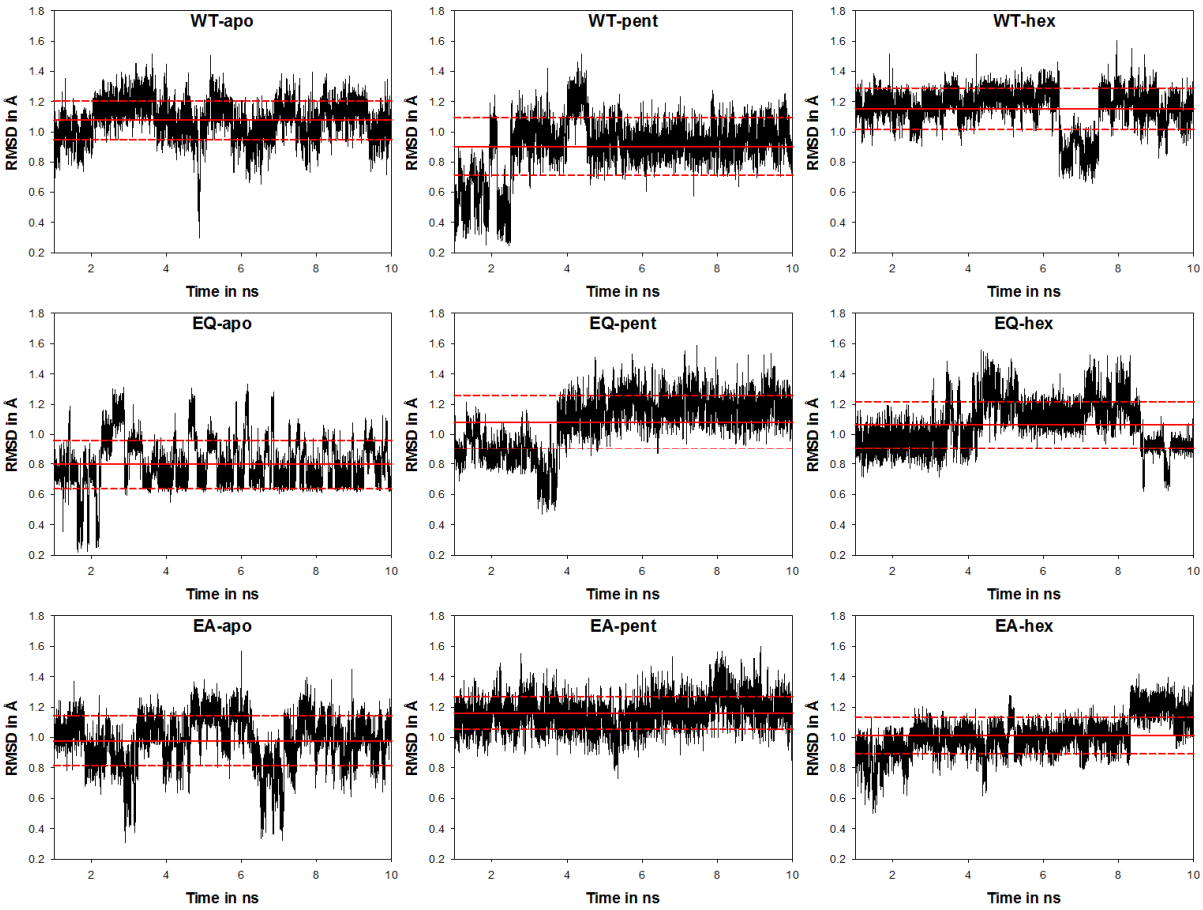

Fig. S 13: Time-dependent fluctuation of RMSD of the glucose binding pocket (residues 469-472) of 1-10 ns simulation.
RMSD were calculated with the *g_rms* tool of the Gromacs/4.5.5 package. One nanosecond at the beginning of each simulation was removed to ensure proper equilibration of fluctuations. Red solid lines show the mean of RMSD with red dashed lines at one standard deviation.

