UNIVERSITÀ DEGLI STUDI DI MILANO
FACOLTÀ DI SCIENZE E TECNOLOGIE

PhD Program in Chemistry - XXXVIII Cycle

Chemistry department

CHARACTERIZATION OF THE INTERACTION BETWEEN HEPARAN SULFATE AND THE SPIKE (S) PROTEIN OF THE LATEST SARS-CoV-2 VARIANTS

Tutor: Prof. Monica Civera

Co-tutor: Dr. Stefano Elli

PhD coordinator: Prof. Daniele Passarella

PhD candidate:

Mandalari Marco

Matr. R13728

- Academic Year 2024/2025 -

PhD Program in Chemistry - XXXVIII Cycle

Chemistry department

CHARACTERIZATION OF THE INTERACTION BETWEEN THE HEPARAN SULFATE AND THE SPIKE PROTEIN OF THE LATEST VARIANT OF SARS-CoV-2

Thesis was written with the financial support of the **European Union - Next Generation EU and the Istituto di Ricerche Chimiche e Biochimiche 'G. Ronzoni'** - CUP no. G43C22002280004

PhD candidate: Mandalari Marco

Matr. R13728

ORCID: 0000-0002-8968-4896

Tutor: Prof. Monica Civera

Co-tutor: Dr. Stefano Elli

PhD coordinator: Prof. Daniele Passarella

- Academic Year 2024/2025 -

"I have no special talents. I am only passionately curious."

Albert Einstein

Table of Contents

Table of Contents **i**

Summary **v**

List of Abbreviations **vii**

Chapter 1. Glycosaminoglycans **1**

1.1 Introduction **1**

1.2 Chemical structure of heparin and heparan sulfate **3**

1.3 Biosynthesis of heparin and heparan sulfate: encoding the microheterogeneity **5**

1.4 Heparan sulfate proteoglycans: structure, localization, and functional implications **9**

1.5 Heparin: Multifaceted Biological and Clinical Properties Beyond Anticoagulation **11**

Chapter 2. The role of glycosaminoglycans in SARS-CoV-2 infection **15**

2.1 The outbreak of the SARS-CoV-2 pandemic **15**

2.2 The SARS-CoV-2 spike protein structure **16**

2.3 The SARS-CoV-2 variants of concern (VOCs) **19**

2.3.1 Structural comparison of the RBD domain in VOCs 24

2.4 The SARS-CoV-2 glycosylation **27**

2.4.1 Glycosylation as a key contributor to protein structural and functional diversity 27

2.4.2 Glycosylation landscape of the SARS-CoV-2 spike protein 29

2.4.3 The functional role of the N343 glycan in SARS-CoV-2 30

2.5 HS acts as a co-receptor in the SARS-CoV-2 cell infection. 31

2.6 Aim of the Project 34

Chapter 3. Methodologies 36

3.1 Experimental methods 36

3.1.1 Saturation Transfer Difference (STD) NMR Spectroscopy 36

3.2 Computational methods 40

3.2.1 Docking Calculation 40

3.2.2 Molecular Dynamics (MD) Simulations 44

3.2.3 Force field 45

3.2.4 Solvation box 48

3.2.5 Periodic Boundary Condition (PBC) and Particle Mesh Ewald (PME) 49

3.2.6 Thermostat and barostat 50

3.2.7 Simulation of STD_0 signals 52

Chapter 4. Results 55

4.1 Unveiling the conformational landscapes of oligosaccharide mimetics of heparan sulfate in unbound state. 55

4.1.1 Monosaccharide ring conformation 57

4.1.2 Inter-glycosidic bond 59

4.2 Step 1: Docking calculations 62

4.3 Step 2: ^{1}H-STD NMR interaction experiment 70

4.3.1 Mapping the glycan binding epitope of the interaction between hexa and Omicron RBD by ^{1}H-STD NMR interaction experiment. 70

4.3.2 Influence of ligand nature on NMR spectral characteristics 70

4.3.3 The binding epitope of hexa in complex with Omicron RBD 73

4.4 Step 3: Molecular dynamics simulations 77

4.4.1 Hexa explores the surface of Omicron and wild-type RBDs. 77

4.4.2 Analysis of the interaction between hexa and the wild-type and Omicron RBDs. 81

4.4.3 The conformational properties of the N343 glycan affect the interaction between hexa and RBDs. 84

4.5 Step 4: NMR-driven analysis of MD runs 92

4.5.1 Selection of WT and Omicron complexes using the NMR and MD data 92

4.6 Conclusions 99

Chapter 5. Materials and methods 103

5.1 Docking calculations 103

5.1.1 Protein preparation 103

5.1.2 Ligand Preparation 104

5.1.3 Docking setup 105

5.1.4 Estimation of the free energy of binding 105

5.1.5 Ligand-protein interactions 105

5.2 All-atom MD simulations 106

5.2.1 System preparation 106

5.2.2 Analysis of the MD simulations 108

5.2.3 Distance-base filtering of MD meta-trajectory. 109

5.2.4 3D model validation by reduced matrix (RedMat) 109

5.2.5 Cluster analysis 109

5.3 ^{1}H-STD NMR experiments 110

Appendix 112

Bibliography 132

Summary

Heparan sulfate is a linear, highly sulfated polysaccharide that belongs to the glycosaminoglycan (GAG) family. The disaccharide repeating unit includes a uronic acid residue (either D-glucuronic or L-iduronic) 1→4 linked to a glucosamine (GlcN). The glucosamine residues can be mono- or polysulfated at the N-, 6-O, and rarely at the 3-O positions, while the uronic acids can be sulfated at the 2-O position. Furthermore, HS displays heterogeneity in chain length and block composition, in which acetyl- and sulfate-rich sequences alternate. HS is covalently attached to the membrane protein (via the O- or the N-glycosylation of a Ser or Asn residue, respectively), forming heparan sulfate proteoglycans (HSPGs). Due to the pronounced density of negative charge of HS, HSPGs play a key role in several biological processes, including cell hydration, intercellular communication, storage, protection, and exchange of biomolecules (e.g., cytokines, chemokines), and the regulation of receptors and proteolytic enzymes (e.g., tyrosine kinase-type growth factor). Moreover, HS is also involved in diverse physiological and pathological events, such as embryonic development, inflammatory response, blood coagulation, and bacterial and viral infection. In the latter context, several studies show that a wide range of pathogens can interact with HS during the early stage of the infection. This aspect has recently drawn significant attention following the outbreak of the SARS-CoV-2 pandemic. Indeed, the virus exploits HS in the extracellular matrix to tether viral particles to the host cell surface, thereby promoting the specific interaction between the cell surface human angiotensin-converting enzyme 2 receptor (hACE2) and the receptor binding domain (RBD) that decorates the capsid of SARS-CoV-2. In fact, this event triggers the fusion between the cell and virus membranes. The RBD region

(and, more broadly, the S1 subunit) contains several positively charged amino acids that are solvent-exposed and form a positive channel in which HS chains can be hosted without overlapping the hACE2 binding site.

The primary goal of this study is to understand how the role of HS in these early molecular recognition events differs in the latest and most widespread variants of SARS-CoV-2, through the application of complementary computational and experimental approaches that combine NMR spectroscopy and different computational tools.

We have selected two hexasaccharides as HS mimetics to probe their interactions with different RBDs using docking calculations. We then focused our study on the Omicron variant and one of the hexasaccharide mimetics, performing molecular dynamic simulations and ^{1}H-NMR Saturation Transfer Difference experiments. Structural insights into its binding modes were obtained using RedMat, an NMR-based analysis tool for the analysis of molecular dynamic simulation trajectories, and they were compared with the ligand interactions observed for the wild-type RBD. This knowledge will contribute to the structural biology knowledge of SARS-CoV-2 and to the development of new antiviral strategies and antiviral drugs.

List of Abbreviations

ARDS	Acute Respiratory Distress Syndrome
AT	Antithrombin
AMBER	Assisted Model Building and Energy Refinement
CFR	Case fatality rate
CoM	Centre of mass
CHARMM	Chemistry at Harvard Macromolecular Mechanics
CS	Chondroitin sulfate
CPPTRAJ	Complementary Portable Program for Trajectory
COVID-19	Coronavirus Acute Respiratory Disease 2019
Cryo-EM	Cryo-Electron Microscopy
CTD1	C-terminal domain 1
CTD2	C-terminal domain 2
CT	Cytoplasmic tail
DS	Dermatan sulfate
GLCE	D-Glucuronyl C5-Epimerase
DIC	Disseminated intravascular coagulation
ER	Endoplasmic reticulum
EXT1-2	Exostosin 1-2
ECM	Extracellular matrix
FF	Force field
Fuc	Fucose
FP	Fusion peptide
Gal	Galactose
4Gal-T	Galactosyltransferase
3Gal-T6	Galactosyltransferase 6

NDSTs	GlcNAc-N-deacetylase/N-sulfotransferases
GlcN	Glucosamine
GlcNS,3S,6S	Glucosamine N-, 3-O-, 6-O-trisulfated
GlcNS,6S	Glucosamine N-, 6-O-disulfated
GNPNAT	Glucosamine-phosphate N-acetyltransferase
GPI	Glucose-6-phosphate isomerase
GlcA	Glucuronic acid
GlcAT-I	Glucuronyltransferase
GFAT	Glutamine-fructose-6-phosphate transaminase
GAGs	Glycosaminoglycans
GPU	Graphic processing unit
GUI	Graphical unit interface
GFP	Green fluorescent protein
GLIDE	Grid-based Ligand Docking with Energetics
HS	Heparan sulfate
HSPGs	Heparan sulfate proteoglycans
HP	Heparin
HBD	Heparin binding domain
HSase	Heparin lyase
HBPs	Heparin-binding proteins
HR1	Heptapeptide repeat sequence 1
HR2	Heptapeptide repeat sequence 2
HTVS	High-Throughput Virtual Screening
hACE2	Human angiotensin converting enzyme 2
HA	Hyaluronic acid
HB	Hydrogen bond
IdoA	Iduronic acid

IdoA2S	Iduronic acid 2-O-sulfated
PPi	Inorganic pyrophosphate
IFNγ	Interferon-gamma
IL-6	Interleukin-6
KS	Keratan sulfate
LMWH	Low-molecular-weight heparin
Man	Mannose
NPT	Isothermal-isobaric ensemble with constant number of particles, pressure, and temperature T
NVE	Microcanonical ensemble with constant number of particles N, volume V, and energy E
NVT	Canonical ensemble with constant number of particles N, volume V, and temperature T
MERS-CoV	Middle East Respiratory Syndrome Coronavirus
MD	Molecular dynamics
MM-GBSA	Molecular mechanics - Generalized Born surface area
GalNAc	N-acetyl-galactosamine
GlcNAc	N-acetyl-glucosamine
NTD	N-terminal domain
NF-κB	Nuclear factor kappa B
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
NST	Nucleotide-sugar transporter
OPLS	Optimized Potentials for Liquid Simulations
XYLT1	O-xylosyltransferase 1
XYLT2	O-xylosyltransferase 2
PME	Particle Mesh Ewald
PPS	Pentosan Polysulfate sodium

PBC	Periodic boundary condition
PGM2	Phosphoglucomutase
PGM3	Phosphoglucomutase
PDB	Protein Data Bank
RBD	Receptor binding domain
RBM	Receptor binding motif
RedMat	Reduced matrix
RMSD	Root means square deviation
RMSF	Root means square fluctuation
STD	Saturation Transfer Difference
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SLC35	Solute carrier family 35
SP	Standard precision
S1	Subunit 1
S2	Subunit 2
TM	Transmembrane domain
TMPRSS2	Transmembrane protease serine 2
GALE	UDP-glucose 4-epimerase
UGDH	UDP-glucose 6-dehydrogenase
UXS1	UDP-glucuronate decarboxylase
AGX1	UDP-N-acetylglucosamine diphosphorylase
UFH	Unfractionated heparin
UDP	Uridine diphosphate
UGP2	UTP-glucose-1-phosphate uridylyltransferase
GALT	UTP-hexose-1-phosphate uridylyltransferase
VOCs	Variants of Concern
VOIs	Variants of Interest

VUMs	Variants under monitoring
XP	Extra-Precision

Chapter 1. Glycosaminoglycans

1.1 Introduction

Glycosaminoglycans (GAGs) are members of a class of negatively charged linear polysaccharides. They consist of repeating disaccharide units composed of either N-acetylated or N-sulfated hexosamine, alternately linked to hexuronic acids (D-glucuronic or L-iduronic acid) or galactose [1]. Based on the sulfation pattern and the type of disaccharide core unit, five different groups can be distinguished: hyaluronic acid (HA), heparin/heparan sulfate (HP/HS), chondroitin sulfate (CS), dermatan sulfate (DS) and keratan sulfate (KS) (**Figure 1**).

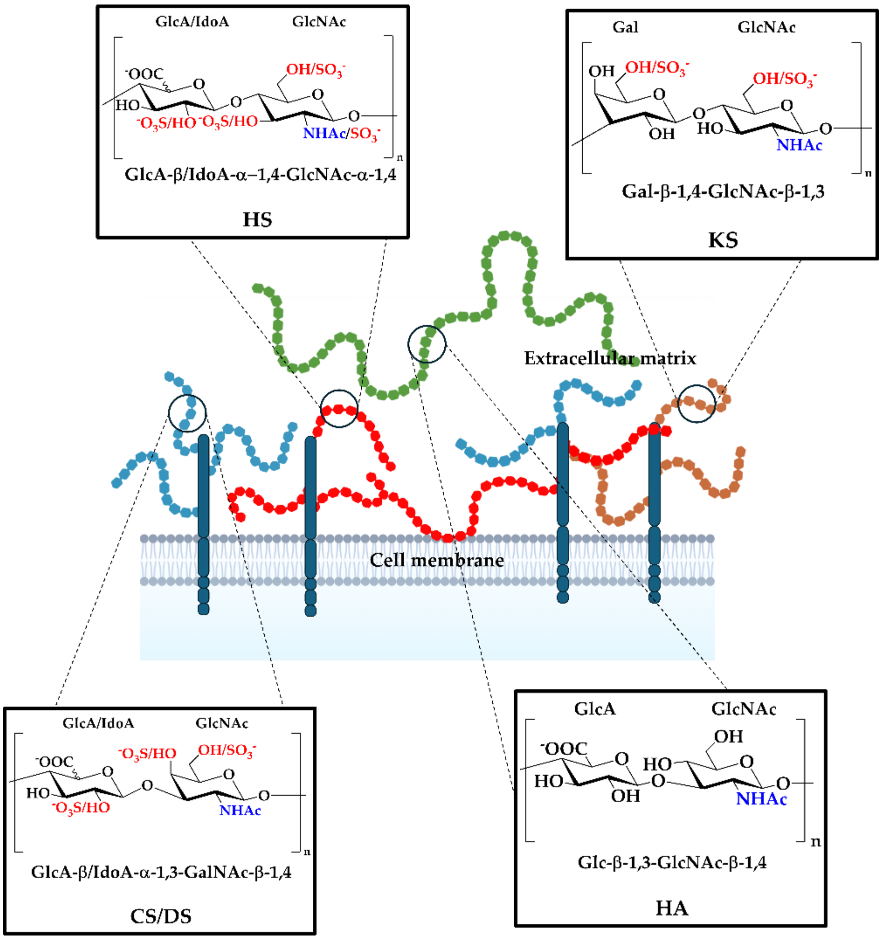
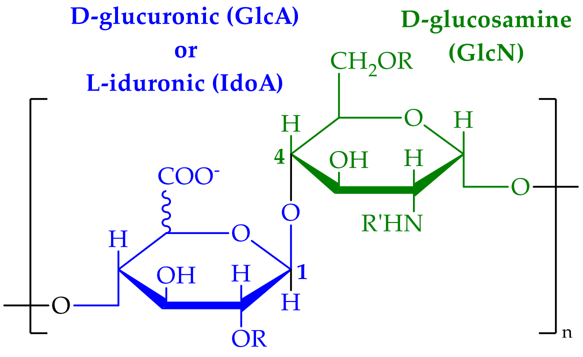


Figure 1 – Schematic representation of principal GAG families and their repeating disaccharide units. CS/DS, HS, HP, and KS are linked to core protein, whereas HA is dispersed in ECM. The potential acetylation and sulfation sites are colored in blue and red, respectively. *Image adapted from Wang, Q., et al. Polymers 2022, Vol. 14, Page 5014 14(22).* https://doi.org/10.3390/polym14225014

Except for HA, they are covalently attached to the membrane or secreted protein, forming proteoglycans [2, 3]. GAGs are abundantly expressed in the extracellular matrix (ECM), on cell surfaces, and in connective tissues of vertebrates, where they play a key role in tissue hydration, elasticity, and mechanical strength. They are also involved in organizing the ECM [4], cell communication (with other cells, external entities or stimuli) and modulating ligand–receptor recognition [5]. Lastly, GAGs actively participate in inflammatory processes by regulating the local availability and distribution of chemokines, cytokines, and growth factors. Due to their ubiquitous nature and the high degree of structural diversity (attributable to numerous chemical modifications), GAGs participate in multiple physiological processes, and altered activity (downregulation or over-expression) can cause the development and progression of various pathologies [1, 5, 6]. Although GAGs have been traditionally associated with the mammals, several lines of evidence indicate that they are also present in invertebrates, ranging from primitive organisms such as nematodes and cnidarians to more complex species, including mollusks, echinoderms and arthropods [7].

1.2 Chemical structure of heparin and heparan sulfate

This thesis primarily focuses on heparin and heparan sulfate (HS) class, which share a common disaccharide building block, consisting of uronic acid D-β-glucuronic (GlcA) or L-α-iduronic (IdoA) residue linked to α-D-glucosamine (GlcN) residue through a 1→4 glycosidic bond (Figure 2).



R=H or SO_3^- and R' = H, Ac, SO_3^-

Figure 2 – Depiction of the common disaccharide building block shared by heparan sulfate and heparin.

Heparin has a long history of use in clinical settings as major anticoagulant drug. Typically, its glucosamine residues are sulfated at the N- and 6-O positions, while the hexuronic acid undergoes sulfation at the 2-O position, forming the characteristic IdoA2S-GlcNS,6S disaccharide unit [8].

In the latter, the presence of the rare 3-O sulfation in the GlcN (GlcNS,3S,6S) has been occasionally observed. In porcine heparin, however, this moiety is present at a concentration of about 3.8 - 5% (molar ratio) [9] and is indispensable for assembling the unique pentasaccharide motif GlcNS,6S–IdoA2S–GlcNS,3S,6S–IdoA2S–GlcNS,6S, which confers to HP its characteristic high-affinity binding to antithrombin (AT). Upon binding, AT undergoes a conformational activation that accelerates its inhibitory activity against thrombin and factor Xa, increasing the rate of inhibition by several orders of magnitude [10].

The structure of HS is similar to heparin, despite the presence of lower degrees of sulfation and epimerization [11]. Whereas HP contains three sulfate groups per disaccharide unit, while HS typically has only about one sulfate per disaccharide unit. This discrepancy in the degree of sulfation results in a different distribution pattern for N-sulfation on GlcN residues: HS is composed of consecutive N-acetylated disaccharide units (NA domains) or alternating N-acetylated and N-sulfated units (NA/NS domains) interspersed with N-sulfated sequences of varying lengths (NS domains), while HP chains are prevalently N-sulfated [12]. Additionally, further distinctions lie in the chain length and uronic acid composition: heparin has a higher percentage of L-iduronic acid, while D-glucuronic acid predominates in the HS chains [13-15] (**Table A. 1**). The intricate nature of heparin and heparan sulfate, arising from their extensive structural variability, is further enhanced by their conformational flexibility, which is due to the presence of iduronic acid residues within the polysaccharide chain. In solution, glucosamine (GlcN), and glucuronic acid (GlcA) residues adopt the chair conformation $^4C_1$ predominantly. In contrast, the iduronic acid (IdoA) residue exists in a dynamic equilibrium between the chairs $^4C_1$, $^1C_4$, and the skew-boat $^2S_0$ conformations [16-19] (Figure 3). The ratio between the conformer populations is influenced by the degree of sulfation of both the iduronic acid itself and the neighboring glucosamine units [20].

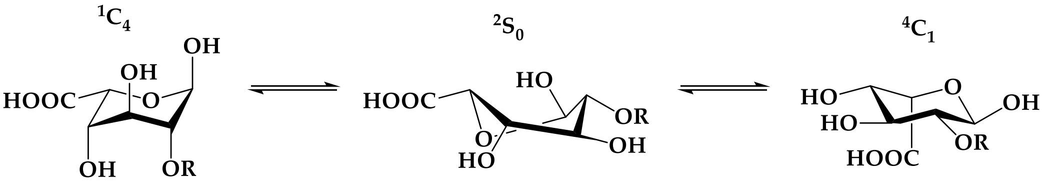


Figure 3 – Conformers of iduronic acid in the heparin/heparan sulfate chains. It exists in an equilibrium between the conformations $^4C_1$, $^2S_0$, and $^1C_4$, according to the degree of sulfation of the surrounding glucosamines.

1.3 Biosynthesis of heparin and heparan sulfate: encoding the microheterogeneity

The structural heterogeneity of HP/HS arises from their specific biosynthesis (Figure **4** and **Figure 5**). Unlike proteins or nucleic acids, this process is not template-driven, meaning that the final structural motifs are not directly encoded in the genome. Instead, a series of enzymatic reactions progressively fine-tune the overall HS/HP structure, resulting in the structural diversity observed.

The process begins in the cytosol, where starting from glucose 6-phosphate and glucosamine 6-phosphate, the five monosaccharides required for GAGs biosynthesis (Xylose [**Xyl**], Galactose [**Gal**], Glucuronic acid [**GlcA**], N-acetyl-glucosamine [**GlcNAc],** and N-acetyl-galactosamine [**GalNAc**]) are synthesized as uridine diphosphate (UDP) derivatives (**Figure 4**), since free monosaccharides are energetically unfavorable in subsequent glycosylation reactions. The first phosphorylation to sugar-1-monophosphate is catalyzed by sugar kinases. Then a UDP-sugar pyrophosphorylase enzyme transfers a uridine diphosphate (UDP) moiety from a uridine-5′-triphosphate to the sugar-1-phosphate, forming the UDP-sugar and inorganic pyrophosphate (PPi). The activated UDP-sugars are then transferred to the Golgi apparatus by specific nucleotide-sugar transporters (NSTs), belonging to the SLC35 protein family. Monosaccharide activation and internalization are essential processes, since most of the enzymes that participate and regulate the biosynthesis are located in the Golgi apparatus (**Figure 4**).

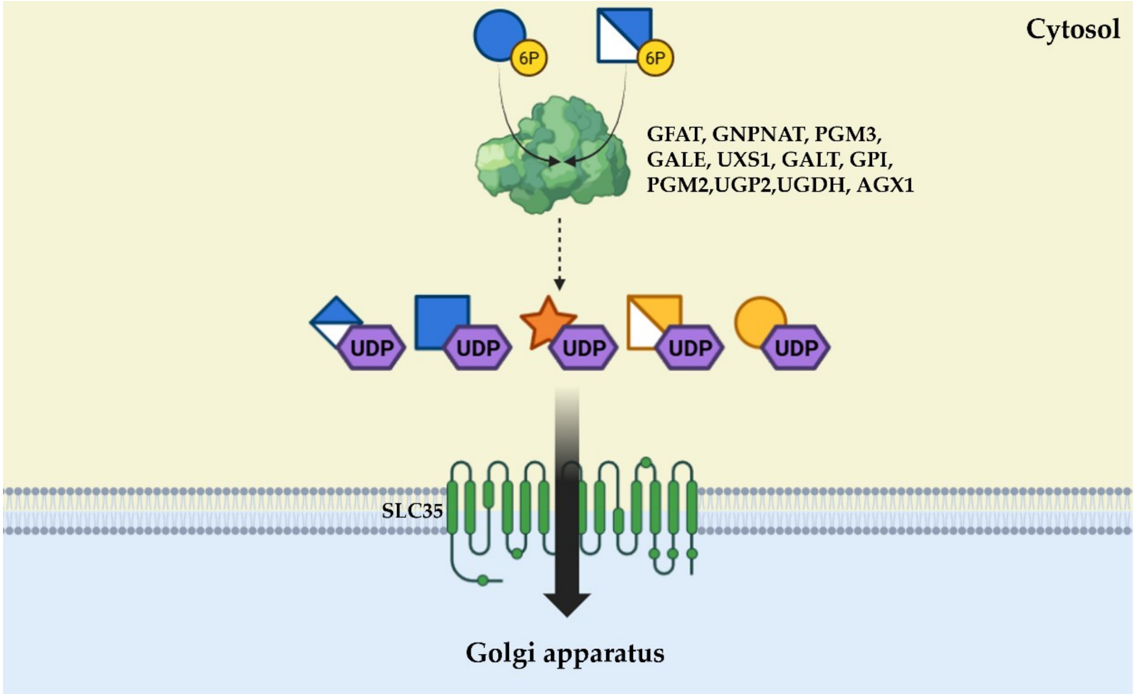


Figure 4 – First step in the biosynthesis of heparan sulfate proteoglycans (HSPGs) and HP chains. Starting from glucose-6-O-phosphate and glucosamine-6-O-phosphate, the corresponding building blocks (glucuronic acid, N-acetyl-glucosamine, xylose, N-acetyl-galactosamine, and galactosamine) are synthesized and transported inside the Golgi apparatus through specific nucleotide-sugar transporters (NSTs). GFAT = glutamine-fructose-6-phosphate transaminase, GNPNAT = glucosamine-phosphate N-acetyltransferase, PGM3 = phosphoglucomutase, GALE = UDP-glucose 4-epimerase, UXS1= UDP-glucuronate decarboxylase, GALT = UTP-hexose-1-phosphate uridylyltransferase, GPI = glucose-6-phosphate isomerase, PGM2 = phosphoglucomutase, UGP2 = UTP-glucose-1-phosphate uridylyltransferase, UGDH = UDP-glucose 6-dehydrogenase, AGX1 = UDP-N-acetylglucosamine diphosphorylase.

Once these building blocks enter the organelles, GAGs synthesis can begin. The first step is the synthesis of an anchor protein sequence, GlcA-β(1→3)-Gal-β(1→3)-Gal-β(1→4)-Xyl-1-O-Ser, a conserved tetrasaccharide attached to the core protein of all proteoglycans that bear GAGs (**Figure 5**). This process is catalyzed by four enzymes that add individual sugar residues sequentially to the non-reducing end of the growing chain. The first reaction catalyzed in this process is the xylosylation of a serine residue, performed by the two O-xylosyltransferases XYLT1 and XYLT2. Subsequently, two galactose residues and one glucuronic acid residue are added by galactosyltransferases β4Gal-T7, β3Gal-T6, and glucuronyltransferase GlcAT-I, respectively (**Figure 5**). This process is essential for the subsequent polysaccharide polymerization, as the hetero-oligomeric complex enzyme exostosin 1-2 (EXT1-2) recognizes this primer sequence as the starting point to chain elongation and begins to alternatively add glucuronic acid (GlcA) and N-acetylglucosamine (GlcNAc) residues. The next phase is the GAGs maturation, which involves the modification of the newly formed GlcA-GlcNAc polysaccharide. During this phase, the GlcA-GlcNAc homocopolymer undergoes N-deacetyltion/N-sulfation, catalyzed by the bifunctional enzymes GlcNAc-N-deacetylase/N-sulfotransferases (NDSTs) (**Figure 5**). Among the four different NDST isoforms, NDST1 is principally involved in the HS synthesis, while NDST2 performs the N-deacetylation/N-sulfation in the heparin chains. This process, regulated by the cellular level of PAPS (3′-phosphoadenosine 5′-phosphosulfate), is a key process in GAGs maturation, since all subsequent enzymes require the presence of GlcNS to promote their respective reaction. Indeed, the $[4\text{GlcA-}\beta(1\rightarrow4)\text{GlcNS}\alpha\text{-}1]_n$ sequences are recognized by D-glucuronyl C5-epimerase (GLCE), which catalyzes the epimerization of D-GlcA into L-IdoA. Since GLCE can also perform the reverse process (from IdoA to GlcA), the configuration of the newly formed IdoA residue is further stabilized through a 2-O-sulfation (IdoA2S), catalyzed by 2-O-sulfotransferase (OST1), as GLCE cannot employ IdoA2S as substrate. In the later stage of the biosynthesis, GlcNS undergoes 6-O-and 3-O-sulfation, processes that are catalyzed by 6-O-sulfotransferases 1-3 (6OSTs 1-3) and 3-O-sulfotransferases 1-7 (3-OSTs 1-7) isoforms (**Figure 5**). The latter modification is the crucial step that confers heparin its characteristic antithrombotic activity. Extensive modification of the nascent GlcA–GlcNAc copolymer through N-deacetylation/N-sulfation, C5-epimerization, and subsequent O-sulfations generates the highly sulfated heparin polysaccharide, whereas partial modification through these processes results in the formation of heparan sulfate. Once the synthesis of the polysaccharide chains is completed, the GAG-decorated proteoglycans are expressed on the cell membrane and involved in additional postsynthetic modifications, namely 6-O-desulfation, which is catalyzed by 6-O-endosulfatases Sulf-1 and Sulf-2 [21] (**Figure 5**).

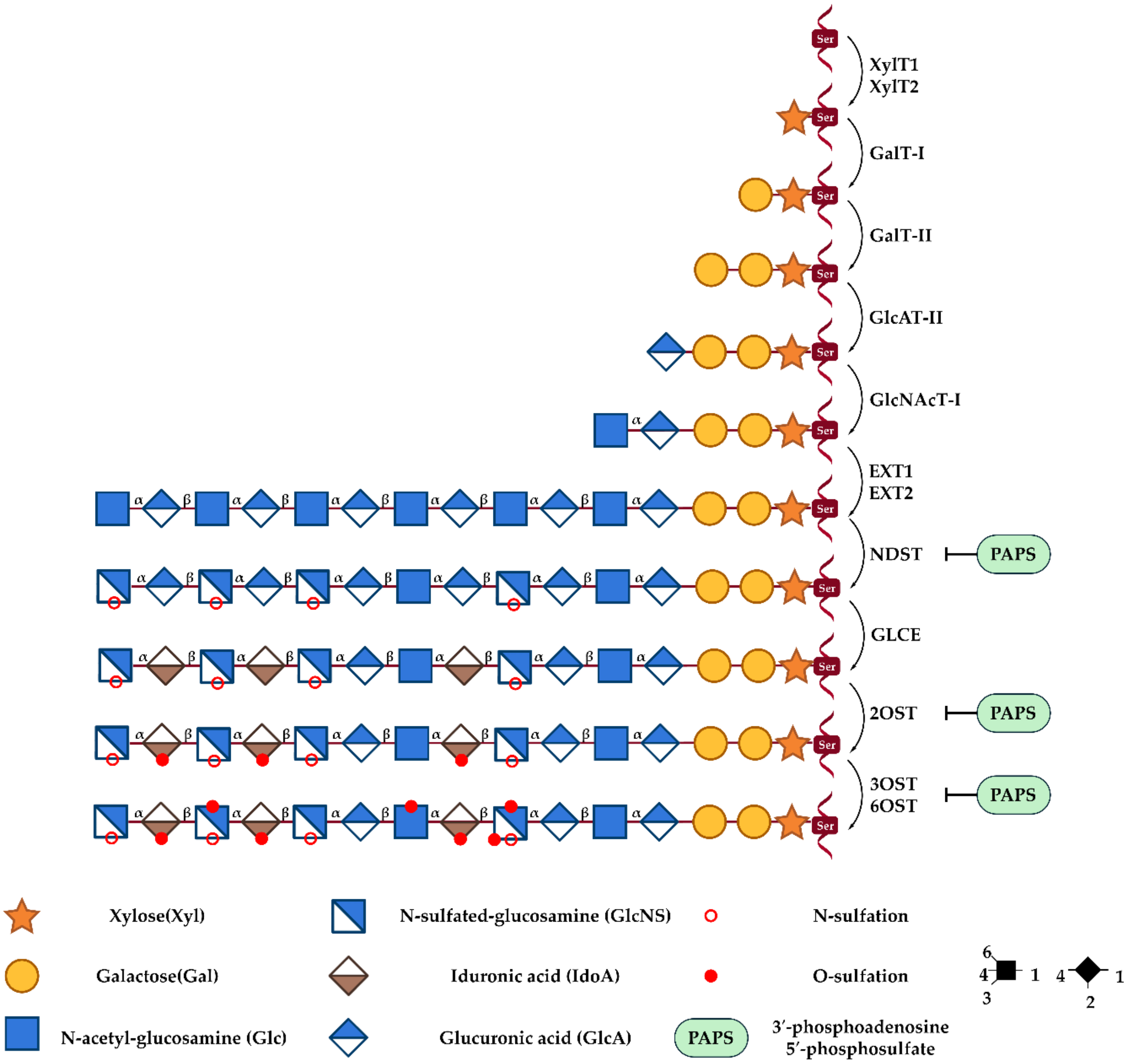


Figure 5 – The elongating polysaccharide chain is covalently linked to a serine residue of the core protein. Following synthesis of the tetrasaccharide linkage region (Xyl–Gal–Gal–GlcA) and the addition of the initial N-acetylglucosamine (GlcNAc), the glycosyltransferase complex composed of EXT1 and EXT2 catalyzes the alternating incorporation of glucuronic acid (GlcA) and GlcNAc residues from their corresponding uridine diphosphate (UDP) precursors. After chain polymerization, the polysaccharide undergoes extensive modification in the presence of the universal sulfate donor 3-phosphoadenosine-5-phosphosulfate (PAPS). These modifications include N-deacetylation and N-sulfation of GlcNAc residues, C5-epimerization of GlcA to iduronic acid (IdoA), 2-O-sulfation of uronic acids, as well as 6-O-sulfation and 3-O-sulfation of glucosamine units.