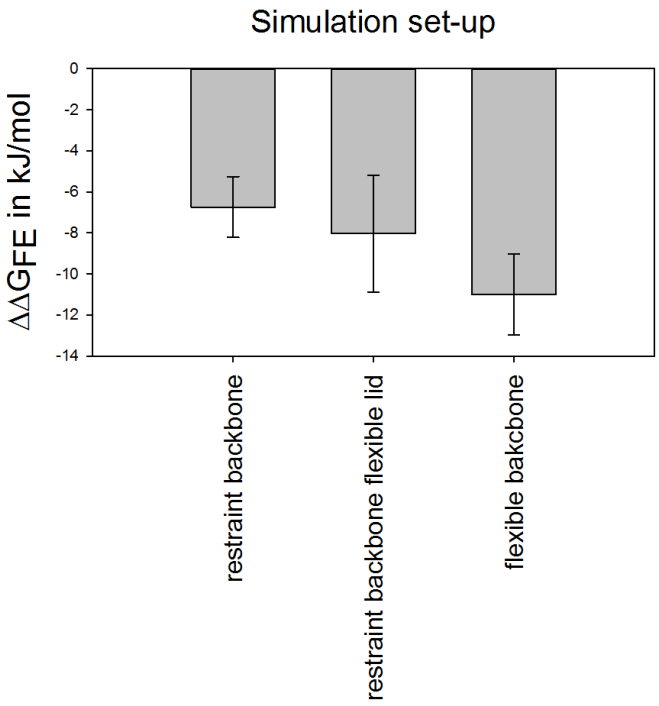

Fig. S 14: Evaluation of free energy calculation simualtion set-ups
The free energy calculations on the water molecule position 0 of E372Q-apo were repeated in five replicates for the three conditions: completely restrained backbone, a restrained backbone, where the lid of the glucose binding groove was kept flexible or a completely flexible backbone without any restraints. The standard deviation of this five replicate measurements resulted in the depicted error bars.

Abbreviations

Å	Ångstrøm 10^{-10} m
aa	amino acid
AcN	acetonitrile
AEBSF	4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride
AIM	auto induction medium
amp	ampicillin
AP	alkaline phosphatase
APS	ammonium persulfate
BAR	Bennett acceptance ratio
BCA	bicinchoninic acid
BCIP	5-bromo-4-chloro-3-indolyl phosphate
B-factor	Debye–Waller factor
BSA	bovine serum albumin
CadA	inducible lysine decarboxylase (*E. coli*)
CBM	carbohydrate binding module
CD	circular dichroism
cfu	colony forming unit
CMD	carboxymethyl dextran
ConA	Concanavalin A
CV	column volume
DA	HK620 TSP D339A
Da, kDa	Dalton g/mol
DAEA	HK620 TSP D339A E372A
DAEQ	HK620 TSP D339A E372Q
ddH2O	pure lab water
DEAE	diethylaminoethyl
DHB	2,5-dihydroxybenzoic acid
DMA	N,N-dimethylaniline
DMF	dimethylformamide
DMSO	dimethyl sulfoxide
DN	HK620 TSP D339N
ΔN	N-terminally truncated form of tailspike proteins
DNA	deoxyribonucleic acid
DNase I	Desoxyribonuklease I
DNEA	HK620 TSP D339N E372A
DNEQ	HK620 TSP D339N E372Q
dNTP	deoxynucleoside triphosphate

dsDNA	double-stranded DNA
DTT	dithiothreitol
E. coli	*Escherichia coli*
EA	HK620 TSP E372A
EADA	Sf6 TSP E366A D399A
EDC	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide
EDTA	ethylenediaminetetraacetic acid
e. g.	example given
EIEC	enteroinvasive *Escherichia coli*
ELISA	enzyme-linked immunosorbent assay
ELITA	ELISA-like tailspike adsorption
EptC	phosphoethanolamine transferase
EQ	HK620 TSP E372Q
EQDN	Sf6 TSP E366Q D399N
ER	endoplasmatic reticulum
et al.	inter alia
FAB	fragment antigen binding
FE	free energy
Fig.	figure
fw	forward
Gal	D-galactose
Glc	D-glucose
GlcA	D-glucuronic acid
GlcNAc	D-N-acetylglucosamine
HPAEC	high performance anion exchange chromatography
HD	high definition
HK620	*Podovirus* HK620
HK620 TSP	tailspike protein of HK620
HPLC	high performance liquid chromatography
HRP	horseradish peroxidase
ial	invasion-associated locus gene
IcsA	outer membrane protein autotransporter in *Shigella*
IDa	all-identity based prediction list
IDe	elitist-identity based prediction list
IL-1	Interleukin-1
ipa	invasion plasmid antigen genes
IPTG	isopropyl β-D-1-thiogalactopyranoside
ITC	isothermal titration calorimetry
IT-TI	independent trajectory thermodynamic integration
kb	kilo bases
kcal	kilocalorie, 4.2 kJ