1.4 Heparan sulfate proteoglycans: structure, localization, and functional implications

In the cellular environment, HS forms covalent complexes with specific "core" proteins. These macromolecules are ubiquitously expressed on the cellular surface and in the ECM in both invertebrates and vertebrate species [22, 23] interacting with surrounding other intra- and extracellular proteins, ions, and biomolecules. Based on the core protein and their location, three HSPG subfamilies are distinguished (Figure 6):

- **Syndecans**, which are embedded in the cell surface via their transmembrane domain and equipped with a short intracellular tail.
- **Glypicans**, whose core protein is anchored to the cell membrane through a glycosylphosphatidylinositol (GPI) anchor.
- **Perlecans**, which are directly secreted into the ECM.

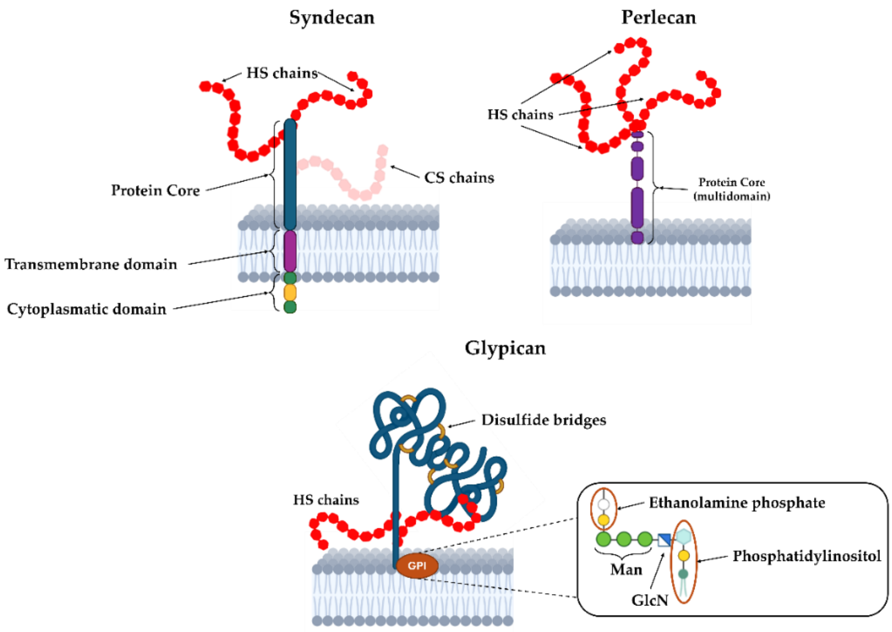


Figure 6 – Schematic representation of the three main subfamilies of heparan sulfate proteoglycans (HSPGs). *Image adapted from Matsuzaka, Y., Yashiro. Biologics 2024, 4, 105-129. https://doi.org/10.3390/biologics4020008*

Syndecans are characterized by a variable ectodomain that carries HS chains, a short hydrophobic transmembrane region, and a cytoplasmic tail. A protease-sensitive site located in proximity of the ectodomain facilitates the process of proteolytic shedding, whereby the extracellular portion is detached from the cell surface. Furthermore, the membrane-bound syndecans have been observed to undergo internalization via the process of endocytosis, which is then targeted for lysosomal degradation. This observation suggests the presence of motifs within the transmembrane or cytoplasmic domains of syndecans that can interact with endocytic machinery. The cytoplasmic domain also harbors peptide sequences that are associated with cytoskeletal proteins and function as substrates for cellular kinases, thereby enabling syndecans to act as signaling molecules. In contrast, members of the glypican family possess an extracellular domain containing GAG attachment sites and 14 conserved cysteine residues that stabilize a compact tertiary structure (Figure 6). In contrast to transmembrane segments, glypicans are anchored to the outer surface of the plasma membrane via a C-terminal glycosylphosphatidylinositol (GPI) linkage. Perlecans are large proteoglycans found in basement membranes and the ECM. They have a protein core of about 500 kDa, which contains several conserved domains. The N-terminal portion of this core is highly glycosylated, carrying long heparan sulfate chains (≈65 kDa each). HSPGs play a key role in a wide range of biological and physiological processes, including the storage, trafficking, and activation of biomolecules (including growth factors, chemokines [24], microbial adhesion factors [25], morphogens, cytokines, enzymes, and ECM proteins). HSPGs are also involved in cell growth, differentiation, adhesion, migration, and apoptosis [26]. Hence, alteration in HSPGs activity or regulation can have a significant impact on normal cellular function, leading to pathological conditions such as inflammation and cancer. Moreover, the long polysaccharide chains of HSPGs can be recognized and exploited by viruses and bacteria during their initial stage of infection [27-32]. This "broad spectrum" of interaction arises from the unique structure of HS: the presence of discontinuous NS regions along the polysaccharide chains allows these macromolecules to bind a wide range of compatible proteins, commonly referred to as HS- or heparin-binding protein (HBP). These proteins are typically rich in basic amino acids (arginine and lysine), which interact electrostatically with the densely negative sulfate regions present on the HS chains. These basic motifs are called heparin-binding domains (HBDs): Cardin and Weintraub [33], for the first time in 1989, identified a correlation between the HD and conserved sequence of amino acids, such as [-X-B-B-X-B-X-] and [-X-B-B-B-X-X-B-X-], where X and B correspond to a hydropathic and basic residue, respectively.

1.5 Heparin: Multifaceted Biological and Clinical Properties Beyond Anticoagulation

As with HS, heparin is also synthesized as part of a proteoglycan, whose core protein is serglycin [34]: a small proteoglycan that is rich in glycine and serine residues. However, while HS is produced virtually in all cell types across the body, biosynthesis of endogenous heparin is restricted to mast cells, where it forms specialized proteoglycans that are involved in the storage, retention, and activation of histamine, inflammatory mediators, and proteases [34]. In contrast, exogenous heparin, which is extracted from porcine and bovine mucosa, is used as an anticoagulant drug to prevent thromboembolic disorder events. However, approximately 70% of extracted heparin lacks the specific pentasaccharide sequence required for anticoagulant activity [35]. This has prompted several research groups to investigate the alternative biological role of heparin beyond its well-established anticoagulant properties [35, 36]. Indeed, the Swiss-Prot database currently lists around 250 entries referring to proteins modulated by heparin, highlighting its diverse non-anticoagulant properties [35].

For instance, heparin has been investigated as an anticancer agent: this property emerged from several clinical findings, observing longer survival of cancer patients in whom heparin was administered for the treatment of cancer-associated thromboembolic disease [37, 38]. Indeed, heparin can prevent the tumor cell-platelet aggregation, through blocking P-selectin on platelet surfaces [39-41].

Heparins can also inhibit angiogenesis and lymphangiogenesis, typically associated with tumor growth and metastasis [40]. It can suppress the VEGF-C/VEGFR-3 molecular pathway, reducing lymphatic vessel formation and the metastatic progression in the lymph node. Additionally, heparin interferes with the CXCL12-CXCR4 signally pathway, which has a critical role in cancer cell migration and homing [40].

Other studies have revealed the potent anti-inflammatory properties of heparin. As previously mentioned, heparin can inhibit selectins, particularly P-selectin and L-selectin [42], which mediate leukocyte recruitment and adhesion to endothelial cells during the initial stage of inflammation [43]. Sulfate groups at the C6 position on GlcN residues of heparin chains can block the initiation of cell adhesion of P-selectin and L-selectin [42]. Heparin has also been found to suppress the activation of nuclear factor kappa B (NF-κB), a key transcriptional activator of inflammatory genes. Recent studies reveal that low-molecular-weight heparin (LMWH) binds with high affinity to pro-inflammatory cytokines such as interferon-gamma (IFNγ) and interleukin-6 (IL-6), inhibiting their action on cellular receptors and the downstream signaling cascades [44, 45].

Finally, increasing attention has been directed towards the antiviral activity demonstrated by heparin. Since heparan sulfate and heparin share a similar structure, the latter can mimic the cellular HS chains of HSPGs, acting as a competitive inhibitor. Thanks to its highly negative charge, heparin can interact electrostatically with amino acids that frequently decorate the viral recognition proteins. This impairs viral attachment, blocking the initial phase of infection [46-49]. Heparin has demonstrated efficacy against diverse viral families, including flaviviruses [50, 51], herpesviruses [52], retroviruses [53], and coronaviruses [46-48, 54-56]. Undoubtedly, the recent outbreak of the SARS-CoV-2 pandemic (that will be discussed in Chapter 2) has prompted several efforts in the detailed elucidation of the role of heparin in preventing viral infections. Exogenous heparin and non-anticoagulant derivatives can directly compete with the HSPGs expressed on the cell surface, which act as a co-receptor, favoring the attachment of the virion to the cellular membranes, facilitating the subsequent binding to human angiotensin converting enzyme (hACE2) (see section 2.5 for further details). The latter aspect was first demonstrated by Clausen et al. [57], who observed that, after the enzymatic degradation of the cellular HS chains of various human cell line (H1299, A549, Hep3B, A375) with heparin lyase (HSase), the binding of SARS-CoV-2 spike protein was dramatically reduced.

These findings were further validated in the infection assay, using pseudo-typed vesicular stomatitis virus (VSV) engineered to express the full-length SARS-CoV-2 spike protein along with Green fluorescent protein (GFP) or luciferase (to monitor the cell infection). In these assays, the infection rate was reduced by approximately 2- to 3-fold in Vero E6 cells and 6- to 7-fold in Hep3B cells following the same HSase treatment.

Parallelly, the viral inhibition potency of various heparin preparations and tissue-derived HS against spike binding was also assessed. Unfractionated heparin (UFH), extracted from porcine mucosa, exhibits the strongest inhibition, reaching IC_{50} in the scale of μM/mL. Notably, the MST-derived heparin and split-glycol heparin (nonanticoagulant heparins) showed a comparable inhibition profile, with an IC_{50} =

0.12 μg/mL and 0.04 μg/mL, respectively, in H1299 cells. These results indicate that heparin's antiviral activity is independent of its anticoagulant function, supporting the potential for safer therapeutic heparin analogues. Hence, in the context of SARS-CoV-2 infection, what emerged is the double role that heparin possesses: it can act both as a decoy receptor and as an entry inhibitor, making heparin and engineered heparin mimetics appealing therapeutic candidates to block early events in SARS-CoV-2 infection by targeting virus–glycan interactions.

Unfortunately, current knowledge regarding the roles of heparan sulfate and heparin chains within the cellular microenvironment remains extremely limited. This limitation stems from the inherent complexity of their interactions, which span a wide range of biological entities, the scarcity of structural data, as there are no crystal structures of such complexes available, and the relatively small number of GAG–protein interactions studied compared to the vast number of proteins that potentially interact with GAGs. As a result, elucidating the roles of HS and heparin in both physiological and pathological contexts remains a critical objective, particularly at the molecular level.

Chapter 2. The role of glycosaminoglycans in SARS-CoV-2 infection

2.1 The outbreak of the SARS-CoV-2 pandemic

In 2019, a new coronavirus emerged in the city of Wuhan, China, the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). This virus was the cause of the Coronavirus Acute Respiratory Disease 2019 (COVID-19) [58, 59] pandemic, declared by the World Health Organization (WHO) in March 2020 [60]. SARS-CoV-2, as the previously known SARS-CoV and Middle East Respiratory Syndrome Coronavirus (MERS-CoV), is a lipid-enveloped positive-sense RNA virus, belonging to the Betacoronavirus genus [61]. The virus enters the body through the upper respiratory tract, affecting lung functionalities, provoking pneumonia, and acute respiratory distress syndrome (ARDS). However, the gastrointestinal, renal, and central nervous systems can also be affected by this pathogen. In 2024, the WHO [62] reported that the total number of confirmed cases was over 771 million, including approximately seven million deaths. The average case fatality rate (CFR) of the virus was estimated at around 1-5% [63], which is notably lower than compared to the rate of SARS-CoV-1 (9.7%) [64] and MERS-CoV (34%) [65]. This discrepancy can be attributed to the fast development and distribution of the vaccine, which was supposed to mitigate the virus diffusion and the adverse effects of the infection [66]. However, due to the relatively fast mutation of SARS-CoV-2, the vaccine shows efficacy only for a limited amount of time. Therefore, the design of new antiviral drugs, as well as new biochemical routes to inhibit the life cycle of SARS-CoV-2, is crucial, involving both applied and basic structural biology investigation, aiming to identify alternative strategies for treating the disease. The following chapter describes the structure and molecular mechanism in the early stage of SARS-CoV-2 infection.

2.2 The SARS-CoV-2 spike protein structure

The SARS-CoV-2 infection begins through interaction between the recognition protein spike (S) and the human angiotensin converting enzyme (hACE2) receptor, which is expressed on the host cell surfaces (**Figure 7**) [67-69]. The formation of hACE2-S protein complex drives the fusion of the virus envelope with the cellular membrane, followed by the release of viral RNA inside the infected cell (**Figure 7**).

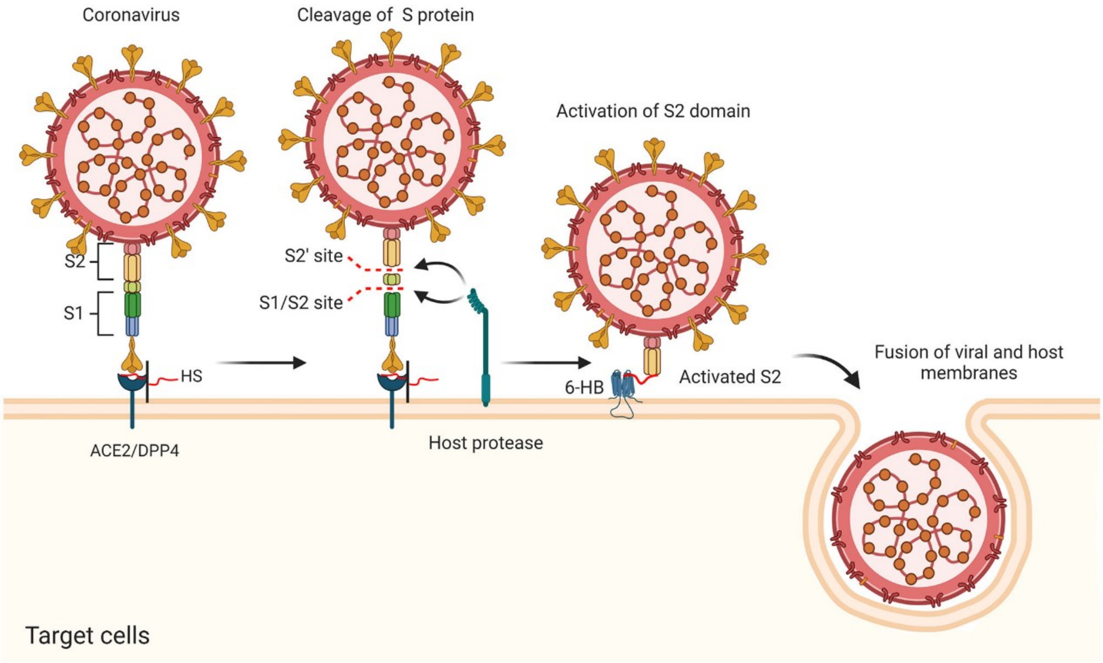


Figure 7 – Infection mechanism of SARS-CoV-2. After this interaction, the viral envelope fuses with the host cell membrane, allowing the viral RNA to enter. Inside the cell, the virus hijacks the host machinery to replicate and produce new viral particles. *Image reprinted from Yu M., Zhang T., Zhang W., Sun Q., Li H., and Li J.-P. (2021). Front. Mol. Biosci.* https://doi.org/10.3389/fmolb.2020.628551

Currently, the structure and function of the SARS-CoV-2 S protein are well-established (**Figure 8a**), due to the rapid effort made to characterize the virus during the initial outbreak of the COVID-19 pandemic. The spike protein is a transmembrane trimer protein that is exposed on the virus surface, forming the characteristic crown that is visible under the microscope [70] (**Figure 8b**).

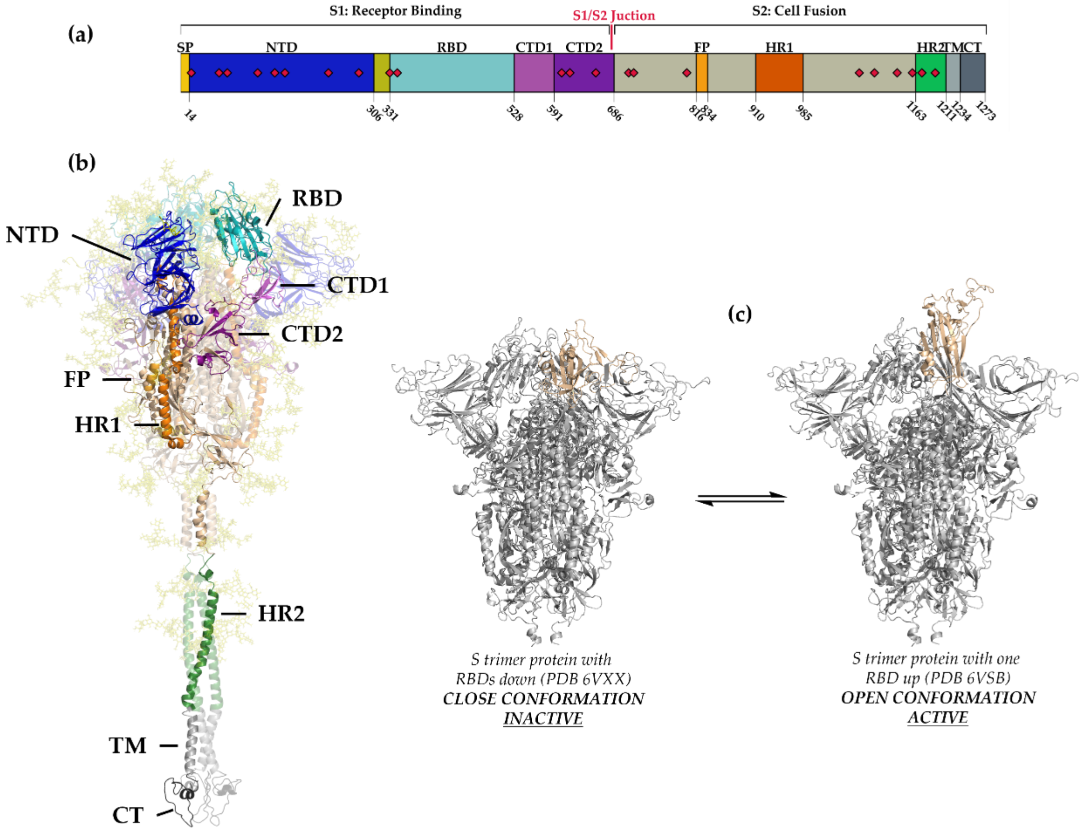


Figure 8 – Schematic representation of the full-length SARS-CoV-2 spike (S) protein organization. (**a,b**) Sequence and 3D structure of the S trimer protein: each subdomain in the 3D structure is colored with the same color code and labelled (NTD = N-terminal domain, RBD = receptor binding domain, CTD1 and CTD2 = C-terminal domain 1 and 2, FP = fusion peptide, HR1 and HR2 = heptapeptide repeat sequence 1 and 2, TM = transmembrane domain, CT = cytoplasmic tail. The 22 N-glycosylation sites are represented as red diamonds on the protein sequence. (**c**) 3D structure of the closed (PDBid: 6VXX) and open (PDBid: 6VSB) conformation, respectively. The RBD region (highlighted in light yellow) is involved in the down-to-up conformational change.

The spike (S) protein is classified as a type I membrane fusion protein, consisting of an extracellular N-terminus, a transmembrane (TM) domain embedded in the viral membrane, and a short intracellular C-terminal tail.

In each spike monomer, two main domains can be distinguished: the subunit 1 (S1 - residues 14–685), responsible for the cell and receptor binding, and the subunit 2 (S2 - residues 686–1273), which participates in the membrane fusion. These two subunits are connected by the S1/S2 junction (**Figure 8a,b**), a multi-basic cleavage site that represents a novel feature of SARS-CoV-2, not observed in other closely related Betacoronaviruses [71].

In turn, the subunit S1 is further divided in the N-terminal domain (NTD, residues 14–305), the receptor-binding domain (RBD, residues 319–541), which includes the receptor-binding motif (RBM, residues 438–508) responsible for the direct interaction with the hACE2 receptor, and two C-terminal domain CTD1 and CTD2 (residues 542–686). The subunit S2 includes the fusion peptide (residues 788–806), the heptapeptide repeat sequence 1 (HR1, residues 912–984) and 2 (HR2, residues 1163–1213), the transmembrane domain (TM, residues 1214–1237) and the cytoplasm domain (CT, residues 1238–1273) [70] (**Figure 8a**).

The RBD is the subunit directly involved in the interaction with hACE2. It can adopt two main conformations (**Figure 8c**):

- **The closed conformation (RBD-down)**, in which the RBD is almost hidden by the plane of the S1 subunit and the RBM adopts a folded position. This conformation is considered inactive; therefore, interaction with hACE2 cannot take place.
- **The open conformation (RBD-up)**, in which the RBD region is exposed to the solvent, making the RBM accessible to external molecules such as hACE2 or glycans like heparan sulfate (i.e., the heparan sulfate of the extracellular matrix of the host cell).

Once the hACE2–RBD complex is formed, the S1/S2 junction can be proteolytically cleaved by the transmembrane protease serine 2 (TMPRSS2) (**Figure 7**). Therefore, the subunit S2, using the fusion peptide, drives the fusion between the viral and cell host membranes, allowing the release of the viral RNA genome into the cytoplasm of the cell, where it can exploit the translational machinery of the host to synthesize the viral proteins and replicate its RNA. Newly assembled virions are then released to infect neighboring cells, perpetuating the cycle of infection [72-74].

2.3 The SARS-CoV-2 variants of concern (VOCs)

Like all RNA viruses, SARS-CoV-2 also has an inherently high mutation rate [75], enabling it to adapt to new hosts and environmental pressures. Indeed, the RNA-dependent RNA polymerase (the enzyme that catalyzes the synthesis of a DNA template from a single RNA strand) lacks an efficient proofreading mechanism. This leads to an accumulation of mutations during viral replication. Additionally, point mutations can occur due to genetic recombination of the virus during the contemporary infection of two distinct SARS-CoV-2 variants.

Consequently, several genetically distinct lineages of this virus have emerged over time, which exhibit characteristics that significantly impact public health measures, diagnostics, treatment efficacy, and vaccine performance.

To systematize global monitoring, the WHO has categorized SARS-CoV-2 variants into three major groups:

- **Variants under monitoring (VUMs)** correspond to all the SARS-CoV-2 variants with genetic modification that are suspected to affect virus features, but the effective phenotypic and epidemiologic impact is uncertain. These variants are closely tracked to determine if they engender any novel risks for public health.
- **Variants of interest (VOIs)** correspond to all SARS-CoV-2 variants with genetic mutations that are known or predicted to confer to the virus potentially dangerous characteristics to human health, such as higher transmissibility, immune evasion, and disease severity. VOIs have been identified as causing significant community transmission in several countries.
- **Variants of concern (VOCs)** are all the SARS-CoV-2 variants that meet the VOI definition and other criteria, such as a rapid increase in disease severity, potential impact on the health system due to a higher number of hospitalized

patients, and a significant decrease in the availability of vaccines to mitigate severe disease. These variants are key models for understanding the molecular and evolutionary mechanisms underlying viral adaptation, and for devising strategies to counteract their impact on the pandemic control efforts. To date, 5 different SARS-CoV-2 VOCs have been identified: Alpha (B.1.17) [76], Beta (B.1.135) [77], Gamma (P.1) [78], Delta (B.1.617.2) [79], and Omicron (B.1.1.529) [80] (**Figure 9** and **Table 1**).

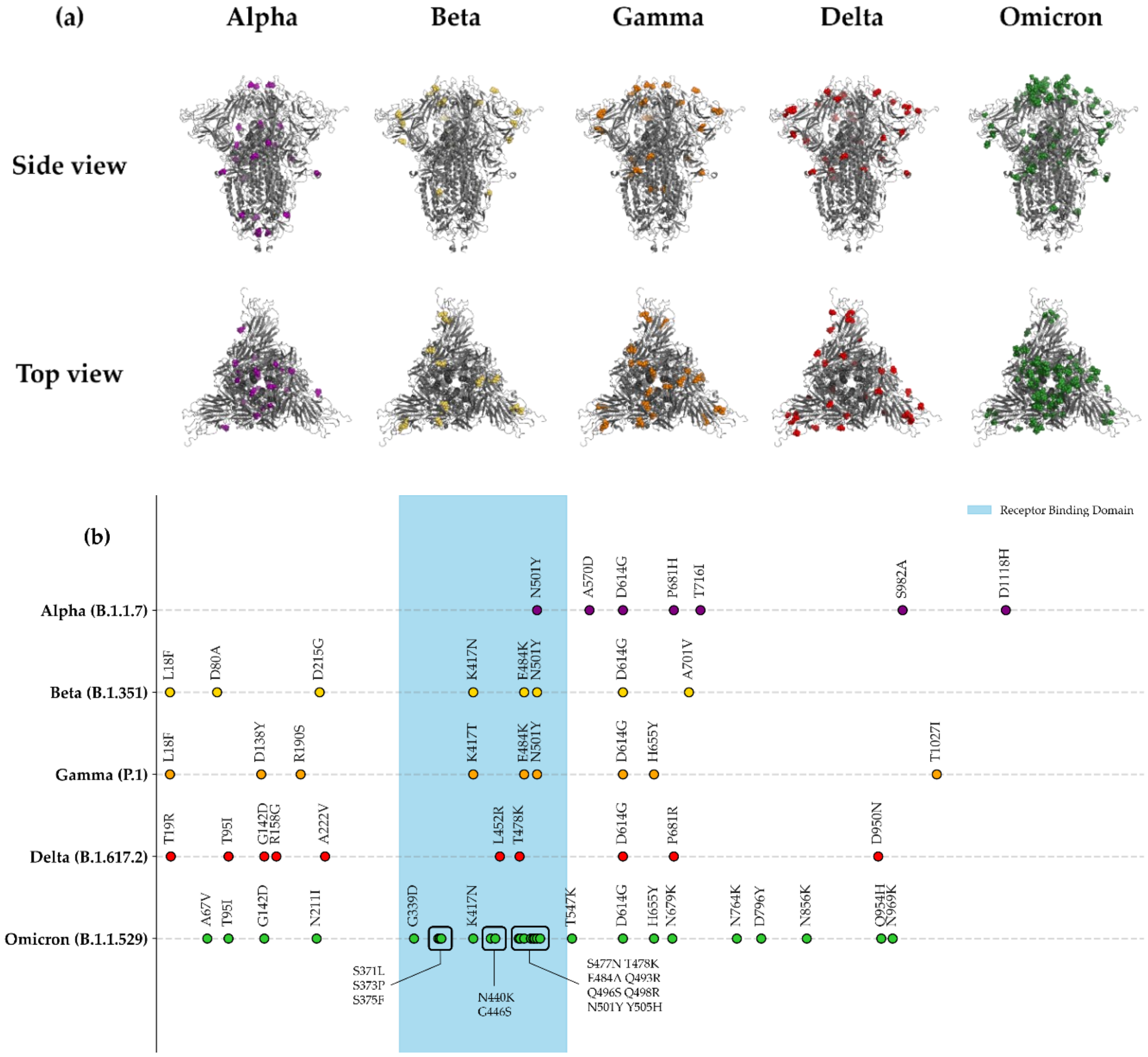


Figure 9 – Schematic representation of spike protein (1-1146 residues) mutations in the five SARS-CoV-2 variants of concern (VOCs). (**a**) Side and top views of the spike protein (grey cartoon), where the mutations are represented as spheres: purple for Alpha, yellow for Beta, orange for Gamma, red for Delta, and green for Omicron. (**b**) Detailed map of spike protein mutations for each variant, using the same color code described above.

The Alpha (B.1.1.7) [76] variant, first identified in the United Kingdom, contained eight amino acid mutations within the spike protein, with N501Y, located in the receptor-binding domain (RBD), acting as the principal driver of enhanced transmissibility presumably via increased affinity towards the hACE2 receptor [81]. The Beta (B.1.351) variant, which has emerged in South Africa, carries the K417N–E484K–N501Y triad within the RBD. It has been demonstrated that E484K and K417N promote the virus escape from the immune system by altering key antigenic residues [82]. Moreover, this mutation triad induces a synergistic effect: the reduction of the affinity between RBD and hACE2 correlated to the K417N mutation is compensated by E484K and N501Y, thereby maintaining the efficiency of interaction between these two subunits [83, 84]. Analogously, the Gamma (P.1) variant from Brazil demonstrated similar RBD substitutions, with the only exception of K417T in place of K417N. Hence, as observed for the Beta variant, the detrimental effect induced by this mutation is compensated by N501Y, ensuring the preservation of structural integrity of the RBD, while optimizing the capacity of this variant to evade the host immune system and the engagement of the hACE2 [85]. The globally dominant Delta (B.1.617.2) variant was defined by L452R and T478K mutations in the RBD, which enhance hACE2 affinity and antibody resistance, respectively [85]. Furthermore, the P681R mutation at the furin cleavage site increases the basicity of the polybasic sequence (residues 681-685, PRRAR), facilitating additional interactions with furin and the S1-S2 cleavage site. This may improve the S1/S2 cleavage efficiency, leading to a higher rate of membrane fusion and internalization and thus greater transmissibility [86]. The Omicron (B.1.1.529) variant represented a significant evolutionary expansion, containing over 30 spike mutations (**Figure 9** and **Table 1**), which stabilize the RBD domain of spike protein, increasing the immune evasion while preserving sufficient hACE2 binding through complex compensatory mutations [82].

It should be noted that the different efficiency in virus replication/virulence across the SARS-CoV-2 variants has been attributed not only to a greater affinity between the RBD and hACE2, but also to an enhanced affinity shown by the mutant spike and heparan sulfate on host cell surfaces. In fact, it was demonstrated that mutations in the S-protein progressively increase their overall formal charge, thereby presumably enhancing electrostatic interactions with both ACE2 and HS [87]. Specifically, the trimeric S-protein (residues 13–1140) also exhibits a change in net charge from +3 in the ancestral wild-type (WT) strain to +18 in Delta, and +24 in Omicron (**Table 1**), reflecting a greater gain in cationic character that is supposed to boost the viral binding avidity toward the hACE2. Cumulatively, these molecular changes underscore the evolutionary pressure of SARS-CoV-2 for maximizing its transmission ability through the coordinated optimization of ACE2 binding affinity, HS binding, immune evasion, and spike processing efficiency.

Table 1 – Summary of mutations across all SARS-CoV-2 VOCs (first column), including the total number of amino acid substitutions within the RBD (2nd column) and within the full S monomer (3rd column), as well as the overall net charge of the spike trimer (4th column).

	Mutation	RBD mutation	Total mutation	Total charge
Alpha (B.1.1.17)	N501Y, A570D, D614G, P681H, T716I, S982A, D1118H	3	7	3
Beta (B.1.135)	L18F, D80A, D215G, R246I, K417N, E484K, N501Y, D614G, A701V	3	9	6
Gamma (P.1)	L18F, T20N, P26S, D138Y, R190S, K417T, E484K, N501Y, D614G, H655Y, T1027I	3	11	15
Delta (B.1.617.2)	T19R, V70F, T95I, G172D, R158G, A222V, W258L, K417N, L452R, T478K, D614G, P681R, D950N,	3	13	18
Omicron (B.1.1.529)	A67V, H69-, V70-, T95I, G142-, V143-, Y144, Y145D, N211-, L212I, G339D, S371L, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, Q496S, Q498R, N501Y, Y505H, T547K, D614G, H655Y, N679K, P681H, N764K, D769Y, N856K, Q954H, N969K	16	35	24

2.3.1 Structural comparison of the RBD domain in VOCs

As reported in paragraph 2.2, the receptor binding domain (RBD) of the SARS-CoV-2 spike comprises residues 319-541 within the S1 subunit (**Figure 10**) and functions as the primary viral interface for host cell recognition, mediating the interaction with the hACE2 receptor. This engagement is essential for viral attachment and subsequent membrane fusion necessary for infection.

The structural characterization of the RBD domain among the SARS-CoV-2 VOCs reveals that mutations in this region can significantly alter both the 3D architecture and functional properties of the spike protein-HS interaction [88, 89]. Although the RBD region maintains a conserved core structure that is composed of five β-sheets surrounded by short α-helices and connecting loops, the RBD is characterized by high sequence and conformational diversity at the RBM, corresponding to a less structured region that folds upon contact with the hACE2 interface.

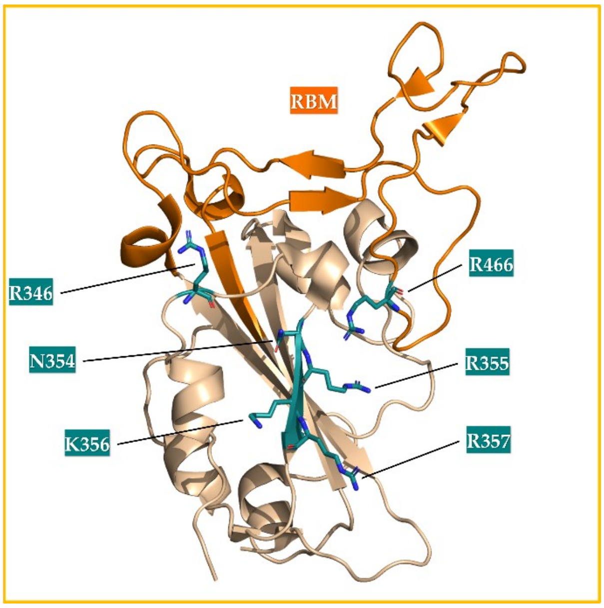


Figure 10 – Enlargement view of the RBD region. The key residues involved in the HS binding are depicted as stick (teal for carbon, red for oxygen, and blue for nitrogen), while the RBM (receptor binding motif) is colored in orange (residues 438-508).