k_{cat}	catalytic rate constant
K_D	dissociation constant
k_{obs}	apparent rate constant
k_{ass}/k_{diss}	association/dissociation rate constant
Kdo	3-deoxy-D-manno-oct-2-ulosonic acid
L,D-Hep	L-glycero-D-mannoheptose
LB	lysogeny broth
LPS	lipopolysaccharide
LptO	outer membrane protein
MALDI-TOF	matrix-assisted laser desorption/ionization-time of flight
mAU	milli absorbance units
MD	molecular dynamics
MER	merged-based prediction list
MES	2-(N-morpholino)ethane sulfonic acid
MetOH	methanol
mIgG	mouse Immunoglobulin G
M_w	molecular mass
MWCO	molecular mass cut off
Mxi	component of the Type III secretion machinery in *Shigella*
NAG	D-N-acetylglucosamine
NaP	sodium phosphate buffer
NBT	nitro blue tetrazolium
NEB	New England Biolabs
NHS	N-Hydroxysuccinimide
NMR	nuclear magnetic resonance
NPT	isothermal-isobaric ensemble
NVT	canonical ensemble
N-WASP	Neural Wiskott Aldrich Syndrome protein
OD	optical density
Omp	outer membrane protein
P22	*Podovirus* P22
P22 TSP	tailspike protein of P22
PBS	phosphate buffered saline
PBST	phosphate buffered saline with Tween
PCR	polymerase chain reaction
PDB	Protein Data Bank
pdbID	Protein Data Bank identity code
PEtN	phosphoethanolamine
PMMA	poly(methyl methacrylate)
POS	position based identity list
PPEtN	pyrophosphoethanolamine

PS	polysaccharide
psi	pound-force per square inch, 6895 Pa
PVDF	polyvinylidene fluoride
QM-MM	quantum mechanics-molecular mechanics
RAM	L-rhamnose
Rha	L-rhamnose
R-LPS	rough lipopolysaccharide
RMSD	root-mean-square deviation
RMSF	root-mean-square fluctuation
RNA	ribonucleic acid
rpm	revolutions per minute, 1/60 Hz
RT	room temperature
RU	repeating unit (for LPS);
RU	response unit (for SPR)
RU_{max}	maximal response unit (for SPR)
rv	reverse
S.	*Shigella*
SCWP	secondary cell wall polymer
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
set1B	enterotoxin 1 gene
Sf2a	*Shigella flexneri* 2a O-antigen
Sf6	*Podovirus* Sf6
Sf6 TSP	tailspike protein of Sf6
SfY	*Shigella flexneri* Y O-antigen
S-LPS	smooth lipopolysaccharide
SNFG	symbolic nomenclature for glycans
Spa	component of the Type III secretion machinery in *Shigella*
spp.	*species pluralis*
SPR	surface plasmon resonance
SR-LPS	smooth rough lipopolysaccharide
ssDNA	single-stranded DNA
ssRNA	single-stranded RNA
Tab.	Table
TAE	tris-acetate-EDTA
TCEP	tris(2-carboxyethyl)phosphine
TEC	tris-EDTA-sodium chloride
TEMED	tetramethylethylenediamine
TEV	Tobacco etch virus
TFA	trifluoroacetic acid
TI	thermodynamic integration

TIP3P, 4P, 5P	water models in molecular dynamics simulations
Tris	tris(hydroxymethyl)aminomethane
TSP	tailspike protein
U	enzyme activity unit, 1 µmol of substrate per min
UDP	uridine diphosphate
UV-Vis	ultraviolet–visible light
vdW	van der Waals
VMD	Visual Molecular Dynamics
(v/v)	volume volume percent
WaaP	Lipopolysaccharide core biosynthesis protein, Heptosyl-1-kinase
wt	wild type
(w/v)	weight volume percent
Wzx	polysaccharide polymerase
Wzz	polysaccharide chain length determinant protein

Amino acids are given as three and one letter code. Abbreviations are unified for singular and plural. Protein variant names are described within the text if not listed here.