In fact, the VOCs accumulate mutations predominantly within this RBM interface region rather than the structural core, with Omicron displaying the most extensive remodeling, harboring 15 RBD mutations relative to the wild-type strain, while two mutations are observed in Delta and one in Alpha [90-92] (**Figure 11**).

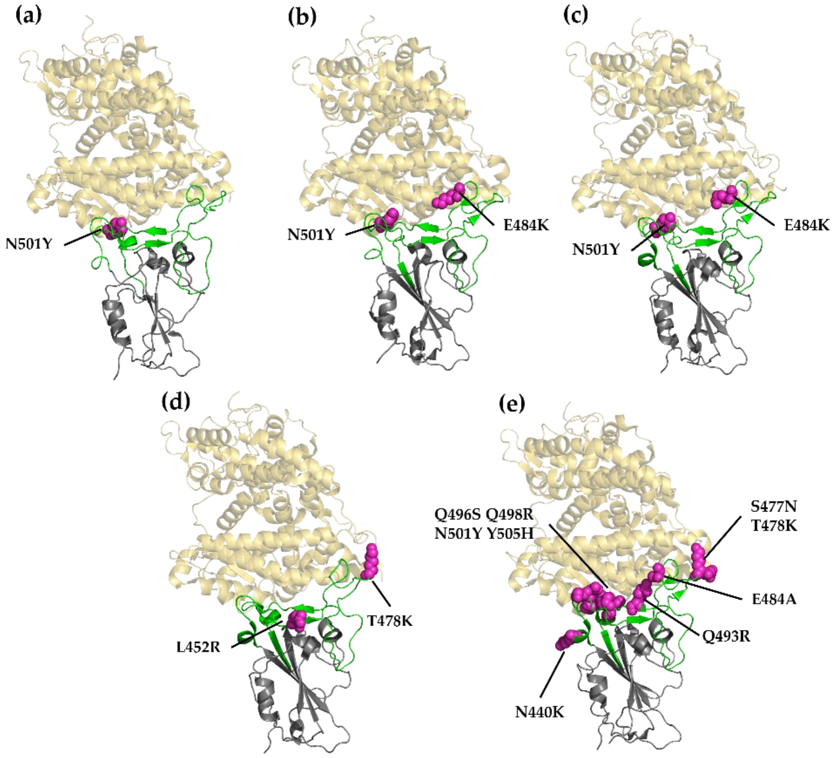


Figure 11 – Structural variability in the RBM among SARS-CoV-2 VOCs. The RBD domain of spike protein from the Alpha (panel a – PDBid: 7FEM), Beta (panel b - PDBid: 7VXD), Gamma (panel c – PDBid: 7V84), Delta (panel d – PDBid: 7V8B) and Omicron (panel e – PDBid: 7WBP) variants is shown in complex with hACE2, is reported. The RBD and hACE2 are represented as cartons, colored grey and yellow, respectively. The RBM within the RBD is highlighted in green, and the residues carrying a mutation in the RBM are labelled and depicted as spheres.

Molecular docking and molecular dynamics simulations predict that these mutations fundamentally reshape the binding interface, enhancing the conformational stability, and increasing the electrostatic attraction and binding affinity in comparison with wild-type strain [89]. Interface analyses reveal that while wild-type strain and Alpha variants maintain 5–7 hydrogen-bonded interactions with hACE2, Gamma and Delta increase this to 9–10 interactions, and Omicron achieves 10–12 polar interactions, including two critical ionic bonds (Q493R-NH2 --- OE1-E35 and Q493R-NH2 --- Oδ1-Asp38 on ACE2), correlating directly with observed clinical transmissibility rates [93]. Interestingly, the increase in affinity between RBD and hACE2 in going from wild-type to Alpha and Omicron variants is figured out to correlate with the observed transmissibility rate of SARS-CoV-2 [93]. This improvement especially involves the latest variant of SARS-CoV-2. The Delta RBD shows that the mutated residues L452R and T478K reinforce the interaction between residue 452 and the loop 469-492, which consequently promotes the engagement of Q493 in hACE2 binding [94]. The Omicron RBD presents the highest binding affinity that correlates with multiple substitutions of uncharged residues to positively charged arginine and lysine at critical positions, including N440K, Q493R, G496S, Q498R, and N501Y. Specifically, Omicron's mutations induce conformational shifts in RBD loops that bring them in proximity to hACE2, increasing root-mean-square deviation to approximately 3 Å compared to wild-type. Furthermore, the forehead mentioned additional ionic interactions increase the binding energy by around 50–75% for Omicron and around 10% for the Delta variant [95, 96].

As described in section 2.2, the RBD of SARS-CoV-2 changes between the closed (inactive) to open (active) conformation (**Figure 8c**), regulating the accessibility of the spike protein to the hACE2. Mutations in the SARS-CoV-2 RBD alter the equilibrium between these two limit conformations [97]. The closed conformation of the wild-type RBD occurs about 55% of the time, limiting access to the ACE2 receptor. The Alpha and Beta variants tip the balance, with a rise in the open conformations by 20% and increase the probability of a favorable contact between the spike protein and hACE2. The Delta variant significantly favors the open form, sometimes transitioning into the closed conformation and adopting an alternative reversed open conformation that promotes the shielding of the spike protein to the immune system. The Omicron variant has the greatest deviation, with approximately 95% of the RBD in the open conformation to maximize the chances to engage hACE2, possibly enhancing the infectivity of the virus. In summary, the equilibrium between these limits conformation of the RBD reflects alterations in SARS-CoV-2 variants, allowing it to optimize the function of the recognition protein spike in terms of S-hACE2 interaction and immune escape [97].

2.4 The SARS-CoV-2 glycosylation

2.4.1 Glycosylation as a key contributor to protein structural and functional diversity

Proteins can be glycosylated. Glycosylation is a highly regulated post-translational modification that involves the covalent attachment of glycans to specific amino acids. This process predominantly occurs in the endoplasmic reticulum (ER) and in the Golgi apparatus, where the joint action of a series of glycosyltransferases and glycosidases catalyze the sequential addition and processing of sugar residues. The two major classes of protein glycosylation are the **N-linked** and the **O-linked glycosylation**.

N-linked glycosylation involves the attachment of glycans to the amid nitrogen of asparagine residues that belong to the consensus sequence Asn-X-Ser/Thr/Cys (with X ≠ Pro) [98]. This modification occurs during protein synthesis in the ER, wherein a preassembled oligosaccharide is transferred to the nascent protein.

The N-glycans share a conserved core pentasaccharide **$Man_3GlcNAc_2$** (chitobiose), but they diverge in their outer chain composition due to different enzymatic processing. These sites can be categorized into three main glycan types [99] (**Figure 12**):

- **High mannose** glycans, in which only mannose (Man) oligosaccharide chains are attached to the chitobiose core.
- **Hybrid glycans**, characterized by the presence of $\beta(1 \rightarrow 2)$ GlcNAc unit or $\beta(1 \rightarrow 2)$ linked to terminal Man residues attached to the chitobiose core

(referred to as antennae). The number of GlcNAc moieties bound to this arm defines the antennary multiplicity of these glycans.

- **Complex-type glycans**, in which a GlcNAc unit is attached to both α-3 and α-6 mannose sites of chitobiose core and the corresponding chains are further decorated with fucose, galactose, N-acetylgalactosamine, and sialic acid.

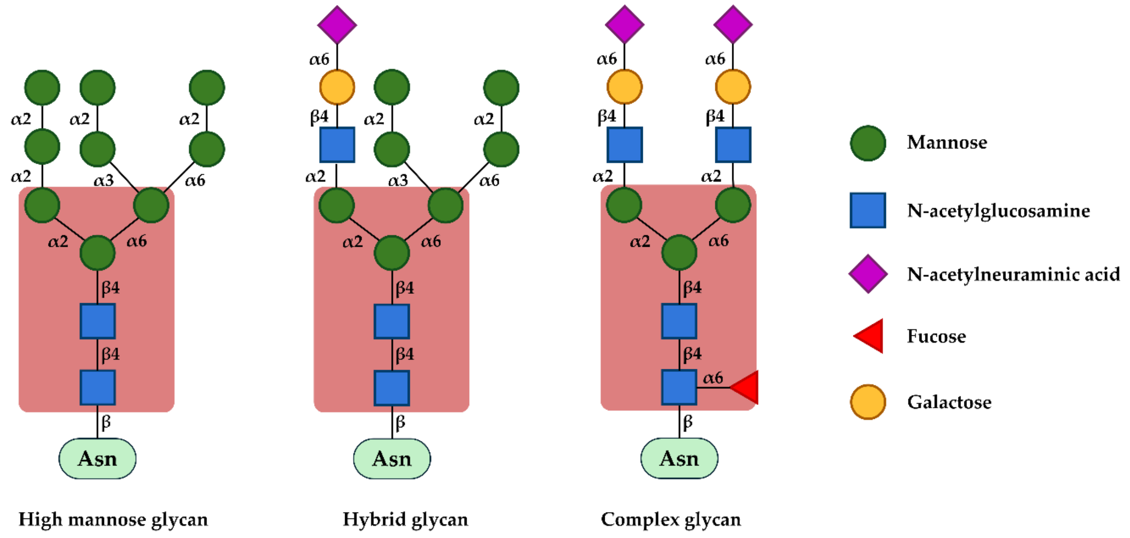


Figure 12 – Schematic representation of the high mannose, hybrid, and complex N-glycans classes. The conserved core pentasaccharide **$Man_3GlcNAc_2$** (chitobiose) is marked with a red rectangle in each glycan. The monosaccharide units are depicted using the SNFG nomenclature [100, 101]. *Image adapted from Gudelj, Ivan & Lauc, Gordan. (2018). Current Cardiovascular Risk Reports. 12. https://doi.org/10.1007/s12170-018-0579-4*

In contrast, in **O-linked glycosylation,** the glycans are bound to the hydroxyl oxygen of serine and threonine residues. Additionally, this modification occurs mainly in the cytoplasm and nucleus [102], does not require a consensus sequence and hence involves the gradual addition of monosaccharide units.

Glycosylation has a strong impact on protein folding, stability, and trafficking, and plays a key role in cell-cell communication, cell adhesion, immune recognition, and signal transduction [103]. Aberrations in glycosylation pathways are associated with different diseases, including cancer [104], blood disorders [105] and autoimmune diseases [103, 106].

2.4.2 Glycosylation landscape of the SARS-CoV-2 spike protein

The SARS-CoV-2 spike protein is heavily glycosylated: indeed, each spike protein monomer contains 22 N- and several O-linked glycosylation sites. All of them form a glycan shield that masks over 40% of its surface, helping the virus to evade antibody recognition [107] (**Figure 13**).

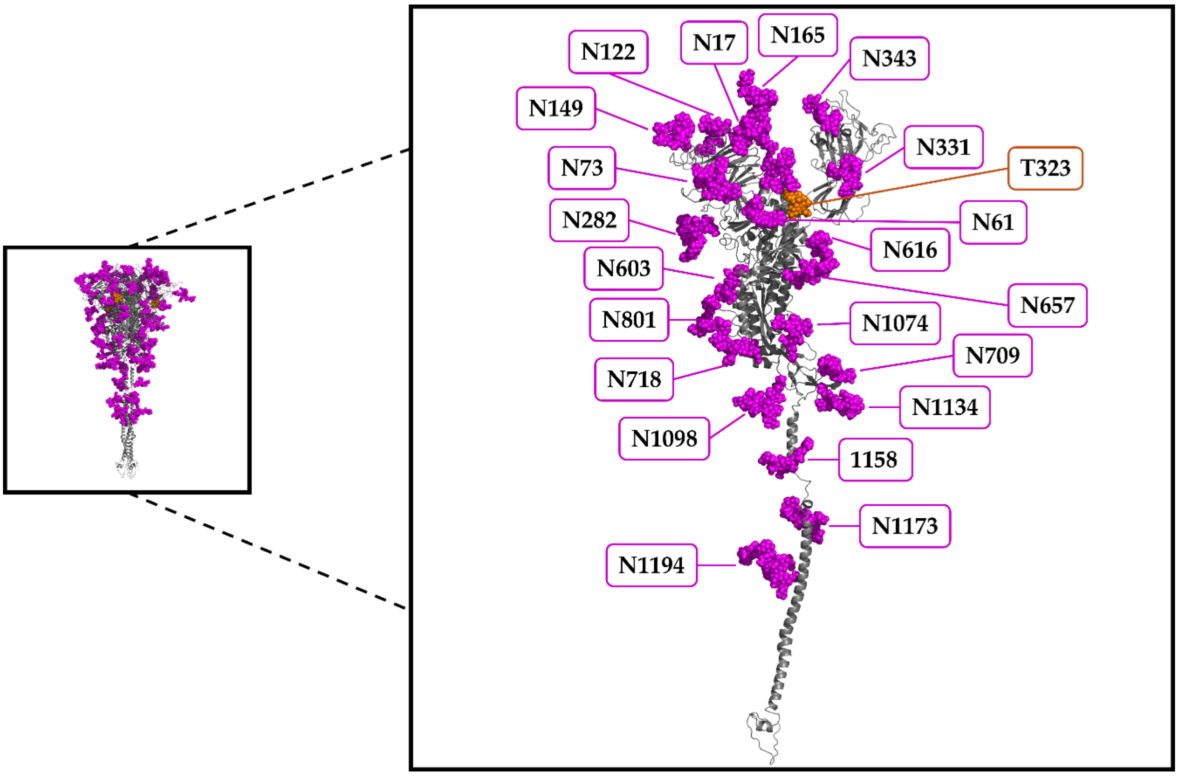


Figure 13 – N- and O-linked glycosylation sites attached to the SARS-CoV-2 spike trimer (coordinates from PDB ID: 6VXX; closed conformation). The spike trimer is shown as a grey cartoon representation, while the 22 N-linked and one O-linked glycans are displayed as magenta and orange spheres, respectively (left panel). The right panel shows an enlargement of a single monomer with site-specific glycan residue assignments based on mass-spectrometric mapping [108].

Several efforts have been made to characterize the composition of each of these N-glycans: N234 and N709 sites are principally occupied by high-mannose glycans, while a mixture of high-mannose and complex-type glycans is found at sites N61, N122, N603, N717, N801, and N1074, while the remaining 14 sites are dominated by complex-type glycans. The N-glycosylation sites located in the RBD region are highly conserved among the SARS-CoV-2 VOCs, especially for residues N331 and

N343, which show a sequence identity almost equal to 100% [99]. While N331 resides in the highly flexible region linking the RBD to the N-terminal domain of the S protein, the N343 glycan stretches across the RBD surface, connecting the two helical regions flanking the central beta-sheet core. This suggested that this glycan could obstruct a substantial part of the domain involved in HS binding within the RBD [109-111].

In contrast, comprehensive characterization of the O-linked glycosylation landscape of SARS-CoV-2 is more challenging. This complexity arises from the absence of a consensus sequence for O-glycans, their high structural heterogeneity, generally low abundance, and the limitations of current analytical methods available for their detection [112, 113]. The most studied O-glycosylation sites are T323 and S325, both of which lie within the receptor-binding domain (RBD) of the S1 subunit. The occurrence of proline at position 322 near T323 justifies the O-glycosylation specificity at this site because proline residues are generally observed near O-glycosylation sites [114, 115].

The other sites of O-glycosylation have also been identified in other VOCs, like sites at T315, T376, and regions near the furin cleavage site. O-glycosylation site occupancy is generally lower than that of N-linked sites and significantly varies among different expression systems and cell lines of the host [114, 116].

2.4.3 The functional role of the N343 glycan in SARS-CoV-2

The various glycan sites, bound to the SARS-CoV-2 spike protein, provide the virus with several important biological functions.

Particularly, the N343 glycan appears to be an essential structural and functional feature for the virus. This complex glycan (see section 2.4.1) is the biantennary core-fucosylated octasaccharide FA2G2 with the structure [GlcNAc^{1} (α1→6) Fuc^{8}] (β1→4) GlcNAc^{2} (β1→4) Man^{3} [(β1→3) Man^{4} (α1→6) GlcNAc^{5}] [(β1→6) Man^{6}(α1→6) GlcNAc^{7}] that works as a gatekeeper, controlling the motion of the RBD from the 'down' to 'up' conformation in the glycosylated spike (S) protein [117]. During this transition, N343 glycan pushes the RBD towards the active state by intercalating specific key residues (F490, Y489, F456, R457). Hence, this gating effect has a crucial role in the equilibrium of the open/closed state in the RBD [110, 117], and antibody escape of SARS-CoV-2 [118]. Indeed, this glycan is highly conserved across all SARS-CoV-2 spike proteins [119]. In the native wild-type strain and in Alpha and Beta variants, the removal or mutation of N343 causes a detrimental rearrangement of the RBD domain, leading to the RBM misfolding from the ACE-2 recognized conformation [110]. However, in Delta and Omicron variants, mutations have reinforced the RBD structure, compensating and reducing the stabilizing effect of the N343 glycosylation, whose absence is less critical for its structural integrity [110]. In summary, the N343 glycan is structurally essential for spike stability, gating function, and conformational flexibility. It acts both as a dynamic shield and as an active participant in spike opening and viral entry.

2.5 HS acts as a co-receptor in the SARS-CoV-2 cell infection.

The disseminated intravascular coagulation (DIC) is a condition frequently associated with severe viral infections and was also observed in SARS-CoV-2 infection [55]. This disease causes a systemic activation of blood coagulation, promoting the formation of microvascular thrombi throughout the body, which progressively damages all the organs. Surprisingly, during the COVID-19 pandemic, it was noticed that patients treated with heparin showed a lower mortality rate than those not treated. Based on this clinical evidence, researchers hypothesized that heparin's anticoagulant activity could improve the clinical picture of COVID-19 disease [46, 55, 120]. Despite that, considering that the recognition protein of several virus (coronavirus [121], flaviviruses [122, 123], HSV

[124]) target the HS chains of HSPGs, and observing also that the spike protein of SARS-CoV-2 presents clusters of positively charged side chains (Arg, Lys) on its S1 subunit, several research groups hypothesized that HS in the ECM plays a role in the early stages of SARS-CoV-2 infection [54, 125-127]. Early studies established that the HS of the host cells is required by SARS-CoV-2 to infect them [57]. Specifically, they hypothesized that the viral spike (S) protein recognizes and binds the polysaccharide chains of HSPGs on the host cell surface, favoring the virion anchoring [32, 57]. It was also hypothesized that HS may serve as a co-receptor, mediating the interaction between the RBD of the spike protein and hACE2, which is located on the cell host surface [128]. This event promotes the next step of the infection, corresponding to fusion between the membranes of the virus and the host cell. In addition, exogenous heparin can affect the ratio up/down conformation of the RBD in the spike, favoring the "up" configuration and consequently promoting these recognition events that precede the virus internalization [47, 120]. Several efforts have been made to uncover the key features of HS-RBD recognition at the molecular level, figuring out the possibility of interfering with the SARS-CoV-2 infection in the initial stages. The process is driven by electrostatic interaction between the positively charged amino acids that form clusters on the accessible surface of the S protein, as well as on the RBD subunit, which corresponds to the forefront tip of the spike, and the negatively charged sulfate of HS chains. Five distinct heparin binding domains (HBDs) have been identified on the spike: three located in the RBD region (**Figure 14**), one at the S1/S2 junction (R682, R683, R685), and another in the S2 site (K811, K814, R815), respectively [46, 47]. However, a complete understanding of the mechanism is still lacking, due to the complexity of the system, namely the presence of charge-based and non-specific interactions, potentially occurring at various sites across spike subunits.

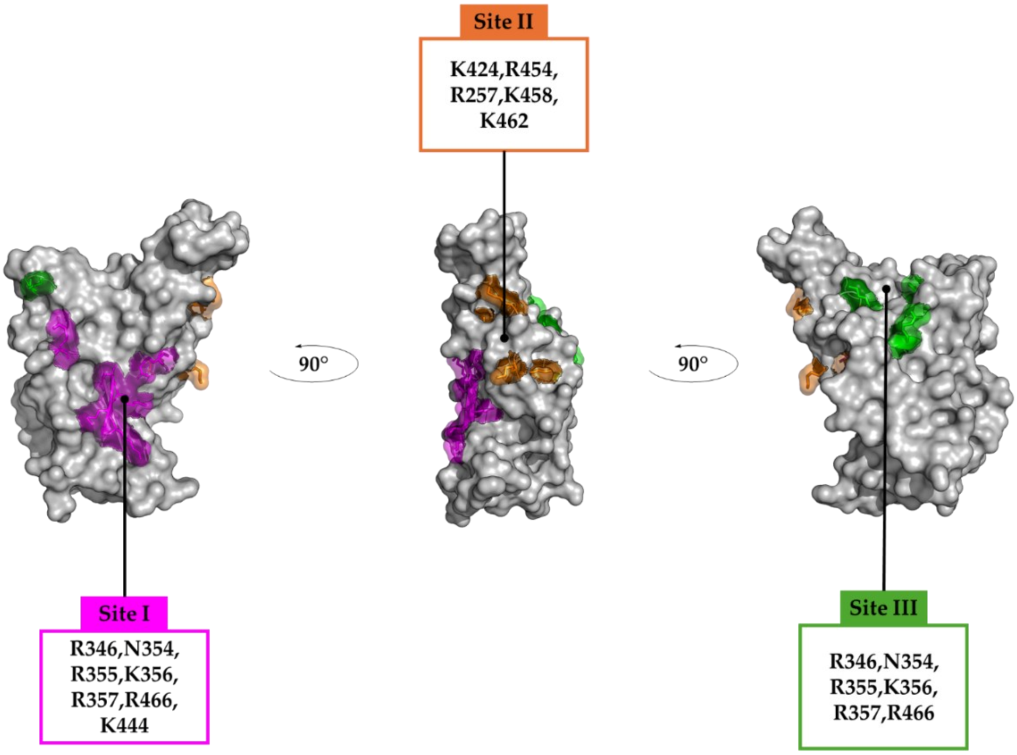


Figure 14 – The three heparan sulfate binding sites (I, II, and III, colored in magenta, orange, and green, respectively) identified in the RBD domain. Each site contains positively charged residues, which complement the negative charge of HS.

2.6 Aim of the Project

This thesis focuses on characterizing the interaction between heparan sulfate oligosaccharides and the RBD region of the spike protein of wild-type, Delta, and Omicron SARS-CoV-2 variants, to understand differences in their ability to bind HS and highlight how the co-receptor role of HS varies among these latter variants. Furthermore, the effect of the N-linked glycan, the biantennary core-fucosylated octasaccharide FA2G2 located at N343, on the ability of RBD to bind HS, was also investigated.

Since HS is covalently bound to the cell surface, its isolation and extraction are particularly challenging. To overcome this, our strategy employed mimetics sharing the same highly sulfated carbohydrate nature but more readily accessible. The selected candidates were HS-like (hexa) and pentosan polysulfate sodium hexasaccharide (PPS), both exhibiting structural features similar to those of cellular HS (see Section 4.1). The former was previously synthesized at the Chemical and Biochemical Research Institute ‘G. Ronzoni’, whereas the latter was designed based on preliminary structural information [129, 130]. An integrated workflow combining NMR spectroscopy and molecular modelling was developed to elucidate a putative interaction model between the RBD region of SARS-CoV-2 and these heparan sulfate mimetics, combining nuclear magnetic resonance (NMR) spectroscopy, docking calculation, and molecular dynamics (MD) simulations.

The workflow consists of 4 steps:

- ***Step 1: Docking calculations*** (Section 4.1)

Starting from the N343-glycosylated form of the X-ray or cryo-EM structures of the wild-type strain RBD and the Delta and Omicron variants, docking calculations were performed to identify the most relevant binding modes of two HS mimetics (hexa and PPS).

- ***Step 2: ^{1}H-STD NMR ligand epitope mapping*** (Section 4.2)

^{1}H-STD NMR experiments between the RBD of the Omicron variant and the HS-like mimetic hexa were carried out to experimentally determine the binding epitope of the hexasaccharide in the RBD-hexasaccharide complex.

- ***Step 3: Molecular dynamics simulation*** (Section 4.3)

Starting from the best-ranked docking poses obtained in Step 1 for the hexa ligand complexed with the RBD model of the Omicron variant, independent MD simulations were performed and analyzed to characterize the hexa–RBD interactions. For comparison, the same protocol was also applied to the hexa-RBD model of the wild-type strain.

- ***Step 4: NMR-driven analysis of MD runs*** (Section 4.4)

MD trajectories of the hexa – RBD complex (step 3) were subjected to theoretical STD NMR epitope mapping analysis, followed by on-the-fly comparison with the experimental STD NMR binding epitopes (step 2). The comparison was performed by using the RedMat approach [131], developed at Prof. Jesús Angulo's group (IIQ-CSIC), was employed during the secondment period carried out in this research group.

Chapter 3. Methodologies

3.1 Experimental methods

3.1.1 Saturation Transfer Difference (STD) NMR Spectroscopy

^{1}H STD-NMR is a very popular spectroscopic technique that is employed for studying low affinity ligand-macromolecule interactions at atomic resolution [132]. This ligand-based NMR technique was first reported by Meyer and Peters in 1999 [133], and leverages the Nuclear Overhauser effect (NOE). The NOE exploits changes in spin-active nuclei signal intensity due to spin polarization transferred from nearby nuclei via cross-relaxation mechanisms. When a protein is selectively saturated via direct radiofrequency (RF) irradiation, magnetization is transferred to a bound ligand by intermolecular NOE, allowing the detection of a binder even in a complex mixture.

An STD NMR experiment consists of two separated 1D NMR acquisitions (**Figure 15**):

- **Off-resonance spectrum:**

A 1H reference spectrum is recorded by applying the saturation pulse at a frequency that is far from the range of protein and ligand signals (typically 40-50 ppm), resulting in a conventional 1H NMR spectrum with unperturbed signal intensities.

- **On-resonance experiment:**

The macromolecule is selectively irradiated with an RF pulse during a precise time (saturation time - t_{sat}; typically, from 0.5 to 6 sec), targeting specific resonances (usually aliphatic or aromatic protons of the protein). The saturation propagates throughout the wide network of intramolecular ^{1}H-^{1}H dipolar coupling of the protein via spin diffusion (intramolecular NOE). If a ligand transiently binds the protein, the magnetization is transferred to ligand protons via intermolecular NOE effects. The closer the ligand protons contact the protein surface, the greater the reduction in signal intensity. Thus, the signal attenuation provides insights into the proximity of ligand atoms to the protein surface.

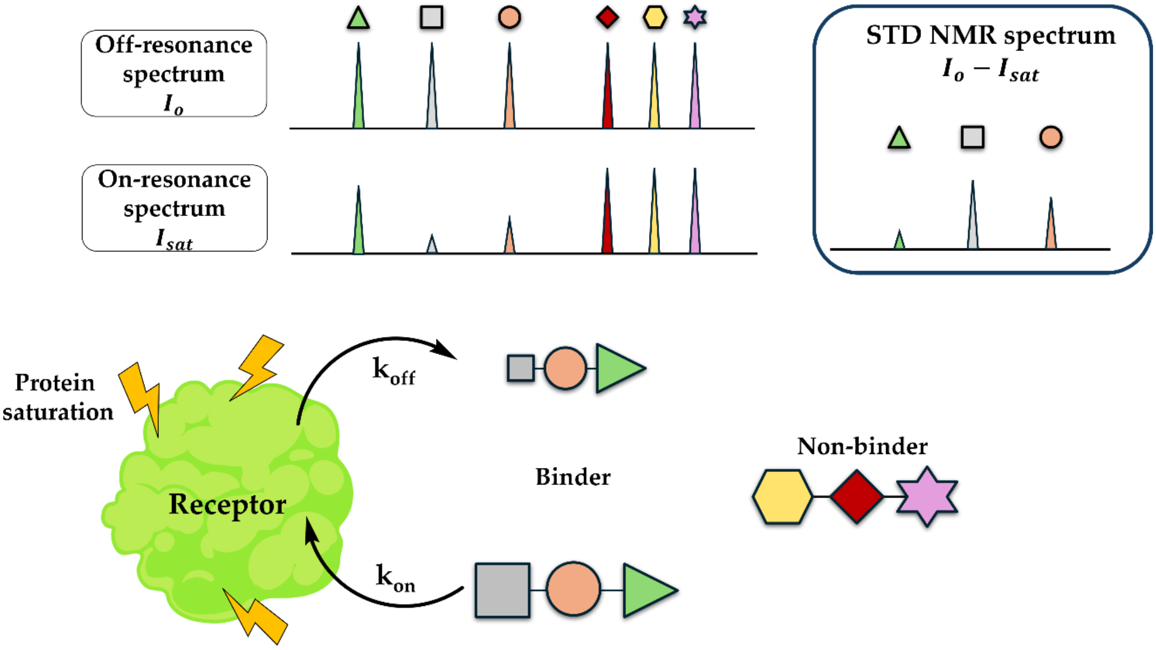


Figure 15 – Graphical representation of an STD-NMR experiment (top) performed of a protein in the presence of a pool of putative binding small molecules (bottom). *Image readapted from Angulo, J., Nieto, P.M. Eur Biophys J 40, 1357–1369 (2011).*

The STD spectrum is obtained by subtracting the on-resonance from the off-resonance spectrum. This subtraction cancels out the peaks of the non-binding molecules and of ligand protons, not establishing closed contacts with the protein. Hence, the STD spectrum only shows the signals of the ligand protons that are in direct contact with the protein. The signal intensities in the STD spectrum can be quantified using the **Equation 1**:

$$\mathbf{I_{STD} = I_0 - I_{sat}}$$

Equation 1 – Equation to calculate the STD NMR intensity of each ligand proton contacting the protein.

where $\mathbf{I_0}$ and $\mathbf{I_{sat}}$ are the signal intensities of the ligand protons in the off- and on-resonance spectrum, respectively. The magnitude of $\mathbf{I_{STD}}$ correlates qualitatively with the proximity of the ligand proton to the protein binding site. High $\mathbf{I_{STD}}$ values indicate that the ligand proton is proximal to the protein site. In contrast, low $\mathbf{I_{STD}}$ suggest that the proton is more distant from the recognition site. To obtain the "so-called" STD NMR epitope map, the STD intensities are normalized against the most intense one (**Equation 2**):

$$\mathbf{STD}_{\%} = \frac{(\mathbf{I_0} - \mathbf{I_{sat}})}{\boldsymbol{I_0^{max}}}$$

Equation 2 – Equation for the normalized STD NMR intensity.

However, the STD intensities obtained at a single saturation time are affected by the differences in longitudinal relaxation rates among ligand protons, as well as artefacts resulting from processes oFsurf protein-ligand rebinding. This can lead to wrong or biased epitopes. To overcome this problem, the initial growth rates approach was applied. It consists in collecting STD experiments at several saturation times in order to build mono-exponential curves of saturation time (t_{sat}) against STD intensity and determine the initial slope of each curve (STD_0).

This approximation was proposed by Mayer [134]: starting from the STD signal, the corresponding build-up curve, $STD(t_{sat})$, can be fitted using the following mono-exponential function (**Equation 3**).

$$STD(t_{sat}) = STD^{max}(1 - e^{-k_{sat} \cdot t_{sat}})$$

Equation 3 – The mono-exponential function employed to fit the STD signal build-up. $STD(t_{sat})$ represents the STD intensity of a given proton measured at different saturation time t_{sat}. The parameter STD^{max} is the equilibrium STD intensity, while the k_{sat} corresponds to the rate constant describing the growth of the STD build-up curve.

As shown in **Equation 3**, The STD signal increase exponentially towards the asymptotic maximum values (STD^{max}), which reflects the relaxation properties of the saturated ligand proton in the free state, with the growth rate determined by k_{sat}.

Determination of the initial slope of the STD build-up curve at short saturation times (ideally extrapolated to $t_{sat} \rightarrow 0$) circumvents the aforementioned artifacts, as the initial growth rate is independent of differential R_1 relaxation and rebinding effects (**Equation 4**).

$$STD_0 = \lim_{t_{sat} \to 0} \frac{dSTD(t_{sat})}{dt_{sat}} = STD^{max} \cdot k_{sat}$$

Equation 4 – Determination of the STD initial slope (STD_0).

The k_{sat} and STD^{max}, parameters can be obtained through least-squares fitting of **Equation 3**, from which the initial slope of the STD build-up curve (STD_0) is calculated

3.2 Computational methods

3.2.1 Docking Calculation

Most biological processes in cells require the interaction between two or more macromolecules. In the human genome, approximately 20,000 protein-coding genes have been identified [135], yet the structure, function, or subcellular localization of only about 60% of their products is currently known [136]. Among these, around 700 genes have a validated ligand [137], while nearly 3,000 proteins are considered potentially druggable, although no active compound have been discovered for them to date [136, 138]. The characterization of such molecular recognition events is essential for understanding the biochemical routes in which these druggable proteins are involved, as well as for their regulation. These challenging tasks can be accomplished through molecular docking calculations, which encompass a wide range of methodologies to determine the hypothetical 3D structure of a ligand-receptor complex and to estimate the corresponding binding affinity between, or more robustly, to estimate ta scale of affinity between a set of ligands and the target protein (receptor). In docking studies, the term 'ligand' usually indicates small molecules of different nature: oligosaccharides, oligopeptides, or oligonucleotides, which will be fitted on a limited portion (active site) of the receptor. The term 'receptor' indicates a macromolecule (frequently a protein) that is supposed to bind the proposed 'ligand'. The docking simulation requires the knowledge of the 3D structure of both ligand and receptor. High-resolution 3D representation of the protein target can be determined by other techniques like Cryo-Electron Microscopy (Cryo-EM) [139], Nuclear Magnetic Resonance spectroscopy (NMR) [140, 141] or X-Ray crystallography [142, 143]. However nowadays, the 3D structure of target proteins and/or their macromolecular complexes can be hypothesized using computational techniques such as homology methods and machine learning driven methods (i.e. AlphaFold2 [144]). The 3D structure and conformation of ligands, particularly in the case of small molecules, are frequently obtained from different experimental sources (X-ray diffraction, NMR spectroscopy). Nevertheless, when such data are unavailable, input conformations can also be generated de novo [145], suing theoretical techniques: quantum chemical methods, molecular mechanics and molecular dynamic simulations. The docking calculation generates (hypothetical) 3D geometries of the ligand-receptor complex, which are commonly referred to as binding poses. These poses correspond to distinct local minima on the free-energy landscape of the predicted complex according to the empirical score function, and in most cases can be considered reasonable approximated geometries for the ligand-receptor complex. The docking process involves two key stages: sampling and scoring. During the sampling stage, the search algorithm [146] (matching algorithm [147-149], incremental construction [150-152], multiple copy simultaneous search [153, 154], Monte Carlo [155, 156], Genetic algorithms [157-159]) explores the possible conformations and orientations that the ligand can adopt within the active site of the receptor.

In the scoring stage, each ligand-receptor complex sampled is evaluated and ranked using a quantitative scoring function. The latter helps to distinguish reliable poses from non-reliable ones, that have been generated during the sampling stage. These two stages are closely interconnected and are performed iteratively. Frequently, predicting the 3D structure of a ligand–receptor complex requires selecting from a set of poses previously generated by docking, where ranking based on the scoring function does not provide a clear indication of the most probable geometries of the investigated complex. Therefore, the introduction of experimental constraints, such as the position of the ligand inside the active site pocket, are crucial to completing this task. Indications about the preferred poses can be obtained by the agreement of selected distances between the predicted ligand-receptor complex, and the corresponding that are estimated using NMR spectroscopy interaction studies or molecular mechanic and molecular dynamic simulation. The accuracy of docking calculations is a compromise between two opposite tasks: an extended sampling of the conformational space of the ligand-receptor system, and an accurate estimation of the intermolecular forces that are required for the estimation of the binding energy. Undoubtedly, the higher the required accuracy, the greater the computational effort and cost, and consequently, the fewer the number of ligands that can be evaluated. As a result, many docking programs adopt a compromise: rough scoring functions, which prioritize speed over precision, are commonly used to guide search algorithms. More accurate scoring functions are then applied for the final ranking of poses. A similar approach can also be adopted in the context of molecular flexibility [160]. Nowadays, the most popular docking programs use a semi-flexible approach, in which the protein is treated as a rigid body, while the ligand is considered flexible and capable of exploring the conformational landscape. Another method is rigid docking, where both the ligand and the protein are treated as rigid bodies, and the interaction is based solely on geometric and spatial complementarity. Finally, the most flexible approach involves a partially or fully flexible protein, while the ligand remains completely flexible.

In this work the docking software Glide [161, 162] , acronym for Grid-based Ligand Docking with Energetics (Schrodinger Inc., LLC, New York, NY) was used.
Glide can sample and score ligand poses through the combination of empirical and force-field terms [161, 162]. Several protocols with different degrees of accuracy can be selected:

- **Standard precision Glide** (SP Glide) is based on a soft scoring function that aims to determine a wide range of ligand binding poses, minimizing all the possible false-positive binders.

- **Extra-Precision Glide** (XP Glide) is based on a hard scoring function, where several penalty factors are introduced to discard those poses that violate the theoretical physicochemical features of protein-ligand interactions [163].
- **High-Throughput Virtual Screening Glide** (HTVS Glide) uses the same scoring function as SP Glide, but it generates a lower number of intermediate conformations by reducing the sampling accuracy and the final torsional refinement.

Typically, while SP and HTVS Glide are used for the virtual screening and hit identification from large compound libraries, XP Glide is more suitable for hit refinement and lead optimization. In the sampling step, the program employs hierarchical filters to sample the conformation, orientation, and position for the ligand within the protein binding site. In this stage, initial ligand conformations are generated by exploring torsional space, and only those with lower torsional energy are selected. Those structures are then positioned within the receptor active site to assess all the possible orientations and potential placements: each pose is evaluated using a precomputed scoring function, which penalizes steric clashes, while recognizing favorable interactions such as hydrogen-bonds, ligand-metal and hydrophobic interactions. Then, the top-scoring poses are minimized with the protein binding pocket, using a molecular mechanism scoring function (OPLSE force field [164]) in combination with a distance-based dielectric model. From this, the three to six lowest poses are further refined through Monte Carlo-based search that samples nearby torsional minima, helping to orient peripheral groups properly and may adjust internal torsional angles as needed. Finally, the minimized poses are rescored using GlideScore (GScore), a scoring function derived from a discretized version of the empirical ChemScore function [165].

$$\begin{aligned}\mathrm{GScore} = {} & 0.05 \cdot \mathrm{vdW} + 0.15 \cdot \mathrm{Coul} + \mathrm{Lipo} + \mathrm{Hbond} + \mathrm{Metal} + \mathrm{BuryP} + \mathrm{RotB} + \mathrm{Site} \\ & + \mathrm{Rewards}\end{aligned}$$

Equation 5 – General GlideScore equation employed during Glide docking.

As reported in **Equation 5**, GlideScore incorporates multiple terms, such as Van der Waals energy (VdW), Colombian energy (Coul), lipophilic (Lipo), hydrogen-bond (Hbond), metal-binding (Metal), and polar interaction (Site) contributions. The function also applies penalties for freezing rotatable bonds (RotB), contributions from buried polar groups (BuryP), as well as penalties for electrostatic mismatches, steric clash and excluded volume (Rewards).

An important drawback of GScore, and more generally, of all scoring functions, is that they do not consider solvation or entropic effects, nor possible perturbative forces that the ligand and the protein can experience upon binding [166-168]. Hence, docking calculations can help to predict the native binding poses, without accurately estimating the corresponding binding affinity. To overcome this limit, a suitable solution is to rescore the final poses determined during the docking calculation using other scoring functions, such as Molecular Mechanics/Generalized Born Surface Area (MM-GBSA) [169, 170].

3.2.2 Molecular Dynamics (MD) Simulations

Molecular Dynamics (MD) simulation is a potent technique in the field of computational chemistry and bioinformatics, which made it possible to explore the structure, function, and dynamics of a biological system. The first MD simulation dates to 1977 [171], when the research group of Karplus performed the simulation of bovine pancreatic trypsin inhibitor (BPTI) in vacuum. Nowadays, thanks to the advent of sophisticated methodology and the development of increasingly powerful computational resources, macromolecules can be simulated in a near-realistic environment by solvating systems in boxes of water and ions.

3.2.3 Force field

In an MD simulation, the motion of a molecule is described by the integration of Newton's equation of motion, considering the masses ($\mathbf{m_i}$) and the position over time $\mathbf{r_i(t)}$ of each atom, the force $\mathbf{F_i(t)}$ acting on them can be determined over time $\mathbf{t}$ (**Equation 6**).

$$\mathbf{F_i(t) = m_i \cdot \frac{\partial^2 r_i(t)}{\partial t^2}}$$

Equation 6 – Newton's differential equation of motion

The integration of **Equation 6** enables monitoring the position the atoms over the time, offering a description from a microscopic point of view, as well as the macroscopic information about the thermodynamic and kinetic properties of the system. Therefore, the simulation generates a trajectory, which is a collection of positions and velocities of all the particles in a simulated time interval. The strength of the MD simulation is that, the microscopic properties contained in this statistical mechanics ensemble can correlated to the macroscopic observables, such as pressure, viscosity, free energy [172].

To solve numerically the Newton equation of motion, the potential energy of the molecules is calculated using the initial coordinates of the system, and an empirical function form, currently known as the 'force field' (FF), that reproduces the covalent and non-covalent forces acting between covalent bonded and non-bonded atoms of the macromolecular system. Upon the selection of their functional form, these empirical potential energy functions require parameterization that are obtained using structural and molecular constraints (optimal bond length, bond angles, dihedral angles, partial charges, and Van der Waals parameters) derived from Quantum Mechanics calculation (QM) and/or experimental data (X-ray, NMR, IR).

Parametrization is usually dependent on a restricted class of compound and/or macromolecules: indeed, FF parameters optimized for proteins can be slightly different from that of nucleic acids, or carbohydrates. To date, several force fields are available [173], which are characterized by a specific set of empirical parameters and potential energy definition.

Both aforementioned FFs calculate the energy using a class I additive potential energy function (**Equation 7**):

$$E_{tot} = E_{bonds} + E_{angles} + E_{dihedrals} + E_{improper} + E_{nonbond} + E_{coulomb}$$

$$E_{bonds} = \sum_{bonds} \frac{K_r}{2}(r - r_0)^2$$

$$E_{angles} = \sum_{angles} \frac{K_\theta}{2}(\theta - \theta_0)^2$$

$$E_{dihedrals} \sum_{dihedrals} K_\chi(1 - cos(n\chi - \delta)$$

$$E_{improper} = \sum_{improper} K_{imp}(\varphi - \varphi_0)^2$$

$$E_{disperive} = \sum_{nonbond} \left(\varepsilon_{ij}\left[\left(\frac{min\, r_{ij}}{r_{ij}}\right)^{12} - \left(\frac{min\, r_{ij}}{r_{ij}}\right)^{6}\right]\right)$$

$$E_{coulomb} = \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}$$

Equation 7 – General equation of a class I additive potential energy that comprises bonded interaction (E_{bonds}, E_{angles}, $E_{dihedrals}$, $E_{improper}$) and non-bonded interaction ($E_{dispersive}$, $E_{coulomb}$) terms. Function of each bonded and nonbonded term in a class I additive potential energy function re reported. $K_r, K_\theta, K_\chi, K_{imp}$ correspond to the bond force constants, the valence angle constants, the dihedral force constants, and the improper force constants, respectively. r_0, θ_0, φ_0 are the equilibrium distance, the equilibrium angle, and the equilibrium improper angle, respectively. n and δ are the multiplicity and the phase angle. $\varepsilon_{ij}, min\, r_{ij}, r_{ij}, q_i, q_j$ and ϵ_0 are the Lennard-Jones well-depth, the minimum interaction radius between atom i and atom j, the distance between atoms i and j, the partial charges of atom i and atom j, and the dielectric constant.

The basic assumption is that the total potential energy is the sum of independent terms, as shown in **Equation 7**, representing the forces that act between atoms through the internal degree of freedom that characterize a bio-macromolecular system: bond distances, bond angles, dihedral angles, and distances between non-bonded atoms.

The bonds stretching (r), angles bending (θ) and improper dihedral angle distortion (φ) are treated harmonically; instead, the torsion angles (χ) are described by a sinusoidal term. Atom-atom repulsion and dispersion interactions are represented by the 6-12 Lennard-Jones term, while the electrostatics is derived from the Coulombic term. Two other types of force fields can also be considered. Class II force fields account not only for bond stretching and angle bending but also for cross-coupling interactions such as bond–bond and bond–angle terms, providing a more accurate description of molecular flexibility (e.g. Polymer Consistent Force Field – PCFF [174, 175]). Class III force fields further extend this approach by including polarization effects, allowing atomic charges to respond dynamically to the local electrostatic environment (e.g. AMOEBA force field [176]), leading to a more refined description of the molecular electrostatic potential field. In contrast, Class I and Class II force fields typically assume that their electrostatic potential is generated by fixed point charges located on the atom. Since heparan sulfate is the subject of this project, selecting the appropriate force field, specifically developed and tailored for carbohydrates, was essential. In 2007, a new derivation of the GLYCAM force field, GLYCAM_06 [177], addressed limitations observed in its first version, particularly related to the low reproducibility of sugar diffusion rates and the radial distribution function. The structure of each monosaccharide was acquired through neutron diffraction experiments, focusing on the conformation and glycosidic linkage. Furthermore, using quantum mechanical calculations, bond and valence angle deformation force constants, dihedral angle rotational barriers, and electrostatic properties were added as new parameters [177], since they are inaccessible experimentally. The calculation of partial atomic charges was achieved through the fitting of QM-derived molecular electrostatic potentials by assigning partial charges to all atoms in a molecule, except for aliphatic hydrogen atoms, which are excluded from the fitting process.

3.2.4 Solvation box

Currently, molecular dynamics (MD) simulations are typically conducted in explicit solvent, namely the solvent is normally described by a fixed number of water molecules, which represent the nearest neighboring environment that surround the ligand-receptor macromolecular complex. To preserve the density of the system, the solute and the solvent are enclosed in a box (the simulation box) that can assume different geometries: cubic, orthogonal, or triclinic, that is characterized by three edges and three angles between edges. Generally, living systems are usually represented in water solutions. Different water models are available, including TIP3P, TIP4P, TIP5P, OPC, and SPC [178-181]. Each of these models represents water molecules using a specific number of interaction points and geometry. While SPC and TIP3P models use three interaction points, corresponding to the actual atoms of the water molecule, TIP4P and TIP5P introduce one or two additional points, respectively, by incorporating dummy atoms to better describe the electrostatic charge of the lone pair electrons that surround the oxygen atom in its typical sp^3 geometry. Although increasing the number of interaction points improves the accuracy of electrostatic representation, it also significantly raises the computational cost, especially since water molecules constitute over 80% of the system's particles and thus account for most of the simulation workload. For this reason, the TIP3P model is frequently selected, as it offers a good balance between computational efficiency and physical realism, with its rigid geometry closely matching the actual water molecule one.

3.2.5 Periodic Boundary Condition (PBC) and Particle Mesh Ewald (PME)

One of the most time-consuming tasks in an MD simulation is the calculation of non-bond interactions. A major problem related to computing these interactions is the high computational cost, since for a system of N atoms, a number of N^2 operations are required for the calculation of all pairwise two body potentials $\boldsymbol{V}(\boldsymbol{r}_i, \boldsymbol{r}_j)$. To overcome this problem, a feasible solution is a cutoff truncation, where interaction over a certain distance cutoff is considered equal to zero or smoothly reduced to zero. In this case the number of pairwise in interaction grows approximately as N.

Currently, the most widely used method for computing electrostatic long-range interaction in a MD simulation is the Particle Mesh Ewald (PME) sum. The central idea is to decompose the electrostatic potential into short-distance and long-distance terms. The former, which decays rapidly with the distance, is computed directly as a cutoff-based summation of pairwise Coulomb component. In contrast, the latter is evaluated in the reciprocal space, by using the Fast Fourier Transform (FFT). To achieve this, the discrete point charges are first interpolated onto a regular three-dimensional mesh to obtain a smooth charge density distribution. The long-range electrostatic potential is then computed as a convolution between this mesh-based charge density and the lattice Green's function. This approach reduces the computational complexity from N^2 to $\mathrm{N} \cdot \log \mathrm{N}$, thereby significantly improving efficiency for large systems without introducing real-space cutoffs for the long-range terms [182].

The application of the PME method relies on the assumption of the existence of an infinitely periodic system, as the FFT of long-range electrostatic requires periodicity. Consequently, PME is typically employed together with Periodic Boundary Conditions (PBC): it consists of replicating the simulation in all the spatial direction, to mimic the bulk behavior and minimizing any edge effect (**Figure 16**). To avoid atoms interacting with multiple images of the same or other atoms (introducing non-physical artifacts), the minimum image convention is applied. This ensures that the inter-atomic distance is computed using the shortest possible vector between a pair of atoms and their periodic images.

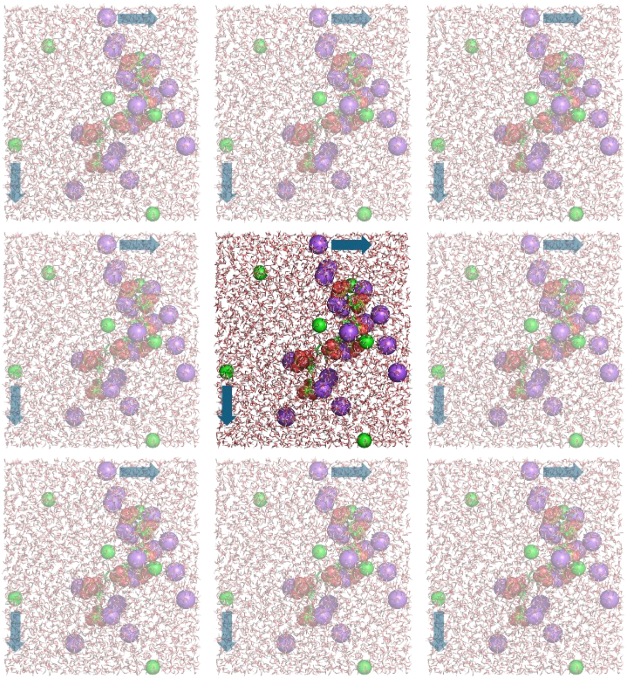


Figure 16 – Schematic representation of periodic boundary conditions (PBC) in two dimensions. When a particle crosses the boundary of the simulation box, a corresponding image particle enters from the opposite side.

3.2.6 Thermostat and barostat

As described in section 3.1.2, the MD simulations are based on the numerical integration of the Newton differential equation (**Equation 6**), which implies the conservation of energy. The correlation between the microscopic description, which involves the coordinates and velocities of the atoms of the system, and the macroscopic, that focus on total energy, temperature, pressure, entropy, free energy, etc. is given by the statistical thermodynamic interpretation. According to this scheme we ideally focus on a representative ensemble in which independent systems are simulated, and the laws of statistical mechanics can be successfully applied correlating the microscopic and the macroscopic variables. From a statistical point of view, the system can be described in a microcanonical ensemble, where the number of particles P, the volume V and the total energy are constant, which typically does not present a common experimental scenario. A more realistic system requires that the MD simulation can be run fixing the average temperature of the simulation box, and the average pressure that is applied on its hedges. Therefore, a thermostat or a barostat algorithm can be coupled to the integrated Newton equations. The result is that the former allows to simulate according to the canonical ensemble (NVT condition, characterized by the number of particles N, the volume V, and the temperature T constant). The application of both thermostat and barostat allows the simulation to sample the isothermal-isobaric ensemble, in which both the average pressure (P) and average temperature (T) are fixed (NPT ensemble.

A briefly description of Langevin thermostat [183] and Berendsen barostat [184], employed in this work, is reported.

- **Langevin thermostat**

The Langevin thermostat [183] maintains the average temperature close to the selected value, controlling the kinetic energy according to **Equation 8**.

$$\mathrm{m_i}\, \partial^2 \mathrm{r_i} = -\nabla_{\mathrm{i}} \mathrm{U} - \mathrm{m_i} \Gamma\, \partial \mathrm{r_i} + \mathrm{W_i(t)}$$

Equation 8 – Modified equation of motion used in the Langevin thermostat. m_i is the mass of atom $\boldsymbol{i}$, $\boldsymbol{r_i}$ is the position of atom $\boldsymbol{i}$ at time $\boldsymbol{t}$, $-\nabla_i U$ is the conservative force applied on atom $\boldsymbol{i}$, Γ is the friction coefficient and $W_i(t)$ is the random force applied at time t.

In detail, at each time step a random force $\mathrm{W_i(t)}$ is applied on the atom $\boldsymbol{i}$ and its velocity is reduced through the application of a friction force of coefficient Γ. This modification mimics the physical situation where a solute particle experiences friction (from being slowed by smaller particles, i.e., the particle of solvent), random thermal forces (from collisions with those particles), as well as the conservative forces that are originated from the force field. In this case the friction forces are applied to decrease the particle velocities to maintain the average kinetic energy and the average temperature of the system.

- **Berendsen barostat**

The Berendsen barostat [184], to maintain the average pressure (P) requires to scale the volume (V) of the simulation box. Since pressure and volume are inversely proportional, the size reduction of the box leads to atoms being more compressed, experiencing an increase of the average pressure. In MD simulations, this result can be obtained rescaling atomic coordinates by a scaling factor λ (**Equation 9**):

$$\mathrm{r}_{\mathrm{i}}^{\mathrm{t}} = \mathrm{r}_{\mathrm{i}}^{\mathrm{t}_0} \lambda^{1/3} \rightarrow \frac{\delta \mathrm{P(t)}}{\delta \mathrm{t}} = \frac{1}{\tau} (\mathrm{P}_{\mathrm{box}} - \mathrm{P(t)})$$

Equation 9 –Atomic coordinate rescaling adopted in Berendsen barostat λ is the scaling factor, while $\tau(ps)$ is the pressure relaxation time.

The algorithm employed in the Berendsen barostat allows a quick equilibration of the simulation box; however, since volume fluctuations are suppressed, it does not properly sample the NPT ensemble. Therefore, it is suitable during the equilibration phase but should be avoided during production [185, 186], during which it is preferable to use the Monte Carlo barostat [187].

3.2.7 Simulation of STD$_0$ signals

In recent decades, several approaches have been developed to exploit binding epitope maps for scoring 3D protein-ligand complexes. Many rely on distance-based algorithms [188, 189] , which do not account for dipole–dipole relaxation and internal dynamics. Others are based on the full relaxation matrix calculation [190, 191], that offer greater accuracy but are computationally demanding, limiting their use for the analysis of MD simulations. To overcome this limitation, the research group led by Jesús Angulo developed the "Reduced Relaxation Matrix" (RedMat) approach [131], a novel method that allows to predict the theoretical initial slopes of the ligand protons for a 3D model of the ligand-receptor complex (generated, for instance, from docking calculations or MD simulations. This approach, implemented as a web server freely accessible to academic users (http://redmat.iiq.us-csic.es/login), is based on the reduced relaxation matrix formalism, which relies on the direct calculation of theoretical STD initial slopes ($\mathrm{STD}_0^{\mathrm{theo}}$) rather than the full theoretical build-up curves, thus significantly reducing the computational cost. RedMat is based on the theory behind the previously known approach: CORCEMA [190], which describes the magnetization transfer in a two-state exchange system (such as a ligand-receptor complex) as reported in **Equation 10**.

$$\mathrm{I}(\mathrm{t}_{\mathrm{sat}}) = \mathbf{I_0} + \left[1 - \mathrm{e}^{\{-(\mathbf{R}+\mathbf{K})\cdot \mathrm{t}_{\mathrm{sat}}\}}\right] \cdot (\mathbf{R}+\mathbf{K})^{-1} \cdot \mathbf{Q}$$

Equation 10 – The complete relaxation matrix for a two-site exchange system employed in CORCEMA. Boldface denotes the matrix character of the corresponding parameter in the equation.

Where **R** is the generalized relaxation matrix of the ligand, **K** is the generalized exchange kinetics matrix, and **Q** is the matrix containing information about the spin system with the species in bound and in free state. By combining **Equation 8**, **9**, and **10**, the theoretical initial slope $\mathrm{STD}_0^{\mathrm{theo}}$ becomes proportional to the reduced relaxation matrix **Q** (**Equation 11**).

$$\mathrm{STD}_0^{\mathrm{theo}} \propto (\mathbf{R}+\mathbf{K}) \cdot (\mathbf{R}+\mathbf{K})^{-1} \cdot \mathbf{Q} \Rightarrow \mathrm{STD}_0^{\mathrm{theo}} \propto \mathbf{Q}$$

Equation 11 – Combination of complete relaxation matrix employed in CORCEMA and STD_0 defined in equation 7.

The latter incorporates cross-relaxation matrix weighted by the thermal intensities of protein protons in the free ($\mathbf{I_{L20}}$) and bound state ($\mathbf{I_{L2'0}}$) (**Equation 12**):

$$\mathbf{Q} = \begin{bmatrix} 0 \\ \mathbf{R_{P1P2}} \cdot \mathbf{I_{L20}} \\ \mathbf{R_{L'P2'}} \cdot \mathbf{I_{L2'0}} \\ \mathbf{R_{P1'P2'}} \cdot \mathbf{I_{L2'0}} \end{bmatrix}$$

Equation 12 – Reduced relaxation matrix **Q**.

where $\mathbf{R_{P1P2}}$ is the cross-relaxation matrix of the unsaturated (P1) and directly saturated (P2) protein protons in the free state, $\mathbf{R_{L'P2'}}$ is the cross-relaxation matrix of ligand protons (L′) and saturated protein protons in the bound state (P2′), $\mathrm{R_{P1'P2'}}$ is the cross-correlation of the unsaturated (P1′) and directly saturated (P2′) protein protons in the bound state. The predicted $\mathrm{STD_0^{theo}}$ selected from the modelled ligand-receptor complex can be compared with the corresponding experimental $\mathrm{STD_0^{exp}}$ values by using the R-NOE factor, i.e. a metric to evaluate the goodness of the agreement between $\mathrm{STD_0^{theo}}$ and $\mathrm{STD_0^{exp}}$ (**Equation 13**), where W_k is a weighting factor which allows to weigh differently the STD intensities.

.

$$R - NOE\ factor = \sqrt{\frac{\left(\sum_k W_k\ (STD_0^{exp} - STD_0^{theo}\right)^2}{\sum_k \left(STD_0^{exp}\right)^2}}$$

Equation 13 – R-NOE factor equation.STD_0^{theo} and STD_0^{exp} represent the simulated and the experimental initial slope of the given proton, respectively.

As reported by the authors, R-NOE factor around or lower 0.3 indicates a good agreement between the set simulated and the experimental $\mathrm{STD_0}$ intensities.

Chapter 4. Results

4.1 Unveiling the conformational landscapes of oligosaccharide mimetics of heparan sulfate in unbound state.

The extraction and isolation of HS, suitable for experimental structural characterization, presents considerable challenges. Cellular HS is a highly heterogenous polysaccharide, varying in chain length, sulfation pattern, and domain organization, making it difficult to obtain homogeneous, structurally resolved samples (see section 1.2 for more details). To overcome these limitations without losing biological relevance, we employ two synthetic hexasaccharides that mimic the key structural and electrostatic features of HS. The first, referred to here as hexa, is a well-defined oligosaccharide designed to reproduce the repeating disaccharide units and sulfation motifs of HS [192] (**Figure 17** and **Figure 21a**). Hence, it consists of an uronic acid, namely α-IdoA2S/ β-GlcA bound to a N-, 6-O disulfate glucosamine through 1→4 glycosidic bond, generating the corresponding hexasaccharide with formula MeO-(4)-α-L-IdoA2S(1→4)-α-D-GlcNS,6S(1→4)-β-DGlcA(1→4)-α-D-GlcNS,6S(1→4)-α-L-IdoA2S(1→4)-α-D-GlcNS,6S-(1)OMe (where IdoA2S, GlcNS,6S, corresponds to 2-O-sulfo- α -L- iduronic acid and N-,6-O disulfate α-D-glucosamine, respectively). The second ligand is a fragment of Pentosan polysulfate sodium [193] (PPS), that is a plant-derived sulfated polysaccharide obtained by chemically sulfonating β-(1→4)-linked xylan extracted for beechwood (**Figure 21b**). PPS is the active pharmaceutical ingredient in Elmiron®, and FDA-approved drug used in the treatment of chronic interstitial cystitis [194]. It was selected since it provides a high density of sulfate groups and a spatial charge arrangement analogous to that found in native HS chain. Beyond its established clinical application, PPS has recently gained attention for its promising inhibitory effects on SARS-CoV-2 replication [130]. Both hexa and PPS carry a remarkable negative charge, arising from the presence of multiple sulfate and carboxylate groups (**Figure 17**).

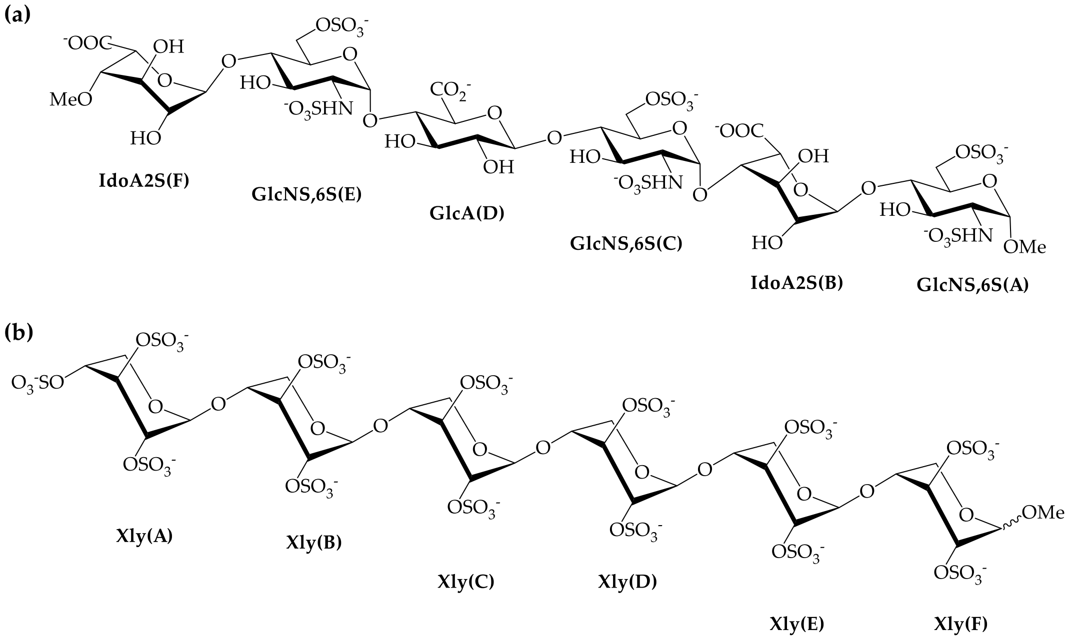


Figure 17 – 2D structure of hexa (panel a) and PPS (panel b). The monosaccharide units are denoted according to their residue names as follows: IdoA2S for 2–O–sulfated iduronic acid; GlcNS,6S for 2-N–,6–O–disulfated glucosamine; GlcA for glucuronic acid; Xyl for xylose. Capital letters are labelled alphabetically from F to A, corresponding to the non–reducing to reducing end, respectively.

This high charge density imposes a certain degree of structural rigidity, forcing each oligosaccharide to adopt a linear and stiff conformation.

To investigate the structural and conformational properties of these ligands in their free state, MD simulations were performed in explicit solvent (three replicas of 200 ns each, under NVT conditions at T=300K), using the GLYCAM_06j-1 force field and the TIP3P water model (see section 5.2 for further details). Throughout the trajectories, the glycosidic torsion angles ϕ_i/ψ_i and the sugar ring puckering were monitored to characterize both the monosaccharide ring geometries and the inter-glycosidic conformation.

4.1.1 Monosaccharide ring conformation

The ring puckering of each monosaccharide ring was characterized using the Cremer-Pople puckering parameters [195], expressed in terms of the corresponding polar coordinates **Q**, **θ** and **ϕ**. These parameters provide a comprehensive description of the three-dimensional distortion of the pyranose ring from the planarity (Q parameter – **Figure 18**), enabling the differentiation between various conformational states such as chair, boat and skew-boat forms (θ and ϕ angles – **Figure 18**).

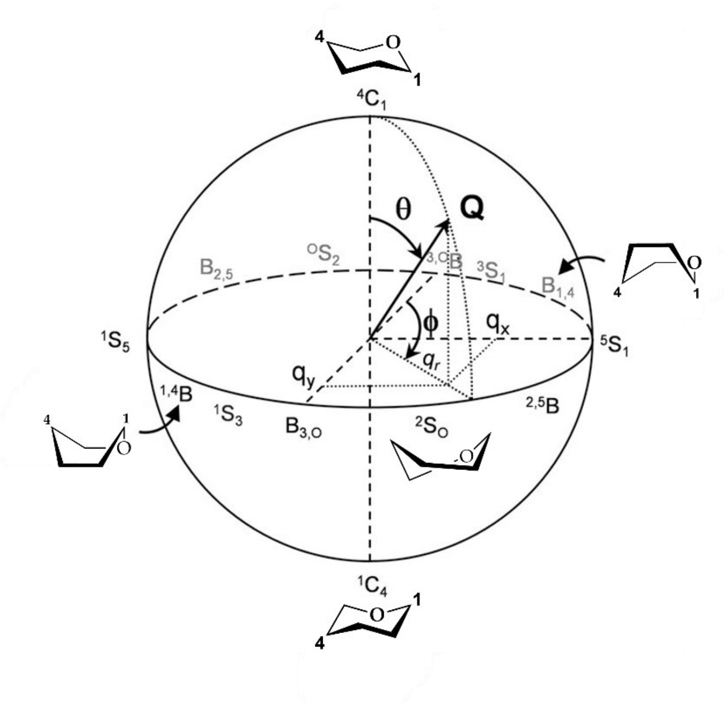


Figure 18 – Cremer-Pople puckering parameters for six-membered rings. The spherical coordinate system (Q, θ, φ) provides a complete description of ring deformation, where Q represents the total puckering amplitude (ranging from 0 for planar rings to ~0.6 Å for typical chair conformations), θ is the polar angle (0° ≤ θ ≤ 180°) defining the type of puckering, and φ is the azimuthal angle (0° ≤ φ ≤ 360°) describing the orientation of deformation. Chair conformations ($^4C_1$, $^1C_4$) correspond to θ = 0° and 180° respectively, with maximum Q values. Boat conformations ($B_{2,5}$, $B_{3,O}$, etc.) and skew conformations ($^1S_3$, $^2S_0$, etc.) occur at θ = 90° with diverse φ values. *Image readapted from C. Rovira. Chem. Sci. 6. https://doi.org/10.1039/C4SC02240H*

The distribution of the puckering angle θ and ϕ obtained from the molecular dynamic trajectories were projected on a bidimensional space generating the Stoddard diagrams (projection of the sphere from the north pole), which facilitated the visualization and classification of the sampled ring conformation according to the Cremer-Pople Parameter (**Figure 19**). In these diagrams, the color code identifies the polar angle adopted by each ring conformation over the simulation: violet for the $^4C_1$ conformation (θ = 0°), yellow for the $^1C_4$ (θ = 180°) conformation, and red for boat (B) and skew-boat (S) conformations (θ = 90°).