Acknowledgement

Willkommen auf der letzten Seite meiner Doktorarbeit oder der ersten, je nach Präferenz.
Die Person, der ich am meisten danken möchte und ohne die diese Doktorarbeit nicht möglich gewesen wäre, ist Stefanie Barbirz, die nicht nur eine gute Chefin und soziale Betreuerin war, sondern auch eine lustige Büronachbarin und großartige Kollegin ist. Danke, Steffi, für all die Begeisterung an diesem Projekt und für alle machmal überwältigenden Ideen, für die Zeit, die du dir in Besprechungen, im Labor, beim Korrekturlesen und bei Vorbereitungen von Präsentationen genommen hast. Danke, dass du mich auf Konferenzen geschickt hast und mich immer wie eine gleichgestellte Wissenschaftlerin nur mit weniger Erfahrung behandelt hast.
Ich danke Prof. Robert Seckler, dass er mir die Möglichkeit gegeben hat, diese Arbeit durchzuführen, dass er allen meinen Bitten entgegen gekommen ist und dass er sich dazu bereit erklärt hat, diese Arbeit zu bewerten. Furthermore I would like to thank Prof. Heiko Möller and Prof. Anne Imberty for accepting to review this thesis.
Danke an Lea Günther, Alexandra Helm und Saskia Buchwald, die ihre Bachelorarbeiten oder Praktika im Rahmen dieser Arbeit durchgeführt haben, die mich als Betreuerin akzeptieren mussten und dabei großartige Ergebnisse produziert haben. Danke an Melanie Anding für ihre Hilfe bei zu langwierigen Optimierungen und für ihren Spaß an meinen Arbeitsplänen. Danke an Jana, Mandy, Simone und Anja als gute Geister des Labors.
Außerdem möchte ich Mark Santer für die Zweitbetreuung dieser Doktorarbeit danken, dass er mir in der Welt der Simulation immer versucht hat weiter zu helfen, mich mit Skripten unterstützt hat und dass er meine Intuition als akzeptable Größe und als Potential zur treibenden Kraft von Wissenschaft akzeptiert hat.
I'd like to thank my collegues from the MPI, especially Yu Kang and Marco Wehle, who introduced me to the world of simulations, and the helpers from the IT and Susi. Special thanks to Fabian Müller-Knapp for providing me with computer scripts as a friend. Thanks to the vesicle and the nucleation groups for surviving with me in workshops in the world of the theoreticians and to my IMPRS fellows, who became my peers, flat mates, conference traveling mates and sport buddies. Special thanks to my friends who were here during this time: Tom, Kristin, Rafa, Ella, Kasia, José, R, ßaasha, Elliot, Selin, Cha, Afro, Arne, Jan the J., Fiona und Meike and everybody I might forget at 3 am ☺ und danke an meine Kommilitonen Fabienne, Sven und Wiebke, die das Leid immernoch aus der Ferne teilen.
Ein ganz besonderer Dank geht aber auch an meine Kollegen, die meine Launen für ungemeine Albernheiten aber auch Labortürknallausraster ertragen haben und die mich besonders in den letzten Tagen unendlich unterstützt haben. Danke an Anja, Nina, Patrick, Ella, Vasil und Tobi fürs Korrektur lesen. Danke an Nina, Tobi, Andreas, Igor, Kerstin, Marlene und Mareike, dass sie die Freuden und Leiden der Tailspikes teilen oder geteilt haben, an Martin für die gute Kultur des Streitens, an Klaus, Sweta und Karolin für die Mensagespräche und an Anja und Patrick für den Kaffee, den Schabernack, das Probleme lösen und für die Erlebnisse im RE1.
Васил, благодария за годината с ти и да съм тукка винаги. Мерси да съм един приятел.
Ich danke meiner Großmutter für ihren und meinen Ehrgeiz, meinen Brüdern für die leichten Schläge auf den Hinterkopf und das Lebärnstraining und meinen Eltern für die Unterstützung und dass sie der Neugierde immer freien Lauf gelassen haben.