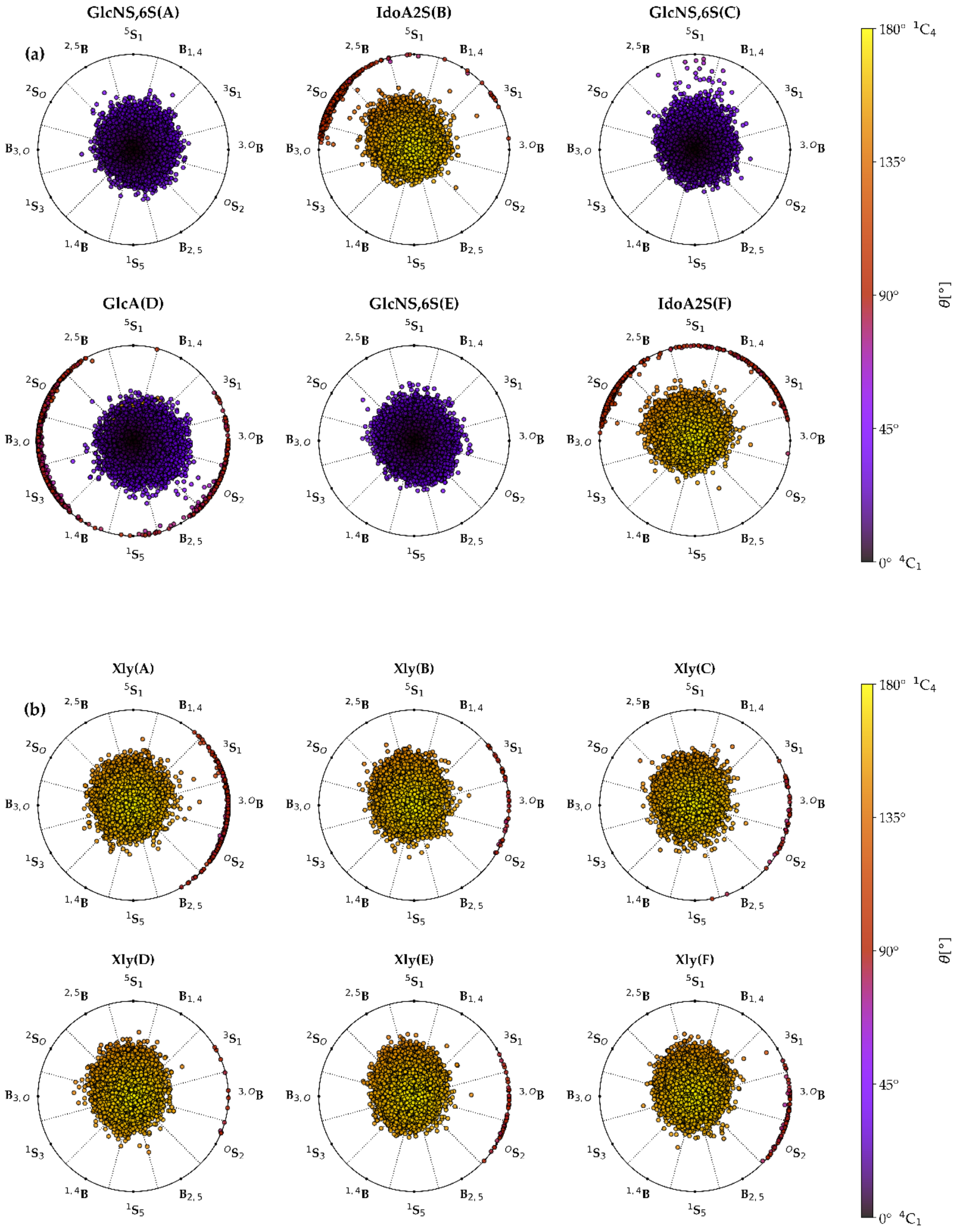


Figure 19 – Stoddart diagram of the ring puckering population of hexa (panel a) and PPS (panel b).

In solution, the glucosamine (A), (C), and (E) of hexa exhibit exclusively the $^4C_1$ chair conformation (**Figure 19a**), while the GlcA (D) populates preferentially the $^4C_1$ (94%) with a minor contribution (6%) of the others(**Table A. 2a**). Instead, the IdoA2S (B) and (F) exhibit greater flexibility, displaying an equilibrium between the $^2S_0$ and $^1C_4$ conformation with a ratio of 98:2 [16, 196].

In contrast, for PPS, all the rings adopt predominantly the $^1C_4$ conformation, with minor population (< 1%) of the $^{3,O}B$, $^3S_1$ ^2S_O conformers (**Table A. 2b**). This result is consistent with previous experimental findings [197-199]. In fact, the presence of bulky, negatively charged sulfate groups induces the xylose residues to assume predominantly the chair $^1C_4$ differently from the un-sulfated Xyl, that populate the $^4C_1$. This conformational inversion is generated by the electrostatic repulsion between the sulfate groups that in PPS occupy the 2-O and 3-O positions on the Xyl residue.

4.1.2 Inter-glycosidic bond

The conformational landscape of each inter-glycosidic linkage was analyzed by calculating the corresponding glycosidic torsion angles, ϕ_i/ψ_i, defined according to the non-IUPAC [200] definitions. Specifically, the ϕ_i and ψ_i are defined by the ***$H1_n$ – $C1_n$ – $O4_n$ – $C4_{(n+1)}$ and $C1_n$ – $O4_n$ – $C4_{(n+1)}$ – $H4_{(n+1)}$*** dihedral angles. Each ϕ_i/ψ_i pair was numbered sequentially, starting from the non-reducing end of the oligosaccharide chain (IdoA2S(F) and Xyl-(A) in hexa and PPS, respectively). The distribution of these dihedral angles was represented by Ramachandran plots, which provide a comprehensive visualization of the conformational space sampled (**Figure 20**).

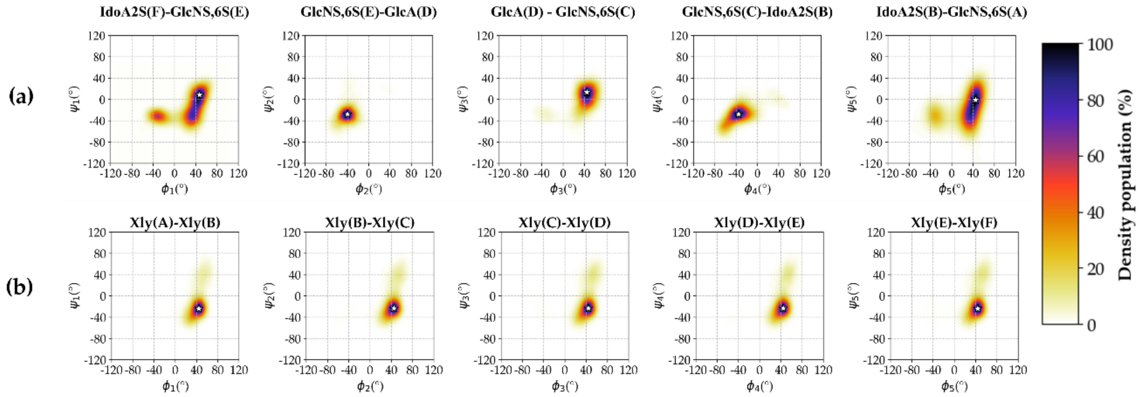


Figure 20 – Ramachandran-type plots of glycosidic angles ϕ_i / ψ_i of unbound hexa (panel a) and PPS (panel b) molecules obtained from MD simulations.

The shape and the extent of the density maps reflect the local flexibility of each glycosidic bond: narrow and well-defined clusters indicate conformational rigidity, whereas broader distribution corresponds to higher torsional mobility. The color scale within the plots represents the population density of each conformational state, thereby highlighting the most frequently sampled ϕ_i/ψ_i states (white star in each plot). The mean values of these torsional angles are summarized in

Table A. 3.

The analysis of the Ramachandran plots (**Figure 20**) indicated that both hexa and PPS are rigid molecules and populate well-defined regions. In the hexa structure (**Figure 20a**), the central residues [GlcNS,6S(E) to IdoA2SB)] predominantly populate a single conformational state closing resembling the input structure [192]
(

Table A. 3). In contrast, the terminal residues, IdoA2S(F) and GlcNS,6S(A), display greater flexibility, as evidenced by the presence of two populated ϕ_i/ψ_i states region, one of them more populated (**Figure 20a**). This increased flexibility arises because these are terminal residues and have more conformational flexibility. Conversely PPS exhibits only a single ϕ_i/ψ_i population across all the residue pairs, indicating an overall stiffer conformation in comparison to hexa (**Figure 20b** and

Table A. 3). This result is consistent with the structural and conformational properties of the ligands, in which the strong electrostatic repulsion between the negatively charged groups (sulfated and carboxylate groups), along with the corresponding steric hindrance, is dominant and forces the molecule to preserve its linear, relatively stiff conformation. It is intriguing to note the charged groups display a diverse spatial arrangement (**Figure 21**). In hexa, the sulfate and carboxyl groups are positioned such that they are oriented in opposite directions along the molecular axis, creating two broad charged faces. Each of these faces is potentially capable of engaging the positively charged patches that are exposed on site I of the RBD of SARS-CoV-2. In contrast, PPS displays a radial distribution of sulfate groups arranged symmetrically around its molecular axis, resulting in a uniformly charged cylindrical surface. This geometry facilitates non-directional electrostatic interactions with basic residues, enabling PPS to make consistent contacts along its entire length with complementary positive regions on the RBD surface.

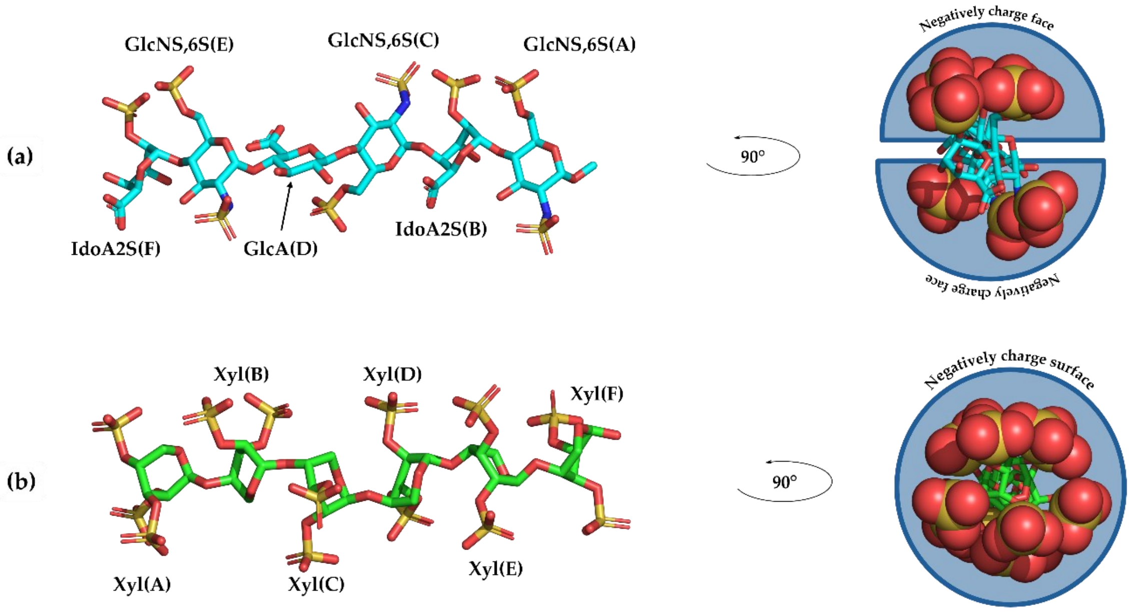


Figure 21 – 3D representative structures of the hexa (panel a) and PPS (panel b). Both ligands are depicted using the stick representation (cyan and green for carbon, red for oxygen, blue for nitrogen and yellow for sulfur). In the view of the molecular axis, the sulfate groups are represented as van der Waals spheres to emphasize their distinct orientations and spatial distributions. For hexa, monosaccharide units are labelled with both their residue name, IdoA2S (iduronic acid 2-O-sulfated), GlcNS,6S (glucosamine N-sulfated, 6-O-sulfated), and GlcA (glucuronic acid, and a capital

letter from F (non-reducing end) to A (reducing end). Instead for PPS the xylose unit (Xyl) are numbered from the non-reducing to the reducing end.

4.2 Step 1: Docking calculations

The results of this section have been published in a scientific paper: Froese, J., ***Mandalari, M.****, Civera, M. et al. Evolution of SARS-CoV-2 spike trimers towards optimized heparan sulfate cross-linking and inter-chain mobility. Sci. Rep. 14, 32174 (2024). https://doi.org/10.1038/s41598-024-84276-5.*

The interactions between the heparan sulfate mimetics, hexa and PPS, and the RBD region of the SARS-CoV-2 spike protein from the wild-type, Delta, and Omicron variants were investigated. Ligand and protein preparations followed the protocols detailed in Sections 5.1.1 and 5.1.2. Each protein model was glycosylated at N343 using the FA2G2 glycan structure obtained from the CHARMM-GUI COVID-19 Protein Archive, which corresponds to a fully glycosylated, all-atom model of the homotrimeric SARS-CoV-2 spike protein in the open conformation (PDB ID: 6VSB). The FA2G2 glycan was covalently attached to each S1-RBD monomer via a glycosidic linkage between the reducing end of GlcNAc1 and the ND2 atom of residue N343. As described in section 2.6, three distinct GAG binding sites were identified on the RBD domain [151] (**Figure 14**). To encompass all these binding regions, which are distributed across a broad area of the RBD surface, the dimensions of the docking grid box were defined to include the entire binding interface. Further details on the grid setup are provided in Section 5.1.3.

Flexible and rigid docking simulations were conducted using the docking precision available on Glide: standard precision (SP) and extra precision (XP).

The rigid docking protocol did not yield any meaningful binding poses. Despite multiple parameter optimization aimed at relaxing the energetic filters, no poses successfully passed the initial scoring phase.

In flexible docking calculations, where both ring puckering and glycosidic bond rotations were sampled, the XP and SP protocols produced notably different outcomes. In XP mode, extensive distortions of the sugar ring geometries were observed, resulting in a loss of native chair conformations and the formation of boat-like or severely distorted ring structures. Conversely, SP flexible docking preserved the overall ring conformations, but induced substantial deviations in the torsion angles around the glycosidic linkages. These deviations exceeded ±30° in relation to both the input geometries and those observed from the predominant conformational state identified in the explicit-solvent MD simulation (see section 4.1.2 - **Figure 20**). For the PPS ligand, these distortions forced the overall molecular structure into an unrealistic U-shaped conformation.

To prevent both ring deformation and glycosidic bond distortion, a semi-flexible docking strategy was employed. In this approach, ring conformations were kept rigid and torsional constraints were applied to all glycosidic linkages. These constraints were defined using 4-atom SMARTS patterns.

Ten docking poses were generated for each ligand and analyzed using Maestro GUI (Schrödinger Release 2023-1). These poses where then rescored using the molecular mechanics generalized Born surface area (MM-GBSA) method [170], implemented on Prime [201, 202] to estimate the binding free energy and refine the ligand-receptor complex ranking.

Interestingly, despite differences in the relative position and orientation of the hexasaccharides on the protein surface, the docking consistently identifies the site I as the primary binding site. This suggests that the local distribution of positive charges in site I (principally arginine and lysine) are the dominant determinants of ligand binding (**Figure 22**).

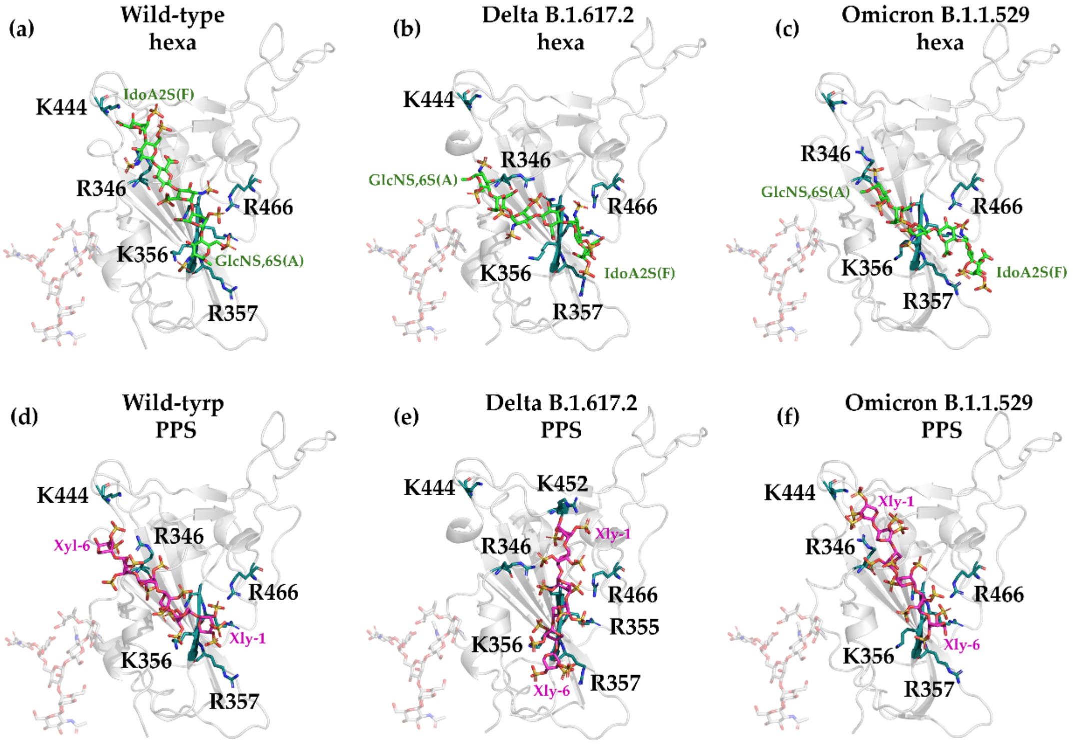


Figure 22 – Top-ranked docking poses of hexa (panels a–c) and PPS (panels d–f) bound to the SARS-CoV-2 RBD from the WT (PDB: 6M0J), Delta (PDB: 7V8B; B.1.617.2), and Omicron (PDB: 7WBP; B.1.1.529) variants. The RBD is shown as a grey cartoon, while ligand atoms are colored as follows: carbon (green for hexa and magenta for PPS, respectively), oxygen (red), nitrogen (blue), and sulfur (yellow). Key binding-site residues of site I, namely R346, N354, R355, K356, R357, K444, and R466—are highlighted using the same atom-based color scheme with deep-teal carbons. For the Delta variant, the L452R mutation in the PPS docking pose is similarly depicted.

In the wild-type RBD, a single docking pose of hexa was obtained with IdoA2S(F) oriented toward R346, while GlcNS,6S(A) interacting with K356 (**Figure 22a**). In Delta (B.1.617.2) RBD, docking calculation predicted an inverted head–tail orientation of hexa relative to the WT model. This reversal brought site I residues into contact with a different monosaccharide arrangement: GlcNS,6S now interacting with R346, while IdoA2S(F) is pointing towards R357 (**Figure 22b**). In all the poses found, the mutated residue L452R do not contribute to the ligand binding. While the total number of contacts did not differ significantly between the WT and Delta – hexa complexes (**Table 2a**), the Delta model displayed a higher number of salt bridges and van der Waals interactions. These additional favorable contacts correlate with a modest improvement in the predicted binding affinity: indeed, the estimated free energy of binding changed from −18 to −26 kcal/mol for the wild-type and Delta RBDs, respectively (**Table 2a**). The increased stability in the latter interaction may reflect subtle conformational differences in the Delta RBD that better accommodate the HS oligosaccharide.

The Omicron (B.1.1.529) RBD exhibited a binding pattern nearly identical to that of the Delta variant, preserving the sequence of the contact between the monosaccharide and site I residue of RBD. (**Figure 22c**). Specifically, in the Omicron RBD – hexa complexes, the ligand binds to site I in two opposite orientations, enabling both of its surface to interact with the surface of the RBD. In first binding mode, which resembles the orientation observed in Delta RBD docking calculation, the GlcNS,6S(A) is oriented towards the hACE2 binding site and proximal to R346, while the IdoA2S(F) is interacting with R357. Conversely, when hexa is docked in the reverse orientation, residue A establish interaction with R355 and R466, where residue F engages with K444, closely recapitulating the docking configuration observed in the wild-type – hexa complex. The interaction with Omicron was markedly stronger, with a greater number of favorable contacts and a significantly improved predicted binding free energy compared to both WT and Delta (**Figure 23** and **Table 2a**).

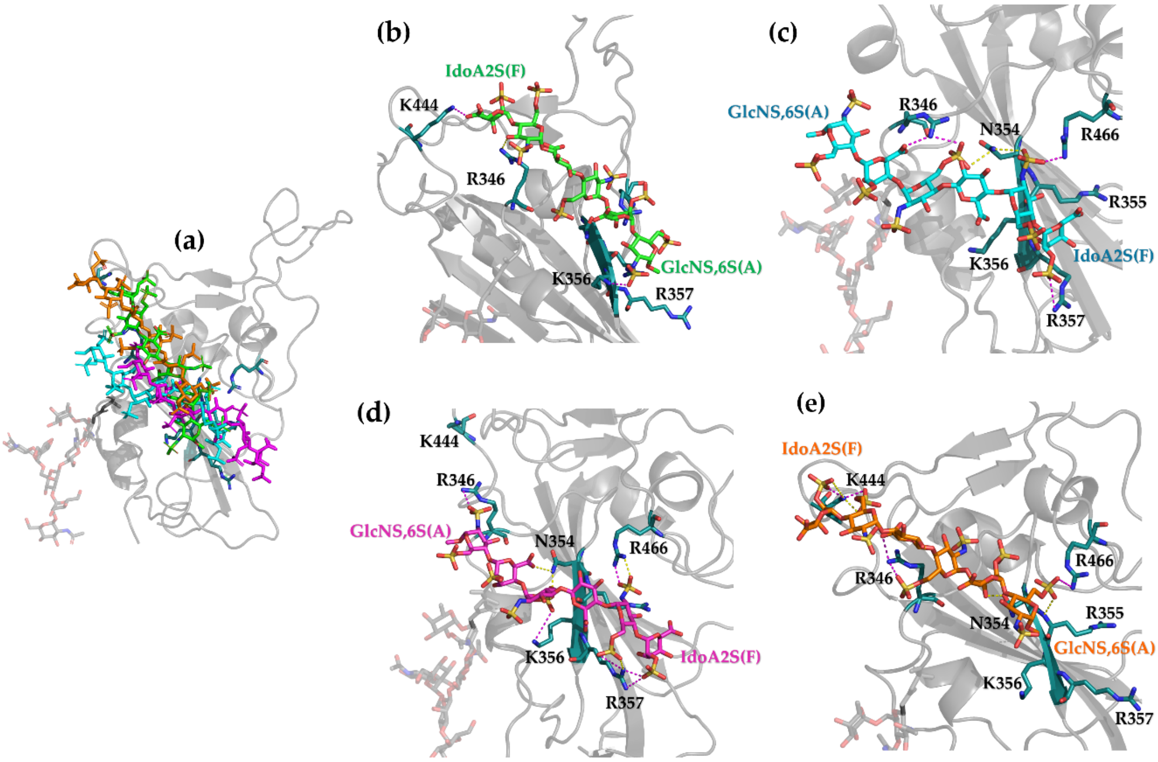


Figure 23 – Superimposition of hexa docked onto the RBD domain of the wild-type (green), Delta (cyan), and Omicron (magenta and orange) SARS-CoV-2 spike variants (panel a). Enlarged view of each complex are shown: Wild type (panel b), Delta (panel b), and Omicron (panel d and e). The RBD region is illustrated as grey carton, while hexa is depicted in stick representation (oxygen in red, nitrogen in blue, and sulfur in yellow). Key interacting residues are shown as sticks with matching color codes (carbon in teal, oxygen in red, nitrogen in blue, and sulfur in yellow). The salt bridges and hydrogen bonds are indicated by magenta and yellow dashed lines, respectively.

Table 2 – Average docking score and MM-GBSA rescoring values, salt bridges, and H-bonds of the ten poses obtained for SARS-CoV-2 RBD variants and hexa (a) and PPS oligosaccharide (b). Regarding wild-type – hexa docking calculation only one pose was calculated. *(#) Only one pose was calculated*

		Docking score (kcal/mol)	**Free binding energy (kcal/mol)**	**Salt Bridges**	**H-bonds**
(a)	**wild-type**	-1.5(#)	-18(#)	2	2
	Delta	-2.4±0.9	-26±9	5	1
	Omicron	-4.4±0.5	-31±5	6	5
(b)	**wild-type**	-4.0±0.3	-24±3	6	6
	Delta	-4.5±0.3	-27±5	8	6
	Omicron	-4.7±1	-44±8	8	7

As previously mentioned, the PPS oligosaccharide is characterized by a radial distribution of negatively charged sulfate groups along its molecular axis (**Figure 21b**). This structural feature enables its sulfate moieties to engage positively charged of site I in RBD surface with an epitope binding potentially different from that of hexa oligosaccharide, which structure and composition resemble heparin and heparan sulfate. When PPS is docked on the wild-type RBD, the docking poses are characterized by the ligand in close proximity to site I (**Figure 22d and Figure 24a**). Interestingly, PPS exhibited a higher number of salt bridges and overall contacts than hexa (**Table 2a-b and Figure 24b**). The charges arrangement appears particularly well-suited for interacting with the basic channel formed along the RBD surface (**Figure 21b**). Notably, the number of PPS–RBD contacts increased progressively across the WT, Delta, and Omicron variant models. This trend was mirrored by a corresponding improvement in predicted binding affinity (**Table 2b** and **Figure 24**). In the Delta and Omicron complexes, the best-ranked docking poses revealed that all residues of site I residues contributed to the engagement of the PPS hexasaccharide, where almost all xylose units are involved (**Figure 24b-c**). Additionally, in case of Delta variant, the mutated residue R452 formed an extra contact with Xyl-1 (**Figure 22e** and **Figure 24c**). The introduction of this contact point may partly explain the higher binding energies observed in Delta, reinforcing the notion that subtle structural changes in the RBD can enhance the accessibility and complementarity of the glycan-binding interface. Instead, the enhanced binding propensity observed for Omicron RBD suggests that this variant presents a more optimally arranged and electrostatically favorable binding platform for the oligosaccharide.

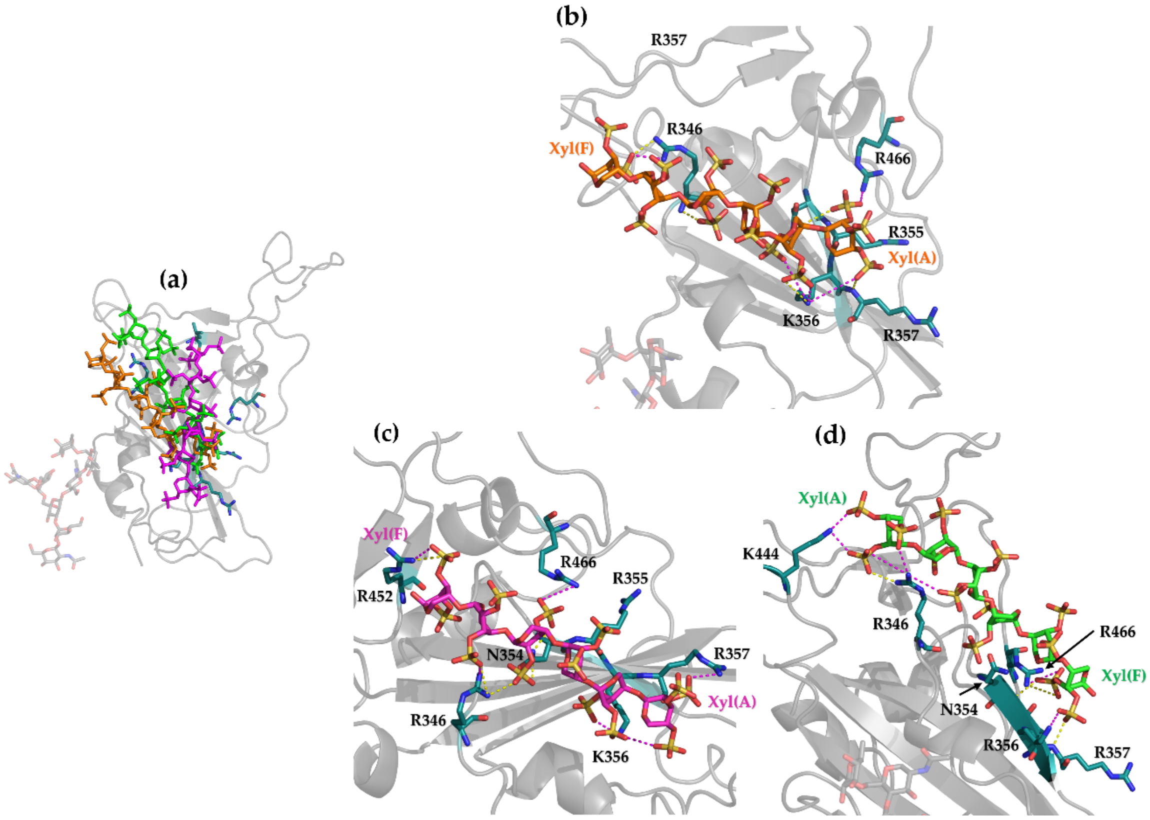


Figure 24 – Superimposition of PPS docked onto the RBD domain of the wild-type (orange), Delta (magenta), and Omicron (green) SARS-CoV-2 spike variants (panel a). Enlarged views of each complex are shown: wild-type (panel b), Delta (panel c), and Omicron (panel d). The RBD region is illustrated as grey carton, while hexa is depicted in stick representation (oxygen in red, nitrogen in blue, and sulfur in yellow). Key interacting residues are shown as sticks with matching color codes (carbon in teal, oxygen in red, nitrogen in blue, and sulfur in yellow). The salt bridges and hydrogen bonds are indicated by magenta and yellow dashed lines, respectively.

Indeed, the wild-type-PPS and Omicron-PPS complexes present the same number of contacted residues (R355, K356, K462, R466 vs. R346, K356, K444, R466), but in the latter, additional short salt-bridge contacts are evident with respect to the sulfate group compared to the former. In fact, in Omicron, the charge-based interactions are closer in general to the protein surface within the expected range for strong salt-bridge or salt-bridge–assisted hydrogen-bond interactions (**Table A. 4**, residue R346 and K444). By contrast, the corresponding contacts in the wild-type occur at slightly longer distances (typically ≥3.6–3.9 Å), and several wild-type interactions extend beyond 4.5 Å, where the electrostatic contribution becomes substantially weaker. In Omicron, the shorter salt bridges fall mostly below 4.0 Å and near 3.0–3.6 Å, well within the geometric criteria for a strong and stable salt bridge that is expected to enhance the binding affinity through an optimization of the electrostatics and directional interactions. On the contrary, the longer distances in the wild type, some above 4.5 Å, indicate weaker and more solvent-exposed ion pairs with strongly reduced energetic contribution. Thus, additional shorter salt bridges in Omicron are consistent with and likely contribute to the increased estimated binding free energy and the higher affinity observed.

4.3 Step 2: ^{1}H-STD NMR interaction experiment

4.3.1 Mapping the glycan binding epitope of the interaction between hexa and Omicron RBD by ^{1}H-STD NMR interaction experiment.

The results of this section and section 4.4 have been published in a scientific paper: ***Mandalari M.****, Parafioriti M., Ni M., Benevelli F., Civera M., Elli S., Guerrini M. Interaction Between Heparan Sulfate Oligosaccharide and the Receptor-Binding Domain of the Wild-Type and Omicron Variant of the SARS-CoV-2 Spike Protein. Biomolecules (2025). https://doi.org/10.3390/biom15091343*

4.3.2 Influence of ligand nature on NMR spectral characteristics

As described in section 3.2.1, ^{1}H STD NMR is a widely used spectroscopic technique for studying interactions in low affinity ligand-macromolecule systems. However, this method cannot be directly applied to investigate the interaction between PPS and the RBD domain, because the ^{1}H NMR resonances are not sufficiently resolved to allow univocal identification of individual residue units. Indeed, the comparison of the NMR profiles of hexa and PPS hexasaccharide (**Figure 25**) illustrated that, in hexa, the anomeric proton of each monosaccharide unit (**H1A** to **H1F**, **Figure 25a**) generates a well-resolved and separated signals, allowing clear assignment of each proton in the spectrum.

In contrast, for PPS, all of the anomeric signals are broad and partially overlapped within a narrow spectral region, preventing unambiguous assignment of the individual residues (**Figure 25b**). Indeed, this spectral crowding is related to the structural features of PPS, namely a homopolymer of xylose 2,3-O-disulfated with the β-D-presence of rare 2-O-branched 2,3-di-O-sulfated 4-O-methylglucuronic acid (MGA), linked to the main poly-xylose chain via a $\alpha(1\rightarrow2)$ glycosidic bond. In the case of the linear poly-xylose chains, the anomeric protons that can be detected in the 2D ^{1}H-^{13}C heteronuclear single quantum correlation (^{1}H-^{13}C HSQC) spectrum are [129, 203]:

- The anomeric proton of reducing end in α and in β conformation (^{1}**Xly**$_{\alpha}$ and ^{1}**Xly**$_{\beta)}$.
- The anomeric proton of the xylose unit adjacent to the reducing end in α conformation (**-**^{1}Xly - ^{1}***Xly***$_{\alpha}$***)***.
- The anomeric proton of non-reducing xylose unit (^{1}***Xly***$_{NR}$).
- The anomeric proton of 2,3-di-O-sulfated xylose unit α1-2 linked to MGA unit (^{1}**Xyl**$_{2MGA}$)
- The anomeric proton of xylose unit α1-2 linked to MGA unit and acetylated in position 3 (^{1}**Xyl**$_{2MGA3Ac}$)
- The anomeric proton of 2,3-O-disulfated xylose unit α1-2 linked to 4-O-sulfated MGA unit (^{1}**Xyl**$_{2MGA4S}$)

Similarly, based on the 2- and 3-OH substitution of the xylose adjacent to the 2-MGA branched Xyl unit, several anomeric signals of the glucuronic acid can also be identified:

- The anomeric proton of MGA linked to a xylose bearing 2-O and 3-O sulfate group (^{1}MGA-Xyl)
- The anomeric proton of MGA linked to a xylose 2-O sulfated and 3-O acetylated (^{1}MGA-Xyl$_{3Ac}$).

The differences in spectra resolution arise from the distinct nature and composition of the two samples. Hexa is a synthetic oligosaccharide, consisting of a single, purified molecular species, which results in sharp, well-resolved NMR signals. PPS, on the other hand, is derived from the fractionation of a heterogeneous mixture of linear and branched polysulfated xylose chains. Although, fractions containing oligosaccharide with a narrow molecular weight, they differ in chain length, arrangement, and sulfation pattern. This structural heterogeneity generates signal broadening and overlaps in the NMR spectra, complicating spectral interpretation and peak assignment. By contrast, the critical requirement for the determination of the binding epitope through ^{1}H-

STD NMR, is the presence of non-overlapping signals, which can be correlated to the structural details and constraints of the molecular recognition event. In this case, the spectral characteristics of PPS, however, preclude this type of analysis.

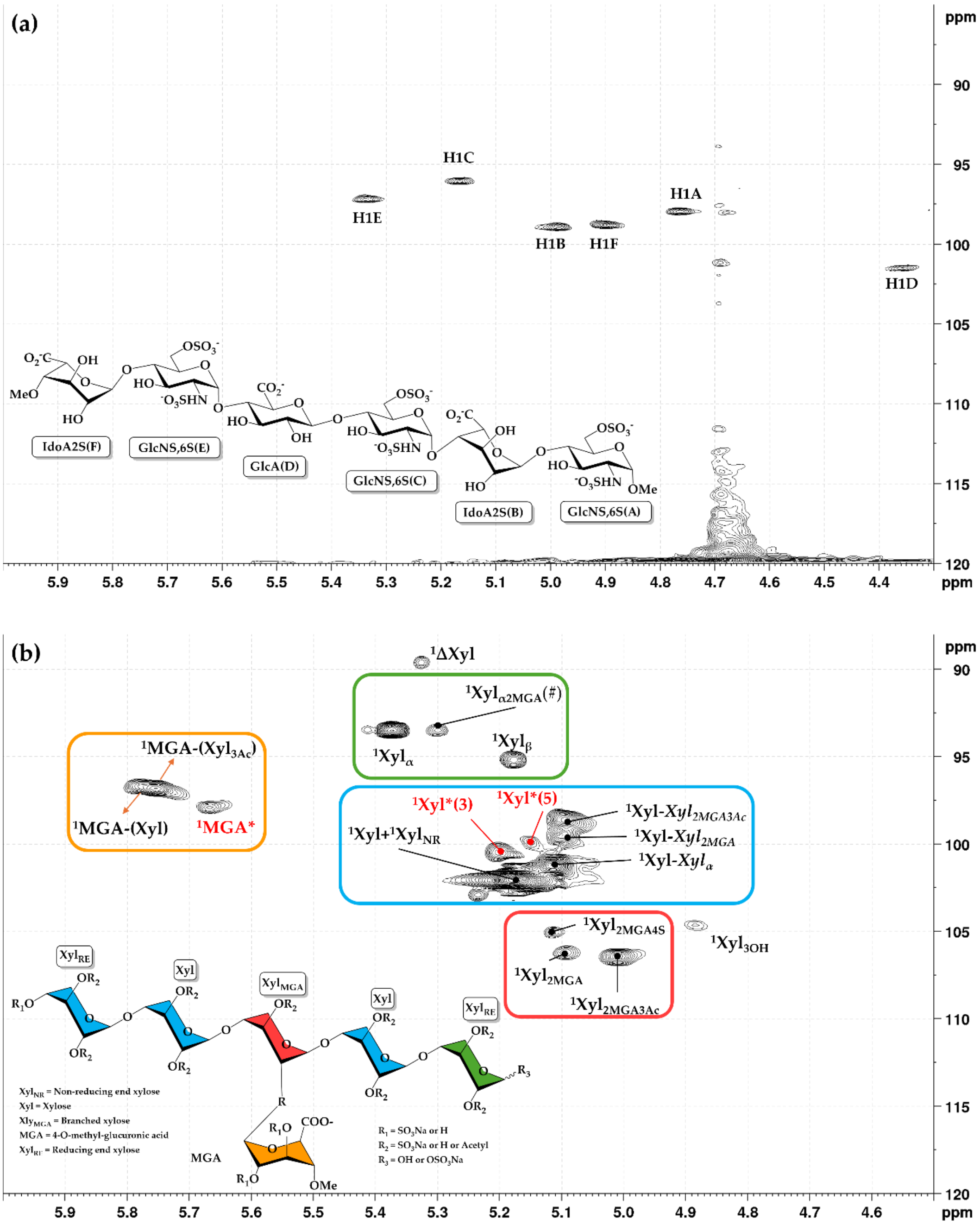


Figure 25 – Comparison between the ^{1}H-^{13}C HSQC NMR spectra of hexa (panel a) and PPS (panel b) in the anomeric spectral region [6-4 ppm for ^{1}H and 85-120mppm for ^{13}C]. The symbol (#) indicates the peak assignment is a hypothesis, requiring further studies to be confirmed. Instead, the red labels indicate the peaks whose assignment is still unknown. (^{1}MGA*, ^{1}Xyl * (3) and ^{1}Xyl * (5) [203]).

4.3.3 The binding epitope of hexa in complex with Omicron RBD

The interactions between hexa and Omicron RBD domain were investigated by ^{1}H-STD NMR experiments. The primary goal was to assess whether the interaction between the hexasaccharide and the RBD occurs, and successively to identify the glycan residues that are in contact with the RBD domain. The STD values at increasing saturation times (for further details, please refer to the section 3.1.2) of the non-overlapping anomeric protons and the two terminal methyl group were measured (**Table 3** and **Figure 26**; see also section 5.3 of the Materials and Methods section), and the STD NMR build-up curves of each was built. The relative STD_0 values were mapped on the ligand structure to obtain the ligand epitope map **Figure 27**.

Table 3 – Absolute STD values for each NMR experiment and normalized STD values at the initial slope (STD_0 %) of non-overlapping protons of hexa (anomeric proton [H1A-H1F], C6 geminal protons of GlcNS,6S (C) [H6C] and the methyl group of non-reducing IdoA2S(F) and reducing end GlcNS,6S(A) [OCH_3(F) and OCH_3(A), respectively].

	Absolute STD values					**Relative STD_0 (%)**
Sat. Time(s)	0.5	1	2	3	5	
OCH_3(A)	0.89	1.43	2.00	2.30	2.58	51%
H1A	0.00	0.41	0.61	0.74	1.07	9%
H1B	0.43	0.69	1.06	1.25	1.38	24%
H1C	0.57	0.87	1.54	1.80	1.97	32%
H6C	1.48	2.04	2.52	2.52	2.84	100%
H1D	1.59	2.28	3.42	3.51	3.74	96%
H1E	0.57	1.05	1.90	2.28	2.55	36%
H1F	0.41	0.63	1.06	1.38	1.36	23%
OCH_3(F)	0.71	1.20	1.83	2.16	2.27	43%

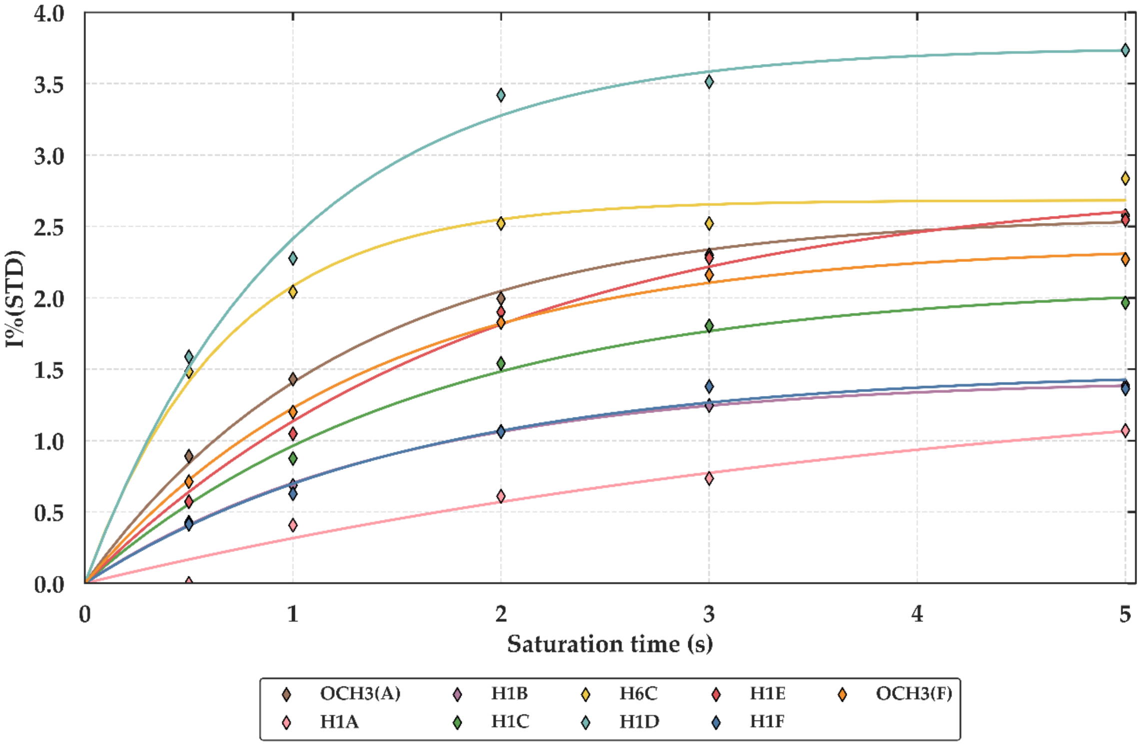


Figure 26 – STD NMR build-up curves for all the monitored hexa protons constructed by plotting the STD factors at different saturation times, ranging from 0.5 to 5 s. Each curve was fitted (colored lines) to the mono-exponential **Equation 3** to derive the ksat and STDmax parameters. The rhombic shaped data points represent the observed STD signal intensities.

Comparing the STD NMR spectra with the reference spectrum revealed that all anomeric protons were affected by the magnetization transfer from the saturated RBD protein, indicating that all the residues of hexa participate in the interaction process (**Figure 27**). Interestingly, the central residues GlcNS,6S(C) and GlcA(D) shows the strongest STD signals: the STD_0% values for protons H6C (the C6 geminal protons of GlcNS,6S(C)) and the anomeric proton H1D were respectively 100% and 96%. In contrast, the remaining protons exhibited lower STD_0 % values, ranging from 10% to 36% (**Table 3**, **Figure 26** and **Figure 27**).

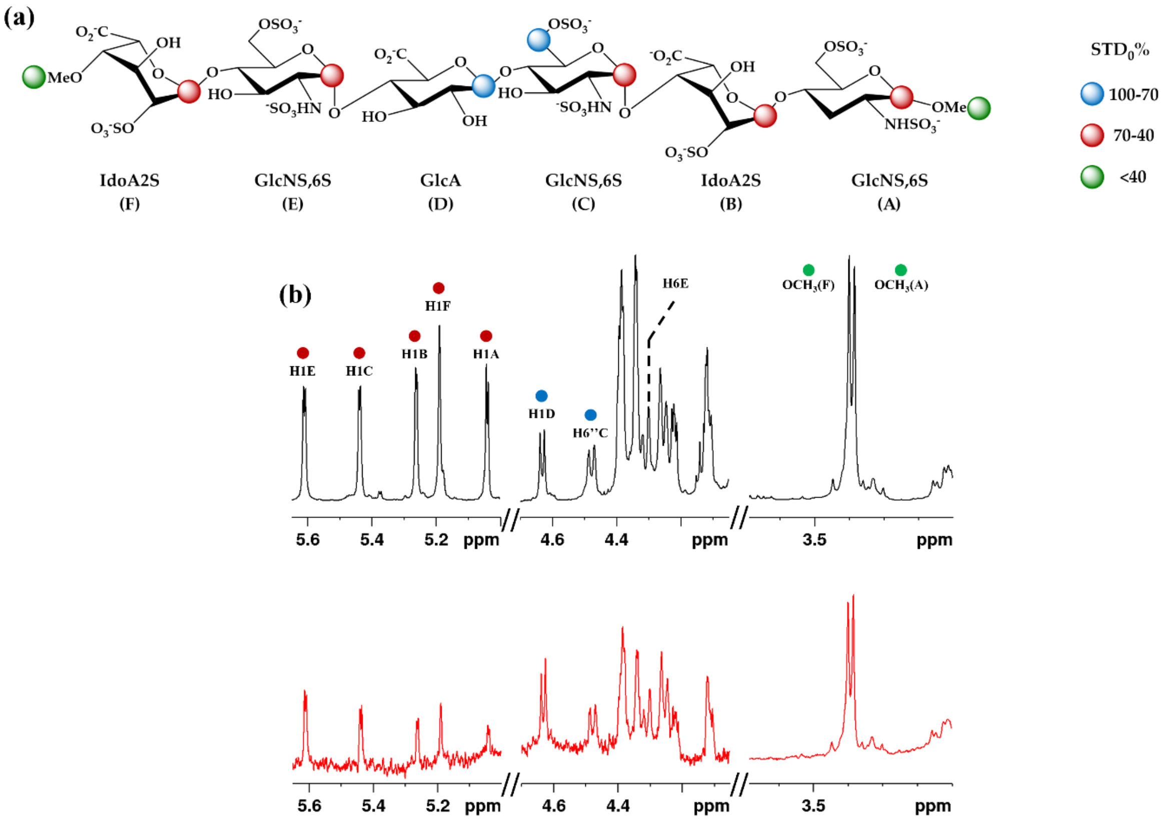


Figure 27 – In panel a, the map of the ligand epitope that interact with Omicron RBD was reported. Panel b shows the off-resonance (black line) and on-resonance (red line) spectra at a saturation time equal of 3 s (T = 293K), with the assignment of the hexa protons indicated above the off-resonance spectrum (b).

Remarkably, as reported in **Figure 27**, the H6E protons (the C6 geminal protons of GlcNS,6S(E)) also showed an appreciable STD intensity. However, the corresponding STD$_0$% value was not determined due to the spectral overlap (**Figure A. 2**).

Since H6C and H6E are approximately oriented in opposite directions in the three-dimensional structure of hexa (**Figure A. 3**), the detection of signals with comparable intensities indicates that the hexasaccharide ligand presumably associates with S1-RBD in two alternative orientations. These orientations represent distinct binding modes in which the glycan engages the protein surface using one of the two available (opposite) surfaces shown in **Figure 17**, that are allowed from the rigid conformations typical of glycosaminoglycans chains such as heparin and heparan sulfate. In this case, either H6C or H6E is oriented towards the RBD interface allowing them to acquire the saturation from RBD in fast exchange conditions. This behaviour is consistent with the multiple poses that are obtained from the docking simulation between hexa and Omicron RBD, and from an earlier study [192], where the interaction involves hexa and the wild-type RBD. In this case, trNOEs (transferred Nuclear Overhauser effect spectroscopy) experiments and docking/MD simulations suggested that the interactions between hexa and the site I of the wild-type RBD, can take place with two different ligand binding modes or orientations [192].

4.4 Step 3: Molecular dynamics simulations

4.4.1 Hexa explores the surface of Omicron and wild-type RBDs.

Selected poses of hexa-Omicron RBD and hexa WT RBD complexes were submitted to multiple independent MD simulations. Precisely, two highest ranked poses, representing the two opposite orientations of the glycan were selected to represent the complex between hexa and Omicron RBD (here labelled as Omi RBD-hexa-I and Omi RBD-hexa-II). A single pose representing the complex between hexa and WT RBD (here labelled as WT-RBD-hexa) was selected (**Figure 28**).

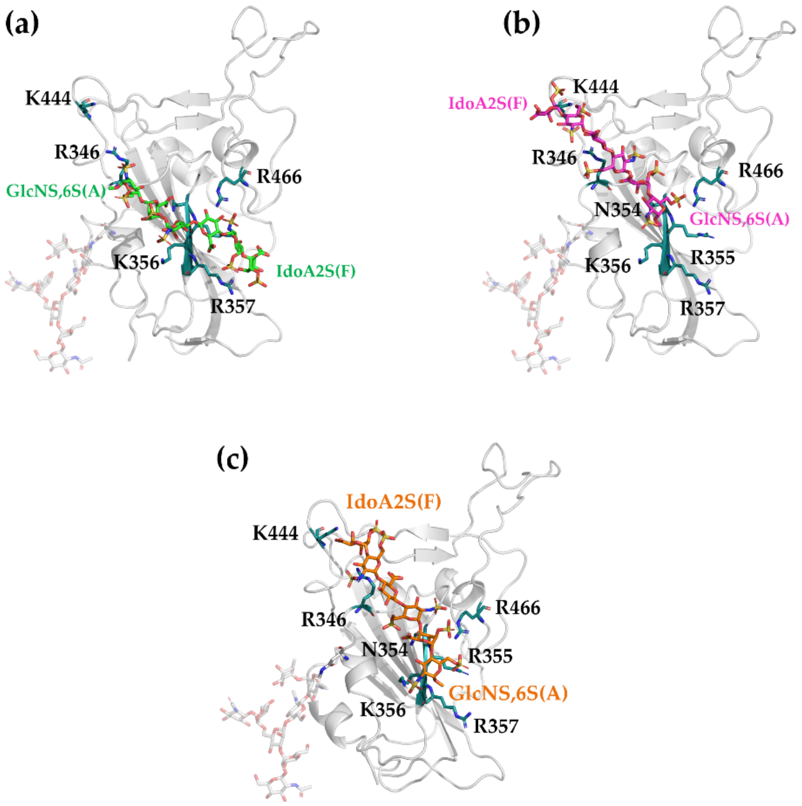


Figure 28 – Top-ranked docking poses of hexa with Omicron RBD domain (Omi-RBD-hexa-I, panel a; Omi-RBD-hexa-II, panel b), and with the wild-type RBD (WT-RBD-hexa, panel c). The RBD protein is illustrated as grey cartoon with the key residues of site I (R346, N354, R355, K356, R357, K444 and R466) are labelled and colored as teal sticks. The hexa is represented by a stick model (carbon in green, magenta, and orange, respectively; oxygen in red, nitrogen in blue, and sulfur in yellow). The non-reducing and reducing ends of the ligand are highlighted with IdoA2S(F) and GlcNS,6S(A) labels, respectively.

For each replica, the root mean square deviation (RMSD) analysis reported in **Figure 29,** highlight a significant variation in the case of the ligand, while the protein remains stable over each production runs, except for replica 1 of Omi-RBD-hexa-II where there was a rearrangement of the 380-390 loop (**Figure 29h**).

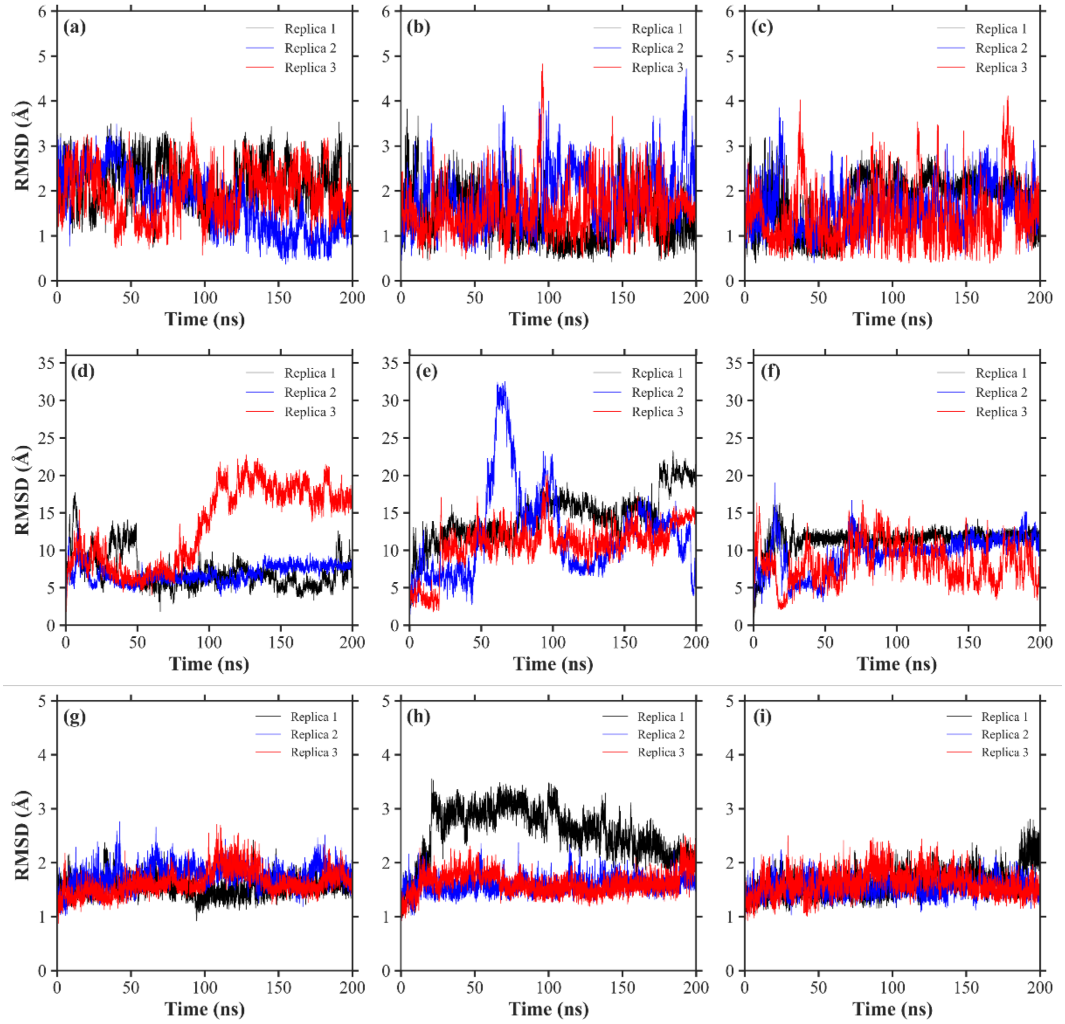


Figure 29 – Root Mean Square Deviation (RMSD) analysis of hexa. The first row shows ligand best-fit RMSD values (calculated after aligning the ligand position to a refence structure, i.e. the first frame); the second row shows the ligand no-fit RMSD, and the third row shows the best-fit RMSD of the RBD protein backbone. The RMSD values were calculated on the Omi-RBD-hexa-I (panel a and d), Omi-RBD-hexa-II (panel b and e), and WT-RBD-hexa (panel c and f) in each replica (black for replica 1, blue for replica 2 and red for replica 3).

Since the ligand maintains the unbound state conformation in all the runs, as indicated by the best-fit RMSD values (calculated after aligning the ligand position to a refence structure, i.e., the first frame ; **Figure 29a-c**), the fluctuation observed in the corresponding ligand unfitted RMSD (where no fitting is applied to the ligand, ranging between 10 and 30 Å ; **Figure 29 c-f**), depends only on the mobility of hexa along the channel. The inter-glycosidic bonds remain essentially stable across all residue pairs, exhibiting similar ϕ_i/ψ_i distribution in all the simulated complexes (**Figure 30** and **Table A. 6**). This result agrees with the intrinsic structure and conformation of hexa, where strong electrostatic repulsions between sulfate and carboxyl groups, combined with steric hindrance, dominate and force the molecule to preserve its linear, stiff conformation.

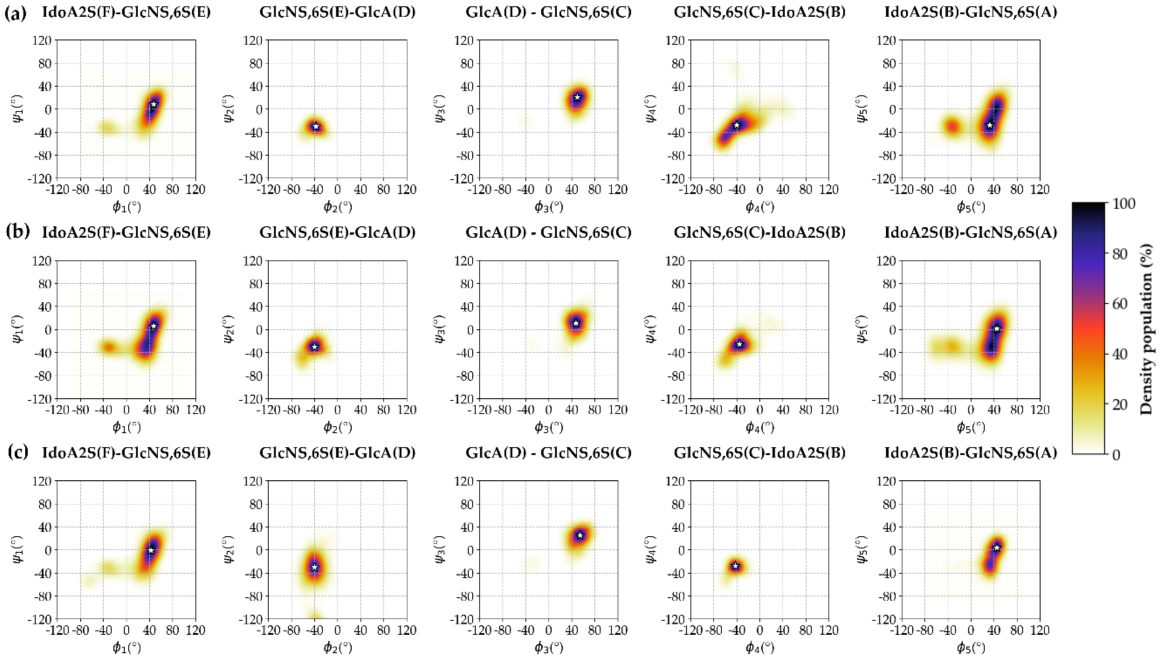


Figure 30 – Ramachandran-type plots of ligand inter-glycosidic dihedral angle ϕ_i/ψ_i in Omi-RBD-hexa-I (panel a), Omi RBD-hexa-II (panel b), and WT-RBD-hexa (panel c) . The ϕ_i/ψ_i were calculated using the non-IUPAC definition. Specifically, ϕ_i and ψ_i are defined by $H1_n - C1_n - O4_n - C4_{(n+1)}$ torsion and $C1_n - O4_n - C4_{(n+1)} - H4_{(n+1)}$, respectively. The population density is reported by using a color gradient (from white to dark blue). The most populated ϕ_i/ψ_i state is highlighted as a white star (see **Table A. 6**).

Therefore, the observed increase in RMSD values reflects solely the motion of hexa, experiencing during the interaction with the shallow and concave surface of site I. Although the sulfate groups of the ligand establish electrostatic interactions with the positively charged side chains, it still possesses kinetic energy to reach the surrounding area near the site I. During the simulations, hexa translates and rotates causing partial detachment from the RBD surface. To collect only those frames, in which the complex is considered formed, the merged trajectory was filtered using the protein-ligand contact information gained by ^{1}H STD NMR.

As reported in section 4.3 (see **Figure 27**), in the binding epitope of Omicron RBD – hexa complex, GlcA(D) and GlcNS,6S(C) exhibits the strongest ^{1}H STD signals, indicating that these residues are, on average, positioned closer to the protein surface than the neighboring disaccharide units. Thus, based on the aforementioned constraint, a Python script was developed and optimized to filter the merged MD trajectories, selecting only those frames in which the disaccharide unit GlcA(D) - GlcNS,6S(C) was located close to the protein surface (see section 5.2.3). By applying this distance-based filtering approach, the fraction of simulated frames in which hexa remain bound to the RBD surface was 73% for Omi-RBD-hexa-I, 53% for Omi-RBD-hexa-II. For WT, previous STD NMR data [192] shows that all the hexa residues (expected for IdoA2S(F)) present high/medium intensity STD NMR signals, indicating that the ligand is in close contact with the protein. Hence the same filtering approach was applied, yielding to 39% of collected frames for WT-RBD-hexa. These frame percentages are consistent with the previous observations [111, 204] (see section 4.2), which indicate that the ligand displays different binding affinities against the two RBDs (**Table 2**). By visual inspection, the wild-type RBD complex was observed to be more susceptible to partial detachment of the oligosaccharide from the protein surface in comparison to the two poses that represent the interaction between hexa and Omicron RBD. This result agrees with the lower binding energy observed in the former case.

4.4.2 Analysis of the interaction between hexa and the wild-type and Omicron RBDs.

The analysis of polar contacts was performed to identify the amino acids residues that remain in persistent contact with hexa (see section 5.2.2 for further details). The frequency of each contact was computed as a percentage of the entire trajectory and visualized in **Figure 31** as contact heatmaps.

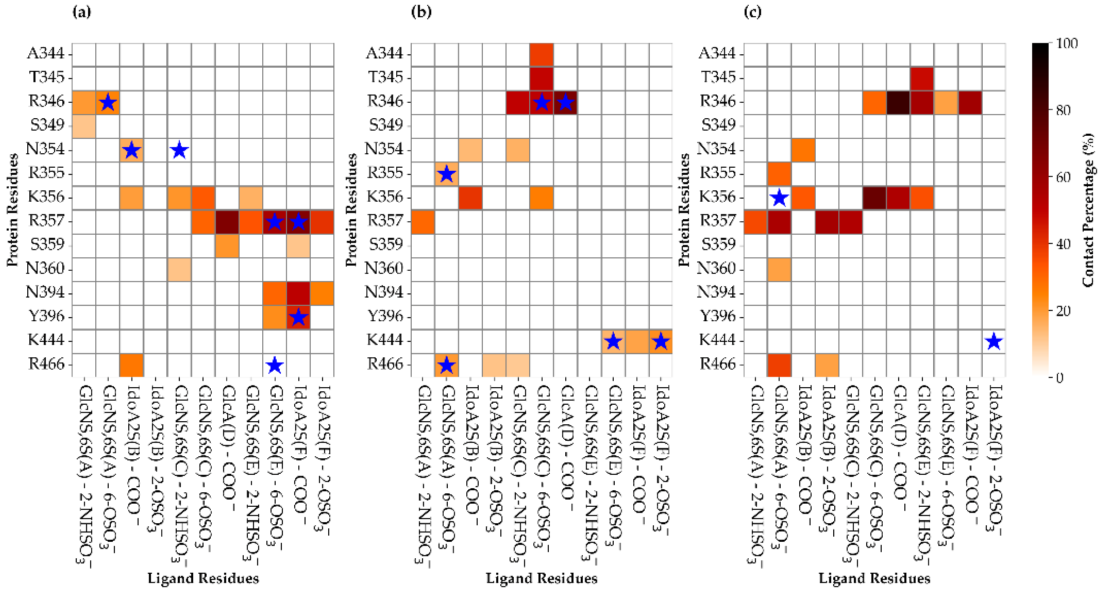


Figure 31 – Contact heatmap of hexa-RBD protein contacts established in Omi-RBD-hexa-I (panel a), Omi-RB-hexa-II (panel b), and WT-RBD-hexa (panel c). The color gradient indicates the contact persistence: lighter tones (white to orange) represent transient interactions, whereas darker tones (dark red to black) denote highly persistent contact. Interaction observed in the initial docking pose (i.e., the starting frame of the MD simulation) are marked with blue stars.

All the contacts observed in the heatmap were categorized as electrostatic or polar, depending on the chemical nature of the participating groups (see section 5.2.2). Electrostatic interactions were assigned to pairs of oppositely charged residues, while polar interactions and potential hydrogen bonds were defined as contact between polar atoms with an average distance ≤ 4.5 Å (**Table A.** 7)

- **Omi-RBD-hexa-I complex.**

A rearrangement of the initial binding epitope in the Omi-RBD-hexa-I complex was observed, and the ligand explores the surrounding area of site I (**Figure 31a**). The residues predominantly engaged in contact with the ligand are R357 (67%), Y396 (43%), K356 (32%), R466 (27%), N394 (25%), and R346 (20%). Among these, R357 exhibits the most frequents interaction, forming electrostatic interaction with GlcNS,6S(C), GlcA(D) and Ido2S(F), with contact frequency ranging between 66% to 33% (**Table A. 7**). Similarly, K356 established new electrostatic interactions with the central region of the oligosaccharide (GlcNS,6S(C)).

Two neighboring residues, S359 and N360, proximal to site I and known to be implicated in GAG-SARS-CoV-2 recognition [205], were also found to interact with the central residues GlcNS,6S(C) and GlcA(D). Additional polar contacts N394 – GlcNS,6S(E) and Y396 – IdoA2S(F) were formed, whereas the initial contact between R466 and GlcNS,6S(E) was lost during the simulation. These findings indicate that hexa extends its interaction beyond those predicted in the initial docking pose. The ligand, through translation and rotational motions, can explore the protein surface, leading to the presence of alterative binding modes.

- **Omi-RBD-hexa-II complex**

In the Omi-RBD-hexa-II complex, most of the contacts present in the docking pose were retained, while several new interactions were also established during the simulation (**Figure 31b**). The residue R346 displayed stable interactions with the central residues GlcNS,6S(C) and GlcA(D), with contact frequencies of 51% and 67%, respectively (**Table A. 7**). These contacts were further stabilized by less frequent, yet relevant, polar interaction such as T346 – GlcNS,6S(C) (46%) and GlcNS,6S(C) – A344 (38%). In addition, residue R356 and R357, located in the core of site I, frequently interact with IdoA2S (B) (40%) and GlcNS,6S(A) (27%).

Notably, the heatmaps for Omi-RBD-hexa-I and Omi-RBD-hexa-II (**Figure 31a,b**) reveal that the ligand interact with two distinct binding epitope of the Omicron RBD surface, each of them corresponding to an opposite face of the oligosaccharide (**Figure 28** and **Figure A. 3**).

- **WT-RBD-hexa complex**

MD simulation predicts that the oligosaccharide in WT-RBD-hexa complex shows greater mobility in comparison to the Omicron RBD-hexa complexes. The glycan translates across site I, show in **Figure 31c**, in which the pattern of persistent contacts deviates from those observed in the docking pose at the beginning of the simulation (blue stars). R346 remains in stable contact with GlcA(D) (85%), IdoA2S(F) (60%) and GlcNS,6S(E) (57%), while K356 and R357 established electrostatic interactions with the central GlcNS,6S(C) residue (72% and 55%, respectively). K356 also displayed weaker interactions with terminal residues GlcNS,6S(E) and IdoA2S(B) (**Table A. 7**). Additional residues, including N354, R355, N360, and 466, shows frequencies (18-57%) with the terminal disaccharide unit IdoA2S(B) – GlcNS,6S(A). Interestingly, several of these contacts, such as R346-GlcA(D)/GlcNS,6S(C), K356-IdoA2S(B)/GlcNS,6S(C), R355-GlcNS,6S(A), and R357-GlcNS,6S(A), were also detected in the Omi-RBD-hexa-II complex, indicating conserved features of the binding interface across the variants.

A central observation derived from both the contact analysis and visual inspection of the MD trajectories concerns the pronounced mobility of the hexa ligand. In all the systems, the ligand exhibits a dynamic pattern of repetitive partial **dissociation and re-binding** events, mediated by the ion pair interaction between the anionic functional groups of the oligosaccharide and the cationic side chain of Arg and Lys residues.

Specifically, the N- and 6-O-SO_3^- groups of GlcNS,6S residue (A, C, and E) and the COO^- group of IdoA2S units (B and F) form the same key positive and/or polar side chain interactions with R346, K356, R357, N394, and N360 in the Omi-hexa-I complex (**Figure 31a**). This indicates that the glycan alternately contacts these residues while exploring the shallow cavity that defines site I. Similarly, in the Omi-RBD-hexa-II complex, the glycan binds in an opposite orientation, allowing its negatively charged groups to interacts with R355 and K356 (**Figure 31b**), whereas in the WT-RBD-hexa complex, the main contacts involve K356 and R357 (**Figure 31c**).

Overall, these results reveal a high degree of ligand mobility, where hexa undergoes combined rotational and translational motions along the positively charged surface of the RBD. This dynamic behavior delineates a more extended and diffuse binding epitope, encompassing a broader region of the protein surface than the size of the hexasaccharide probe. Such mobility (which leads to the formation of a complex pattern of interactions) combined with the conformational rigidity of the ligand (which reduces the entropic penalty upon binding) facilitates a balance between the enthalpic stabilization and entropic freedom, contributing to the overall binding features of the glycan-RBD complexes.

4.4.3 The conformational properties of the N343 glycan affect the interaction between hexa and RBDs.

Interestingly, the FA2G2 glycan in Omi-RBD-hexa-I and Omi-RBD-hexa-II displays a significantly reduced mobility in comparison to that shown by WT-RBD-hexa. (**Figure 32**).

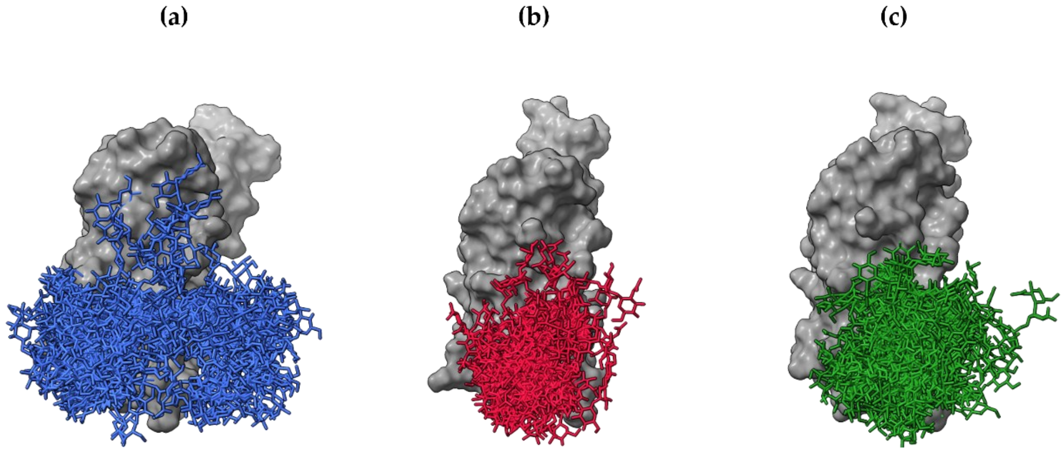


Figure 32 – Sampling of FA2G2 orientations over the merged trajectory of WT-RBD-hexa (panel a), Omi-RBD-hexa-I (panel b), and Omi-RBD-hexa-II (panel c). Snapshots were taken every 10 ns, yielding a total of 60 structures. The protein is depicted as a grey surface, while the FA2G2 glycan is represented as sticks: blue for WT-S1-RBD-hexa, red for Omi-RBD-hexa-I, and green for Omi-RBD-hexa-II, respectively.

Indeed, in the latter case, the FA2G2 octasaccharide adopts two preferred orientations, characterized by χ_1 dihedral angle of the N343 side chain. This angle, defined by the $C - C_\alpha - C_\beta - C_\gamma$ atoms, populated with similar probability two distinct rotameric states: the *trans* (~160°) and the *gauche*$^+$ conformation (~70°) (**Figure 33a**).

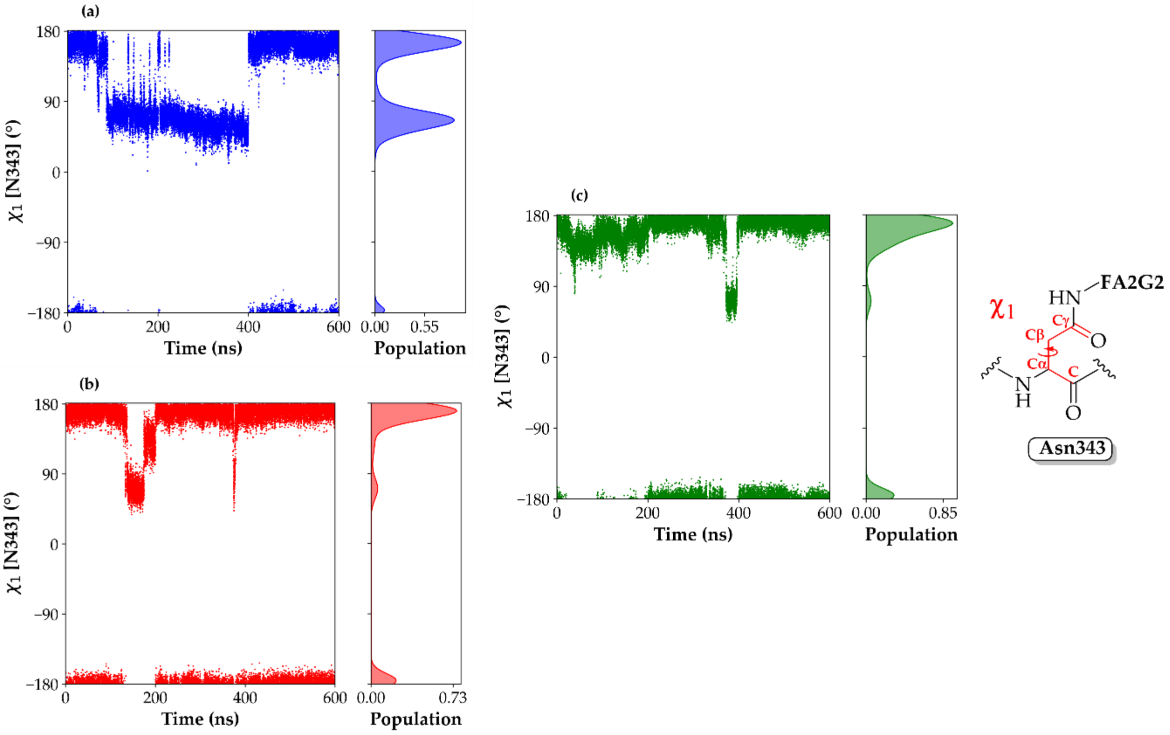


Figure 33 – Scatter plots of the χ_1 torsion angle of the N343 side chain (reported on the right) are shown for WT-RBD-hexa (panel a, blue), Omi-RBD-hexa-I (panel b, red), and Omi-RBD-hexa-II (panel c, green). The corresponding histogram distributions are displayed along the y-axis.

Interestingly, the orientation that FA2G2 assumes according to the rotameric state of χ_1 can interfere with the site I of RBD, eventually hiding it. When χ_1 is in the *trans* state (~160°), the FA2G2 occupies a pocket adjacent to the α-helix (residues 366–371) and the neighboring loop (residues 372–375). In this conformation (**Figure 34a**), the glycan remained spatially separated from heparin-binding site I without hiding it 99.

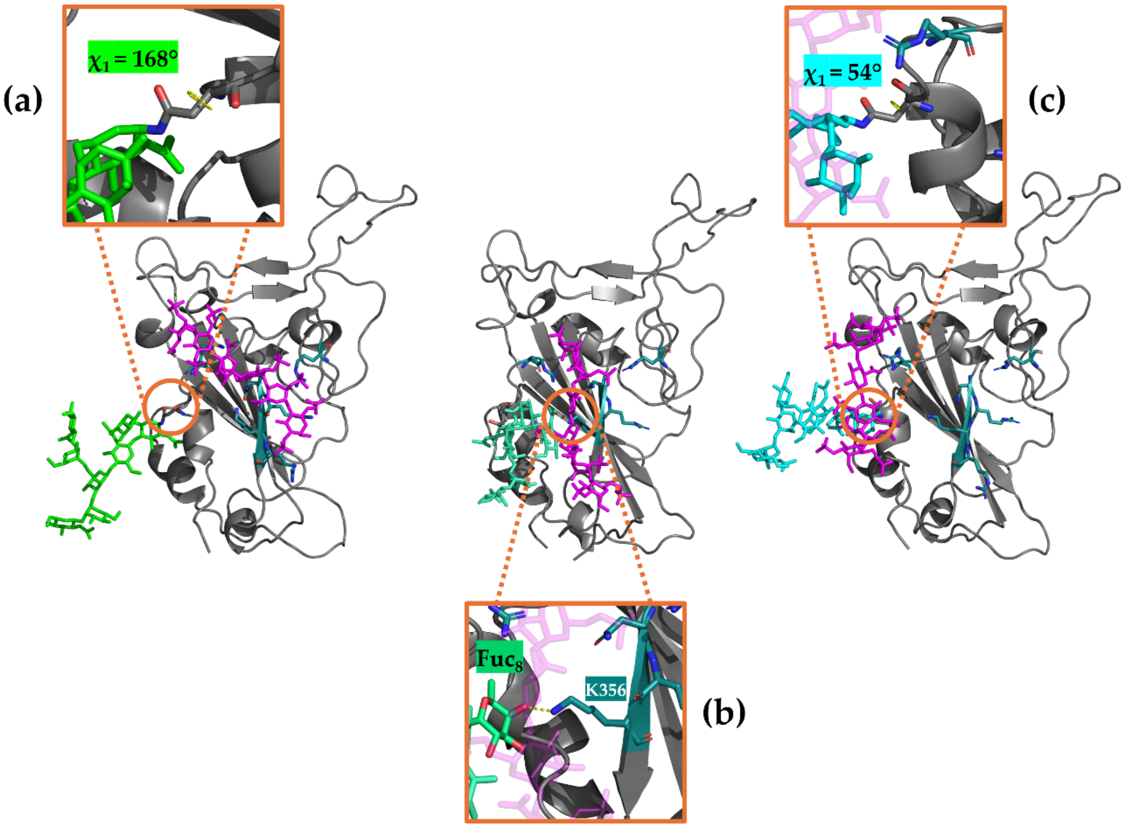


Figure 34 – Representative snapshots of WT-RBD, highlighting the two observed X1 population (**a, b**) and the perturbative effect of FA2G2 on hexa binding (c). The protein is shown as a grey cartoon, and binding site residues R346, N354, R355, K356, R357, and R466 are highlighted in teal. The FA2G2 octasaccharide is represented as sticks: green indicates the population in anti-conformation, while greencyan and cyan corresponds to the population in **gauche**$^+$ state°. Enlarged views show the χ_1 dihedral angle in the *trans* conformation (panel a), the interaction of FA2G2 with K356 at site I, and the *gauche*$^+$ conformation (panel c).

By contrast, when χ_1 is in the *gauche*$^+$ state (~70°), the FA2G2 glycan reoriented toward site I, establishing molecular contacts with both the bound hexasaccharide (hexa) and key site I residues, particularly K356 (**Figure 34b**). In this conformation, FA2G2 affects the hexa in bound state with site I of RBD, inducing a partial displacement of ligand from its binding site (**Figure 34c**).

In stark contrast, FA2G2 in Omi-RBD-hexa-I and Omi-RBD-hexa-II complexes displayed only the *trans* state (**Figure 33b-c**). In this case the χ_1 populate preferentially the trans conformation (χ_1 ~170°) (**Figure 35a,** green colored).

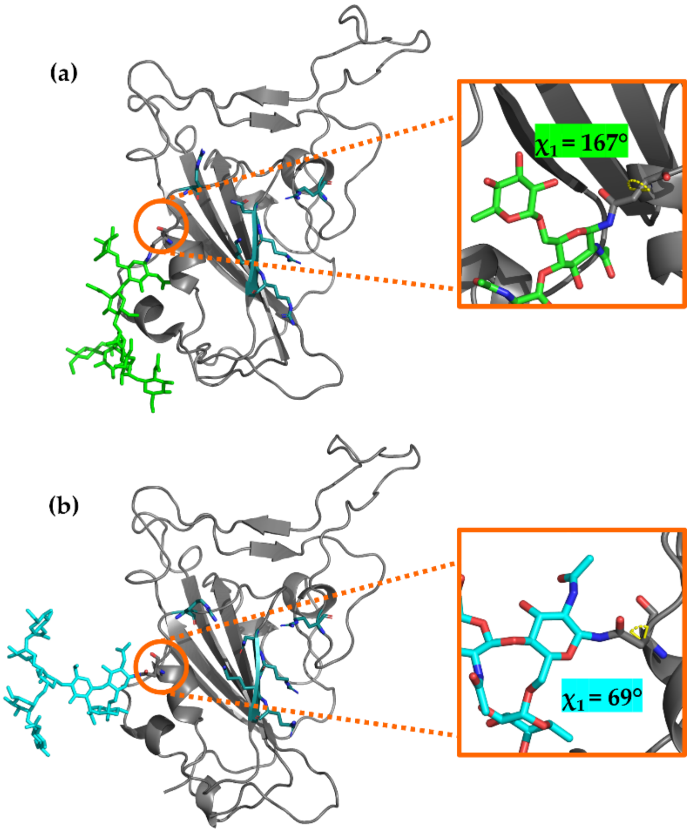


Figure 35 – Representative snapshots of Omi-RBD-hexa-I, highlighting the two observed X1 population (**a, b**). The protein is shown as a grey cartoon, and binding site residues R346, N354, R355, K356, R357, and R466 are highlighted in teal. The FA2G2 octasaccharide is represented as sticks: green indicates the population in anti-conformation, while greencyan and cyan corresponds to the population in *gauche*$^+$ state. Enlarged views show the χ_1 dihedral angle in the *trans* conformation (panel a), and the *gauche*$^+$ conformation (panel b).

This conformational preference positioned the FA2G2 proximal to the 372–375 loop, thereby eliminating competitive interference with hexa that occupy the site I of RBD. This stable configuration is depicted by the green-coloured glycan in **Figure**

35a. A minor conformational substate (population <10%) was observed with χ_1 ~70°, in which the FA2G2 octasaccharide became solvent-exposed and lost the stabilizing contacts with the RBD surface, as shown by the cyan-colored glycan in **Figure 35b**. The bimodal orientation of FA2G2 observed in both RBDs appears induced by the presence of hexa. Indeed, the MD simulations of wild-type and Omicron apo RBD (i.e. without ligand), using the same MD protocol (see section 5.2.2), show a comparable population of χ_1 (**Figure 36**), with a small (1%) percentage of *gauche*+ only in wild-type RBD.

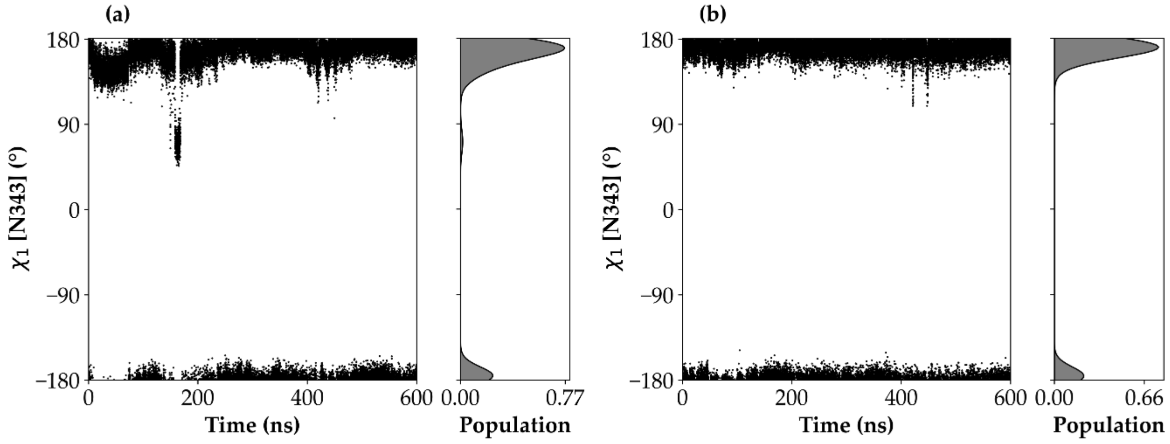


Figure 36 – Scatter plots of the χ_1 torsion angle of the N343 side chain (reported on the right) are shown for the apo form of wild-type RBD (panel a) and Omicron RBD Omi-RBD-hexa-I (panel b). The corresponding histogram distributions are displayed along the y-axis.

The hexa in bound state with wild-type RBD shows a greater population of the *gauche*$^+$ state in comparison to the corresponding apo system: this corresponds to enhanced perturbative effects on the bound hexa binding (**Figure 34c**). Differently, mutations adjacent to site I in hexa - Omicron RBD, and Apo Omicron RBD systems, restricts the motion of N343-linked FA2G2 and narrows its conformational landscape.

A rationale for this reduced mobility is correlated to the G339D mutation in the Omicron variant of RBD. In the WT RBD, the G339 residue lies between the glycan at N343 and site I residues, imposing a minimal steric or electrostatic barriers to the movement of FA2G2 (**Figure 37a** and **Figure A. 5a**).

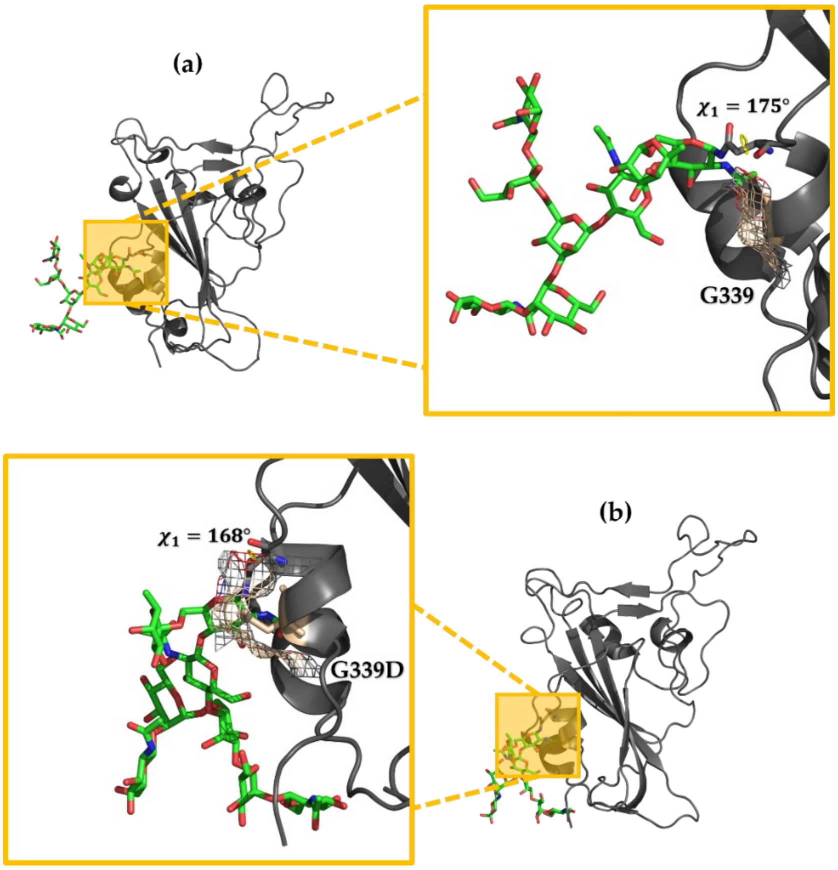


Figure 37 – The G339D mutation in Omicron S1-RBD reduces FA2G2 glycan flexibility. The RBD glycosylated domain is shown as a grey cartoon, with the FA2G2 glycan depicted as sticks (green for carbon, red for oxygen, blue for nitrogen). Residue 339 is highlighted in light pink (Gly in the wild-type, Asp in the Omicron variant) and the occupancy is depicted as mesh surface. (**a**) In the wild-type, the small, apolar G339 allows the FA2G2 glycan to sample multiple orientations, transitioning between *trans* and *gauche*$^{+}$ conformations. (**b**) In Omicron, the bulkier, negatively charged D339 forms steric and electrostatic interactions with the FA2G2 residues (GlcNAc, Man and Fuc), restricting its motion to the *trans* state.

Upon mutation to aspartic acid (G339D) in Omicron, a larger and negatively charged side chain is introduced into this region (**Figure 37b**). This side chain forms persistent interactions with the GlcNAc, Man and Fuc units of the N343-linked glycan, thereby physically and electrostatically constraining its range of motion (**Figure A. 5b-c**). These interactions are persistent throughout the simulation, anchoring the glycan and preventing excursions toward site I of the RBD. Consequently, in Omicron the FA2G2 glycan becomes confined to a narrower set of conformations (a phenomenon directly observable in the distribution plots) consistent with a smaller root mean square fluctuation (RMSF) values for the affected loop and helical residues near the glycosylation site (**Figure 38a**). In contrast, in wild-type RBD, the higher mobility of hexa enhances the flexibility of the adjacent protein region, resulting in higher RMSF values within the same area (**Figure 38b**).

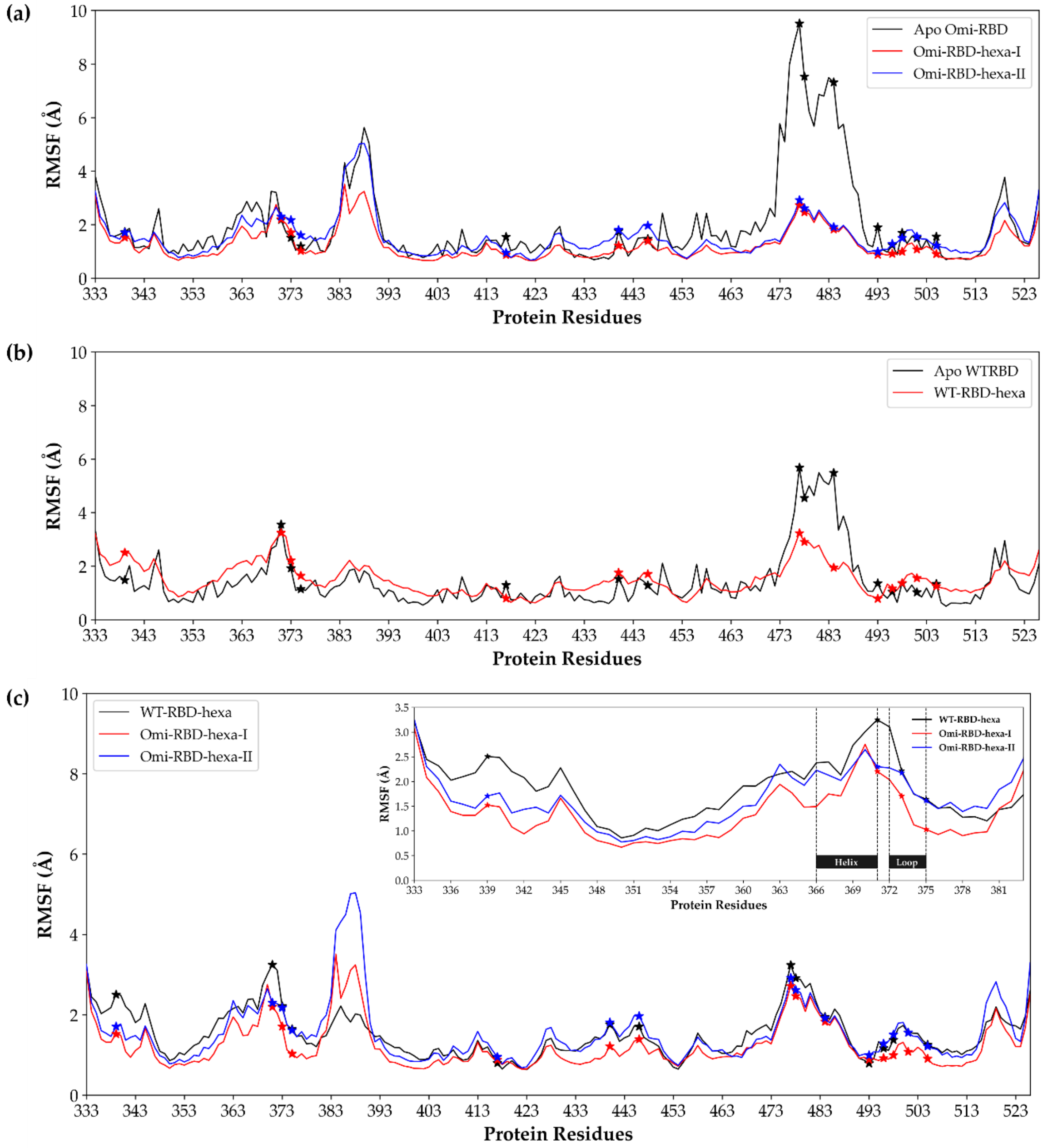


Figure 38 – Root means square fluctuation (RMSF) profile of the RBD backbone calculated on the merged trajectories of Omi-RBD in its apo form, Omi-RBD-hexa-I, Omi-S1-RBD-hexa-II are shown

in panel a; the corresponding profiles for the apo form of WT RBD and WT-S1-RBD-hexa systems are shown in panel b. Panel (c) compares the RMSF of the three complexed RBD systems: Omi-S1-RBD-hexa-I (black line), Omi-S1-RBD-hexa-II (blue line), and WT-S1-RBD-hexa (red line). The mutations present in the RBD region of SARS-CoV-2 Omicron variant (G339D, S371L, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, Q496S, Q498R, N501Y, Y505H) are indicate with a star to highlight their position.

Moreover, additional Omicron-specific substitutions S371L, S373P, and S375F further reinforce local hydrophobicity (**Figure 38c**). These observations align with previous computational analyses predicting that the Omicron variant exhibits smaller RMSF values compared to the wild-type RBD, suggesting the formation of a more stable and compact protein structure [206-210].

However, another important observation that emerges from the RMSF analysis is the stabilizing effect induced by hexa binding in the loop 470-490, located in the distal area of RBM (**Figure 10**). Indeed, the presence of the ligand has an important impact on this disordered region, which results in being more rigid compared to the apo form (**Figure 38a-b)**. This increased stiffness of the flexible loop 470-490 may represent a key allosteric mechanism, whereby the binding of hexa promotes a stabilization of selected conformations in a region remote from the direct ligand-binding interface. Such allosteric stiffness is consistent with literature evidence showing that heparan sulfate binding promotes the adoption of the energetically favorable "open" RBD conformation, characterized by reduced conformational dynamics and enhanced competence for hACE2 recognition [205]. Hence, also in the RBD domain alone, the stabilization of this previously disordered loop may therefore facilitate the transition toward a conformation with higher affinity for the human receptor, while simultaneously reducing the entropic penalty associated with hACE2 binding, may lead to an improved interaction of the RBD and hACE2 receptor.

4.5 Step 4: NMR-driven analysis of MD runs

4.5.1 Selection of WT and Omicron complexes using the NMR and MD data

To determine the putative *"3D atomistic model"* that describes the interaction between the hexa and RBD of SARS-CoV-2 in Omicron variant, a synergetic integration of the experimental and theoretical approaches was performed. The acquisition of the ^{1}H-STD NMR spectra provides us all the information on the ligand binding epitope, thereby suggesting that the oligosaccharide engaged in proximal contact with the surface of Omicron RBD (**Table 3** and **Figure 27**). In parallel, MD simulations generated an ensemble of geometries, representing multiple binding modes with which hexa can interact with site I of RBD. The integration of these two complementary sources of information about our system was done by the simulation of selected ^{1}H-STD NMR signals, followed by their comparison with the corresponding experimental spectra. This analysis require the application of an software RedMat [211], developed by the group of Prof. J Angulo (IIQ-CSIC), thereby yielding an experimentally validated structural model of the Omicron RBD–hexa complex.

- ***Structural characterization of Omicron RBD–Hexa complexes in agreement with STD$_0$ data.***

The STD$_0$ intensities of the central residues of hexa (from GlcNS,6S(E) to IdoA2S(B) were simulated and compared with the corresponding experimental counterpart (see section 5.2.4). The terminal units of hexa, GlcNS,6S(A) and IdoA2(F), were excluded due to their high conformational mobility as evidenced in **Figure 30**. The determination of R-NOE values for both Omi-RBD-hexa-I and Omi-RBD-hexa-II systems, reveals a heterogenous level of agreement: only 25% of the conformations sampled for the Omi-RBD-hexa-I complex and 18% for Omi-RBD-hexa-II complex exhibit a R-NOE factor below 0.3 (**Figure 39**).

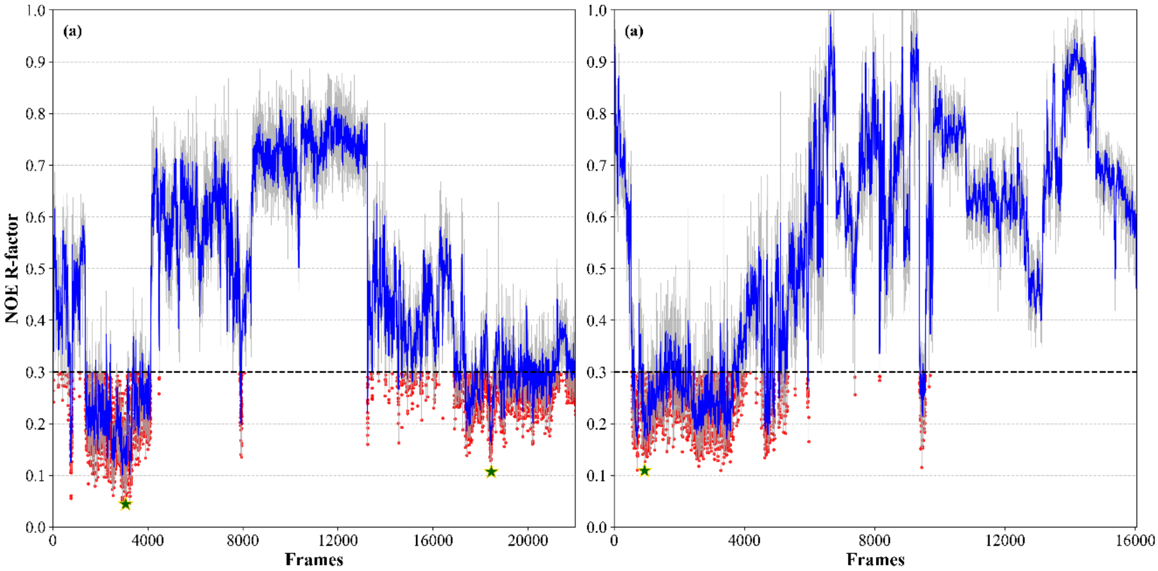


Figure 39 – Evolution of the R-NOE factor in RBD-Omi-hexa-I complex (panel a) and RBD-Omi-hexa-II (panel b) filtered trajectories. The grey line represents the R-NOE values in each frame, while the blue line shows the rolling average (window size = 10). Structures with R-NOE values below 0.3 are marked with red dots. The frame with the lowest R-NOE factor is highlighted with a green star.

To determine the dominant and most experimentally consistent binding modes, cluster analysis was performed on those frames with R-NOE factor ≤ 0.3 in Omi-RBD-hexa-I and Omi-RBD-hexa-II trajectories. For each identified cluster, the structure with the lowest R-NOE was selected as the representative cluster geometry and is depicted in **Figure 40**.

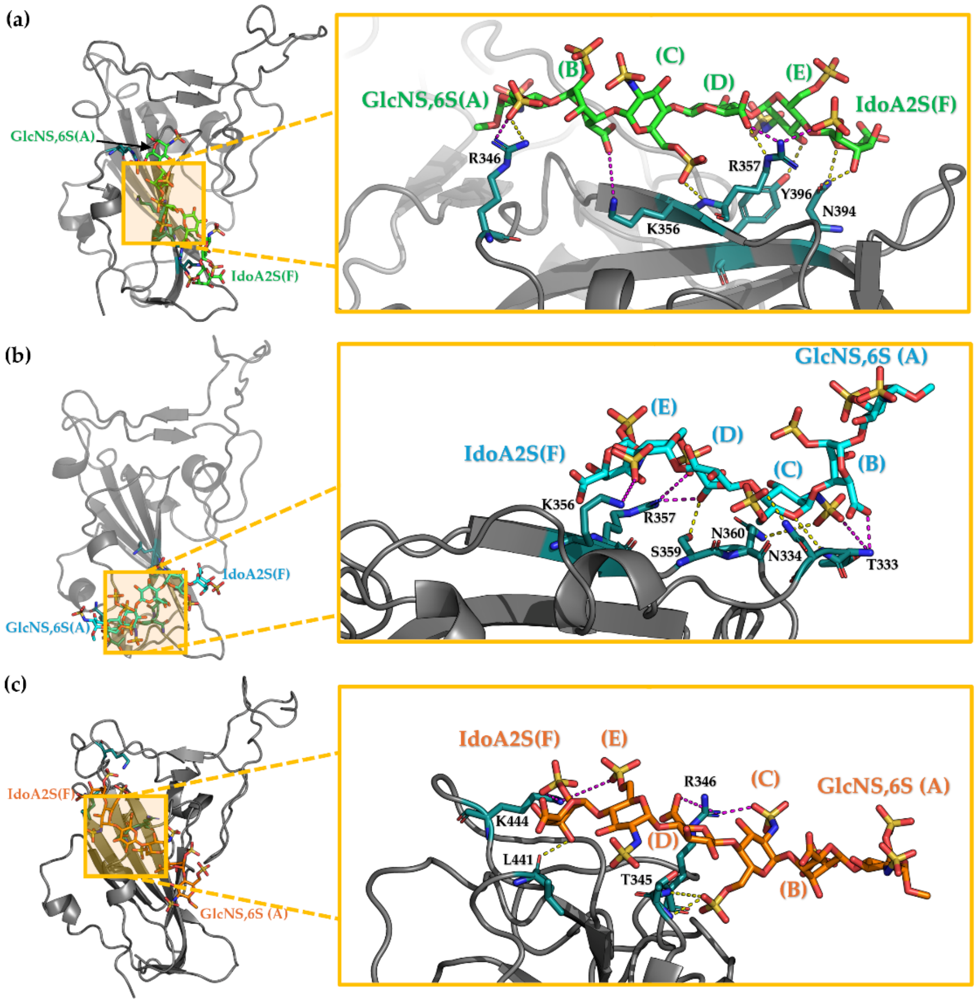


Figure 40 – Representative structures of the main clusters of the Omicron RBD and hexa systems: the first and second clusters of Omi-RBD-hexa-I (panels a and b) and the main cluster of Omi-RBD-hexa-II (panel c). These conformations correspond to R-NOE factors of 0.04, 0.11, and 0.11, respectively. Hexa is depicted as sticks, with green and cyan carbons for the two clusters of hexa-I and orange carbons for hexa-II. Atoms of oxygen, sulfur and nitrogen are colored red, yellow, and blue, respectively. Each monosaccharide unit is labelled (A-F), and the protein backbone is shown as a grey cartoon. Residues involved in the interaction are represented as teal sticks, with the magenta indicating salt bridges and yellow representing hydrogen bonds. The FA2G2 octasaccharide, linked to the N343 and all hydrogen atoms are omitted for clarity.

The cluster analysis of Omi-RBD-hexa-I complex identifies two major families, with 43% and 39% populations, respectively (**Table A. 8**). They correspond to two alternative binding modes.

In the first cluster (43%), the ligand occupied the central region of site I (**Figure 40a**). The $C6\text{-}OSO_3^-$ of GlcNS,6S(C) forms a hydrogen bond with R357 backbone, while The GlcA(D) carboxylate group engages in a salt bridge with the same residue. In addition, GlcNS,6S(A) established a salt bridge with R346, while IdoA2S(F) forms a hydrogen bond with N394.

In the second cluster (39%), hexa is located in a region near site I, establishing contacts with residues S359 and N360 (**Figure 40b**). In this alternative binding mode, the C6-OSO_3^- of GlcNS,6S(C) forms a hydrogen bond with N334, while the 2-$NHSO_3^-$ is hydrogen-bonded to T333 and N360. Additionally, the COO- group of GlcA(D) engages via a salt bridge with R357 and a hydrogen bond with S359.

For the Omi-RBD–hexa-II system, one cluster was identified, with relative populations of 73%. (**Table A. 8**). The most populated cluster (**Figure 40c**) corresponds to a pose consistent with the original docking configuration. In this geometry, IdoA2S(F) is oriented toward the ACE region, where the 3-OH group forms a hydrogen bond and the backbone carbonyl of L441, together with ionic interactions involving the 2-OSO_3^- and the K444 residue. Additionally, the $NHSO_3^-$ and COO^- groups of GlcNS,6S(C) and GlcA(D), respectively, engage in ionic interactions with R346, while the C6-OSO_3^- group of residue C establishes a hydrogen bond with the backbone NH of T345.

Collectively, these findings demonstrate that the Omicron RBD–hexa complexes adopt multiple binding modes consistent with the experimental STD_0 data. In all cases, the central disaccharide moieties GlcNS,6S(C)-GlcA(D) show the greatest involvement in the interactions with the RBD surface. Therefore, the existence of different binding poses (**Figure 40a-c**) that are found in accord with the selected ^{1}H-STD signals indicates that this interaction is characterized by a low degree of specificity, and is found consistent with the conclusions drawn by Parafioriti et al. [192].

➢ *Structural characterization of WT RBD–Hexa complexes in agreement with STD data.*

In a previous study [192], the interaction between hexa and WT-RBD were characterized by performing a single ^{1}H STD NMR experiment (saturation time equals to 3 s). We decide to employ two different approaches: one is based on the cluster analysis of the merged metatrajectory of WT-RBD system in order to define and observe the agreement with the previously acquired STD data. In the second one, we employ STD values for filtering the MD simulation with RedMat, aware of the fact that, since they are not STD$_0$ values, this may result in less accurate filtering. Using the first approach, two representative clusters were identified upon cluster analysis (**Figure 41**).

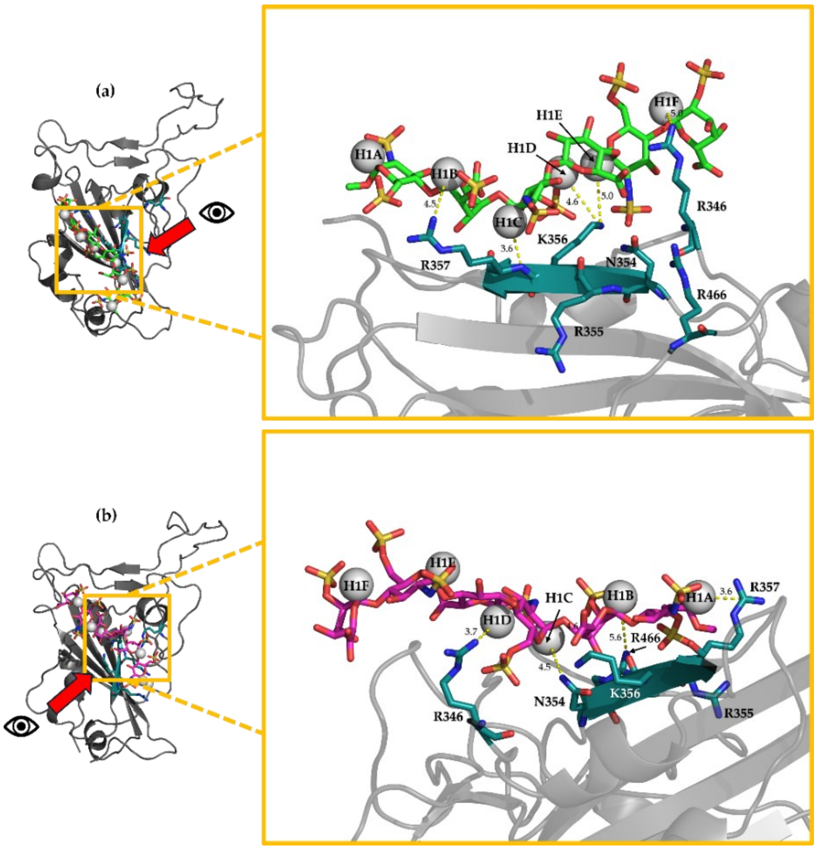


Figure 41 – Graphical representation of the two identified cluster of WT-RBD-hexa complex (panels a and b, respectively). An enlarged view of the binding interface is shown. The hexa molecule is represented as sticks, with carbon atoms colored green and magenta for Cluster I and Cluster II of WT-S1-RBD-hexa, respectively. Red, yellow, and blue indicate oxygen, sulfur, and nitrogen atoms.

The WT RBD protein is illustrated as grey cartoon, and the key interacting residues (R346, N354, R355, K356, R357, and R466) are shown as sticks (carbon in teal, oxygen in red, and nitrogen in blue) and labelled. Anomeric protons are labelled, and the distances to the nearest RBD residues are reported.

The most populated cluster (48%) corresponds to ligand binding at site I (**Figure 41a**). In this binding mode, GlcNS,6S(A) and IdoA2S(B) interact with R357, while GlcNS,6S(C) and GlcA(D) form electrostatic contacts with K356. Moreover, GlcNS,6S(E) and IdoA2S(F) interact with R346 and R345, respectively. The second cluster (20%) features solvent exposure of IdoA2S(F) and GlcNS,6S(E), whereas residues GlcNS,6S(A)–GlcA(D) occupy the upper portion of site I (**Figure 41b**), forming electrostatic interactions with R357/R466, K356/R466, and K356/R346, respectively. These binding arrangements are in good agreement with structural constraints that were found previously [192].

In the second approach, the filtration via RedMat method, identify only a small ensemble of frames (equals to 4 poses) with a R-NOE below the 0.3 (**Figure 42**).

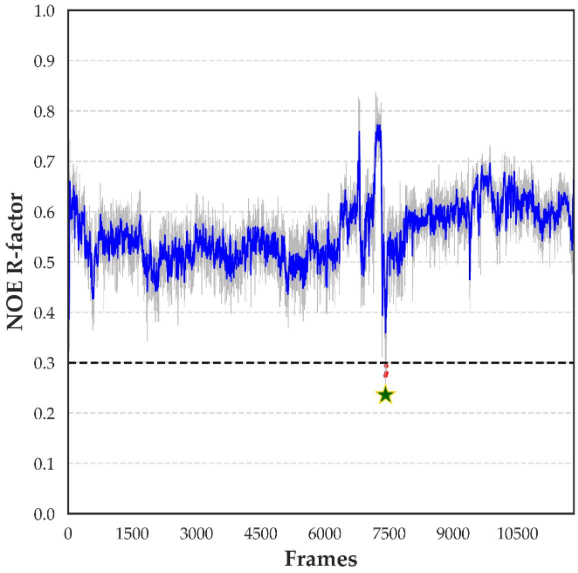


Figure 42 - Evolution of the R-NOE factor in RBD-WT-hexa complex filtered trajectories. The grey line represents the R-NOE values in each frame, while the blue line shows the rolling average (window size = 10). Structure with R-NOE values below 0.3 are marked with red dots. The frame with the lowest R-NOE factor is highlighted with a green star.

In this small ensemble of poses very similar to each other (RMSD < 2Å) , the ligand occupies the upper portion of site I, same region observed for the second cluster determined in the first approach (**Figure 43** and **Figure A. 6**). In this binding mode, the $NH\text{-}SO_3^-$ and $6\text{-}OSO_3^-$ sulfate groups of GlcNS,6S(A) engage R357 and R466, the COO^- carboxylate group of IdoA(B) form a salt bridge with K356. Moreover, the $6\text{-}OSO_3^-$ group f GlcNS,6S(C), the COO^- group of GlcA(D), the $NH\text{-}SO_3^-$ of GlcNS,6S(E) and COO^- of IdoA2S(F) establish a electrostatic interaction with R346, as previously observed in the contact heatmap (**Figure 31c**). Importantly, in all poses determined using both approaches, the distance between the anomeric protons and the protein surface remains below 5 Å. This suggests that efficient saturation transfer from the RBD to the oligosaccharide in the WT–RBD–hexa complex is feasible, consistent with previous STD NMR observations [192].

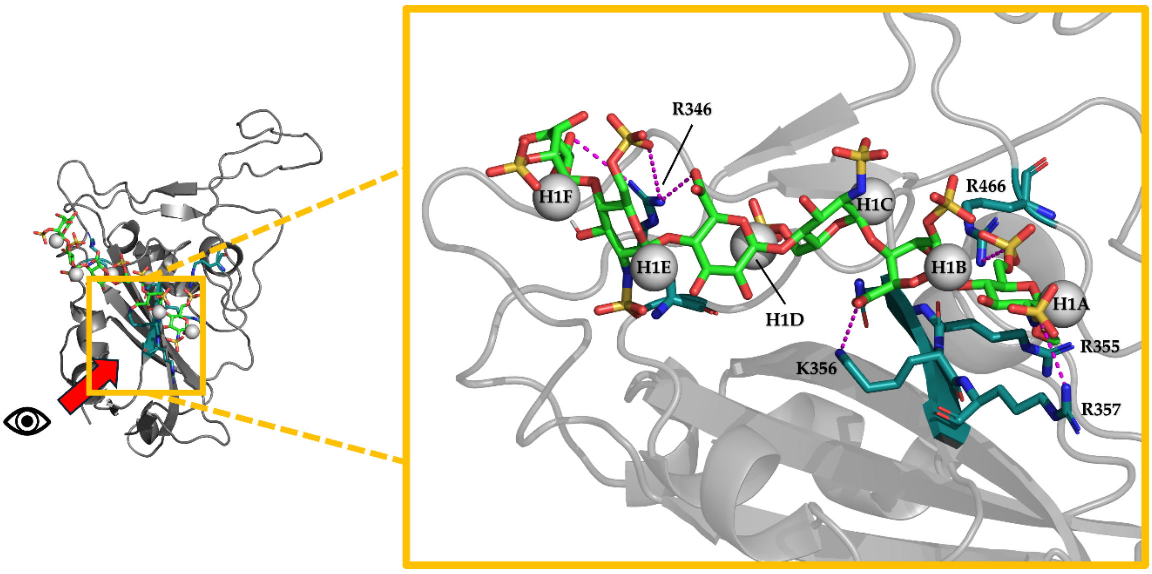


Figure 43 – Graphical representation of binding mode of WT-RBD-hexa complex in. An enlarged view of the binding interface is shown. The hexa molecule is represented as sticks, with carbon, oxygen, sulfur and nitrogen colored in green, red, yellow, and blue, respectively. The WT RBD protein is illustrated as grey cartoon, and the key interacting residues (R346, N354, R355, K356, R357, and R466) are shown as sticks (carbon in teal, oxygen in red, and nitrogen in blue) and labelled. The magenta dashed lined indicates the formed salt bridges. The FA2G2 octasaccharide, linked to the N343 and all other hydrogen atoms are omitted for clarity.

4.6 Conclusions

Heparan sulfate (HS) plays a pivotal role in SARS-CoV-2 infection. It acts as a co-receptor, facilitating the attachment of the viral particles on the host cell surface, promotes the specific interaction between the tip of the spike protein (RBD) and the cell surface hACE2, activating the fusion between the viral and cell membranes, while simultaneously stabilizing the open conformation of the spike protein. This is a prerequisite for the engagement RBD-hACE2 and cell internalization. The evolutionary path that increased the viral infectivity in the latest variants correlates with a higher net positive charge within the spike protein architecture, presumably altering the capacity of RBD to bind HS in the extracellular matrix. Notably, mutations across the variants are concentrated on the ACE2-binding interface, while the structural integrity of site I is preserved.

This thesis characterizes the interactions between hexa and PPS, which are oligosaccharides mimetics of the HS, and the RBD of the SARS-CoV-2 spike protein using docking, molecular dynamics simulations and ^{1}H-STD NMR spectroscopy. The investigation elucidates the epitope binding between the two macromolecular surfaces that become in contact, reproducing the intermolecular forces that acts between them, and showing how an HS oligosaccharide sense and explores the surface of the N343 glycosylated RBD.

Docking calculations revealed systematic enhancement of the estimated binding energy starting from wild-type, followed by Delta and then Omicron variants binding to hexa and PPS (**Figure 22** and **Table 3**). Despite variations in ligand positioning and orientation across variants, the site I (R346, N354, R355, K356, K357, K444, and R466) emerged as the principal HS binding interface of the RBD (**Figure 14**). This site clusters several positive charged (Arg, Lys) and polar residues that are solvent exposed and attracts poly-anionic oligosaccharides through electrostatic and hydrogen bonds interactions. The Omicron variant exhibits particularly intriguing binding behaviour, allowing the hexasaccharide probe to adopt two opposite orientations within site I, showing how HS oligosaccharides potentially engage the site I of RBD using both its two majors ‘sticky' surfaces.

For PPS specifically, docking results reveal a conserved binding mode across all tested RBDs, forming an extended network of electrostatic interactions and hydrogen bonds. This enhanced binding network correlates with increased affinity for sulfated glycans across SARS-CoV-2 variants.

^{1}H-STD-NMR experiments revealed dominant interactions with central moieties of the hexasaccharide probe, particularly GlcNS,6S(C) and GlcA(D), which exhibited the highest intensity of the saturation transfer difference.

Multiple independent molecular dynamics simulations show that the hexasaccharide explored site I of the RBD, undergoing partial detachment, rebinding (e.g. wild-type) alternating rotation, and translation movements (e.g. Omicron) that allow him to probe the site I and the surrounding regions without approaching the ACE2 binding interface. The integration of experimental NMR data with molecular dynamics simulation through the RedMat analysis identified three principal interaction modes between hexa and Omicron RBD, indicating that this interaction is characterized by low specificity, and involve the central disaccharide moiety, while terminal residues displayed greater variability and solvent exposure. A slightly different binding epitope is predicted when hexa binds the wild-type RBD, where cluster analysis of bound frames revealed interactions extending across the entire ligand, with all sugar units positioned within 5 Å of the protein surface.

An important distinction between wild-type and Omicron variants emerges from the different conformational behaviour of the biantennary N343-glycan FA2G2, which correlates with a distinct population of *trans* and *gauche*$^+$ states of the dihedral angle χ_1, that control the orientation of this glycan toward the surface of the RBD. In wild-type RBD, the FA2G2 glycan exhibits higher mobility and propensity to shield the site I. This prevents its interaction with an approaching hexa or affecting an already established contact hexa-RBD. In contrast, the G339D mutation in Omicron RBD confines FA2G2 between helix 366–371 and loop 372–375, stabilizing its *trans* state. Additional point mutations (S371L, S373P, S375F) reinforce the secondary structure of RBD. An unexpected and important observation involves the stabilizing allosteric influence of hexa binding on the distal loop 470-490 region of the RBD, whereby this flexible loop undergoes significant stiffening upon binding, indicating a possible long-range conformational communication that induce a favourable transition toward ACE2-engaged conformations, while simultaneously reducing the entropic penalty associated with primary receptor engagement.

In summary, this study suggests that the tip of the recognition protein (spike) of SARS-CoV-2 potentially binds poly-anions such as HS of the extracellular space trough low specificity multi-valent interactions. These multivalent interactions observed in Omicron and wild-type RBD are reflected by different modes in which HS chains engage the RBD occupying different points and orientation on its surface. Furthermore, the Omicron RBD showing a reduced interference that the FA2G2 glycan exerts on the accessibility of site I, supports its greater affinity for HS oligosaccharides compared to wild-type RBD. Interestingly, this study suggest also that sites on the protein surface that binds HS require solvent exposed positive patches (Arg, Lys, and His at appropriate pH values), that are uninterrupted by negatively charged residues (Asp, Glu), these last introducing repulsive components, that in the sum of intermolecular forces weaken the ability to attract poly-anionic chains.

However, the use of isolated RBD domains and short HS oligosaccharides may overestimate the conformational flexibility of the FA2G2 glycan and oversimplify the distributions of the intermolecular forces driving this recognition. In physiological contexts, the complete S1 subunit of the trimeric spike protein is surrounded by extended heparan sulfate chains at the target cell surface. The binding behaviour in this environment likely differs substantially from that observed in isolated RBD systems, reflecting enhanced cooperativity and multivalent interactions that cannot be directly extrapolated from this model of limited size.

This investigation establishes how the recognition of heparan sulfate by the recognition protein S corresponds to a critical component of SARS-CoV-2 pathogenesis. The findings reveal how viral evolution has systematically optimized HS binding capacity through coordinated sequence mutations and glycosylation remodelling. These changes facilitate enhanced accessibility to cellular co-receptors while preserving ACE2 engagement. The molecular insights gained provide a foundation for rational design of therapeutic interventions based on the inhibition of the early infection events: viral attachment and entry, in this case poly-anions mimetics of HS are suggested as potential anti-viral drugs to investigate further.

Chapter 5. Materials and methods

5.1 Docking calculations

5.1.1 Protein preparation

For the three variants of SARS-CoV-2 (wild-type, Delta and Omicron), the RBD region in the open prefusion conformation (RBD up) in complex with the hACE2 receptor, was taken from Protein Data Bank (PDB) database [212]. In detail, for wild-type and Omicron variant (BA 1.1529 subline), the X-ray structure were selected (PDBid: 6M0J [213] and 7WBP [213] respectively), while for Delta variant (B.1.617.2 subline) a cryo-EM structure (PDBid: 7V8B) was used. Then, for each system the N-glycosylation site at N343 was added. This N-glycan correspond to the FA2G2, a biantennary core-fucosylated octasaccharide, with the formula [GlcNA(α1–6)Fuc](β1–4)GlcNAc(β1–4)Man[(β1–3)Man(α1–6)GlcNAc][(β1–6)Man(α1–6)GlcNAc]. This oligosaccharide, that has been demonstrated to be the most abundant sequence found in all SARS-CoV-2 variants [119], was retrieved from the CHARMM-GUI [214] COVID-19 protein archive [215]. Indeed, this library contains structures deriving from the all-atom modelling and simulation of fully glycosylated full-length homotrimeric SARS-CoV-2 spike protein in open conformation (PDBid: 6VSB [216]) in a viral membrane. Hence, after isolating the simulated glycan structure, it was added to each RBD protein by connecting the glycan reducing end with the ND2 atom of Asn343. Lastly, each glycosylated protein was prepared using the "*Protein preparation*" wizard [217] of Schrödinger-Maestro (release 2023-1). After removal of all the crystallographic water, the hydrogen atoms were added and the disulfide bonds between cysteines were formed. The pKa values for protein residues were calculated using the PROPKA method at pH 7.4 [218, 219]. Hydrogen bonds were subsequently optimized using the exhaustive sampling option. The structures were then relaxed through restrained minimization, converging the heavy atoms to an RMSD of 0.30 Å, employing the OLPS_20051 [220] force field. The final structure was used to generate the grid for docking calculations.

5.1.2 Ligand Preparation

The 3D structure of the hexasaccharide, referred here as to hexa, was retrieved from a previous published work [192]. In detail, the oligosaccharide α-L-IdoA2S(1→4)α-D-GlcNS,6S(1→4)β-D-GlcA(1→4)α-D-GlcNS,6S(1→4)α-l-IdoA2S(1→ 4)α-F-GlcNS,6S-OMe was design in agreement with the experimental heparin structure (PDBid: 1HPN[18, 221]), considering the glucosamine and the glucuronic acid (GlcNS,6S and GlcA, respectively) residues in the $^4C_1$ chair conformation, while the iduronic acid (IdoA2S) in $^1C_4$ chair conformation. The glycosidic dihedral conformation were set as follows: ϕ_1/ψ_1 = 41°/14°, ϕ_2/ψ_2 = -40°/-30°, ϕ_3/ψ_3 = 60°/30°, ϕ_4/ψ_4 = -39°/-33°, ϕ_5/ψ_5 = 41°/14°, in order from the non-reducing to the opposite reducing end as reported previously [192]. The glycosidic dihedral angles are defined as $\phi_i = H1_n \rightarrow C1_n \rightarrow O4_n \rightarrow C4_{n+1}$ and $\psi_i = C1_n \rightarrow O4_n \rightarrow C4_{n+1} \rightarrow H4_{n+1}$]. The net charge of the ligand was -11.

The structure of the hexasaccharide pentosan polysulfate, referred here to as PPS, was generated from a common xylose unit designed in a previous work [130]. Indeed, in the model, all the 2,3-di-sulfo-Xyl residues were set $^4C_1$ chair conformation, as were observed in NOEs spectra (unpublished data) and in agreement with the paper of Schwartz et al. [198]. This conformation leads to the orientation of the sulfated groups in an axial position, thereby minimizing the strong repulsion observed between those charged groups. All the residues are equipped with two sulfate groups at position C2 and C3, except for the non-reducing end monosaccharide, which has an additional O-sulfation at position C4, resulting in a global net charge of -13.

5.1.3 Docking setup

Starting from the glycosylated RBDs, the cubic grid of the receptor was centered on the geometric centroid calculated selecting the residues R346, N354, R355, K356, R357, R466, that correspond to the RBD residue involved in the HS binding. In addition, due to the wide area of the binding site, the dimension of the outer box and the inner box were set at 55 and 25 Å, respectively. The docking calculations were performed applying the standard precision scoring function with the OPLS_2005 [220] force field. To ensure accurate conformational sampling, the ring conformations and nitrogen inversion degree of freedom were not considered, and all φ (phi) and ψ (psi) torsions around glycosidic bonds were kept fixed to the input value. The number of poses generated during the initial docking stage was increased to 10.000, and then 5.000 of these were retained for the subsequent energy minimization step. All other parameters were left at their default settings.

5.1.4 Estimation of the free energy of binding

The analysis of the contacts was performed using the Interaction Fingerprint tool available in Schrödinger [222]. A hydrogen bond was considered present if the distance between the hydrogen bond donor (HBD) and acceptor (HBA) was ≤ 2.8 Å, and if the minimum angles between the donor and acceptor were ≥ 120°. For salt bridges, a maximum distance of 5 Å between oppositely charged groups was used to define a favorable interaction.

5.1.5 Ligand-protein interactions

Contact analysis was performed using the Interaction Fingerprint tool available in Schrödinger [222]. A hydrogen bond was considered present if the distance between the hydrogen bond donor (HBD) and acceptor (HBA) was ≤ 2.8 Å, and if the minimum angles at the donor and acceptor were ≥ 120° and 90°, respectively. For salt bridges, a maximum distance of 5 Å between oppositely charged groups was used to define a favorable interaction.

5.2 All-atom MD simulations

5.2.1 System preparation

Molecular dynamics (MD) simulations were performed on the top-ranked poses that were generated by docking. In the wild-type system, the single obtained pose was employed, whereas two distinct binding modes were selected for the Omicron RBD system. They are characterized by an inverse orientation along the binding site. The Amber18 [223] software was used for the systems preparation , MD simulations and analysis.

The RBD protein and glycans (hexasaccharide ligand and FA2G2 glycan) are described using the Amber (ff14SB) [224] and the GLYCAM-06j-1 [177] force field, respectively. Each system was solvated in cubic water box hedge of 15 Å, using the TIP3P [178] as water model. The net negative charge of the complex was neutralized by adding an equivalent number of Na^+ ions. Then, additional Na_+ and Cl^- were added randomly within the box to mimic the ionic strength of a physiological environment ([NaCl] = 0.15M). A cut-off of 8 Å was applied for non-bonded interactions, and long-range electrostatics were treated using the Particle Mesh Ewald method [225, 226]. Each system underwent energy minimization through three consecutive steps, each consisting of 20 cycles of steepest descent algorithm followed by 880 cycles of conjugate gradient minimization. Harmonic restrains were applied to the atoms of the solute preserving input geometry of hexa-RBD complex, during these two steps ($k = 50\ kcal/mol \cdot Å^2$ and $k = 10\ kcal/mol \cdot Å^2$, respectively), while the final minimization step was performed without restraints to allow full relaxation of the system.

Subsequently, the systems were heated to 300K over 100 ps under NVT conditions using a Langevin thermostat [183] (collision frequency $\gamma = 1\ ps^{-1}$) with a time step of 0.5 fs. Hydrogen atoms were constrained using the SHAKE algorithm [227], and a positional restraint of 10 $kcal/mol \cdot \text{Å}^2$ was applied to solute atoms during this phase. To equilibrate the system density, an initial 100 ps equilibration was carried out in NVT conditions with restrained solute atoms (5 $kcal/mol \cdot \text{Å}^2$), followed by a second equilibration step under NPT conditions using the Berendsen barostat [184] (pressure P = 1 atm, time step t = 0.5 fs, relaxation time τ_P = 2 ps). A final 100 ps equilibration step was performed under NVT conditions without restraints.

After equilibration at T = 300K, production MD simulations were conducted in triplicate under NVT condition to maintain constat the volume of the simulation box. The temperature was controlled using the Langevin thermostat [183]. Each production run lasted 200 ns with a time step of 2 fs. All three replicas were characterized with the same coordinates but with different initial velocities, which sampled randomly from a Maxwell-Boltzmann distribution at 300 K. Coordinates were saved every 20ps, resulting in 10.000 frames per simulation. The three independent trajectories were merged into a single meta-trajectory, which includes a total of 60.000 sampled frames.

Both hexa and PPS hexasaccharide, and the glycosylated WT and Omicron RBDs in their unbound (apo) form were also characterized through three independent MD simulations, following the same protocol previously employed for the hexa complexed with the viral protein.

For hexa, in the first replica, the initial conformation of IdoA2S, GlcA and GlcNS,6S residues, along with the inter-glycosidic dihedral angle, were set to the values described in section 4.1.2. For the other production run, the input geometry was obtained by adding or subtracting 10° to each ϕ_i/ψ_i pair from the initial geometry of the first simulated hexa.

Regarding PPS, in the three independent replicas, the model designed in previously work [130]. For all the systems, all trajectories were merged into a single trajectory of 600 ns, comprising 60,000 sampled geometries for subsequent analysis.

5.2.2 Analysis of the MD simulations

The production MD simulations that involve models of hexa and RBD in unbound state and as hexa RBD complexes, were analyzed using CPPTRAJ toolbox [228] on Amber18 [223]. The root means square deviation (RMSD) and root means square fluctuation (RMSF) were calculated on the backbone atoms of the RBD (C, CA, N, O, H) and the backbone atom of the oligosaccharide (C1, C2, C3,C4, C4,C5, O5, O4) for each independent replica, using the initial geometry and an average geometry as reference for RMSD and RMSF, respectively. The conformation of the oligosaccharide hexa were analyzed, at level of each monosaccharide trough the calculation of the spherical coordinates θ, ϕ, q according to Cremer-Pople approach for six term rings [195]. The corresponding Stoddart diagrams were generated using NumPy (v.2.2.2) [229] and Matplotlib (v.3.10.0) [230] libraries. Furthermore, the conformational analysis was extended to the glycosidic dihedral angles ϕ_i/ψ_i and represented as Ramachandran plots, in which the color gradient is proportional to the density of states, indicate qualitatively the most populated states for each glycosidic linkage. The Ramachandran plots were generated using the seaborn package (v. 0.13.2) [231].

The contact analysis was performed using the *nativecontacts* application in CPPTRAJ, setting a 6 Å cut-off. Contacts were classified as either ionic or polar interactions based on the types of heavy atoms involved. Interactions between a positively charged protein group (Lys and Arg) and a negatively charged ligand group (SO_3^- or COO^-) were identified as electrostatic (ionic) interactions. All other contacts involving polar atoms with an average distance of ≤ 4.5 Å were classified as potential hydrogen bonds or polar interactions.

5.2.3 Distance-base filtering of MD meta-trajectory.

Each meta-trajectory of the hexa–RBD complexes was filtered to select frames in which the oligosaccharide hexa remained in a bound state with the corresponding RBD. The definition of the bound state was guided by the ligand epitope map obtained from the ^{1}H-STD NMR experiments. In this case, the central residues of the hexa, GlcA (D) and GlcNS,6S (C), were observed to have the strongest STD$_0$%, that correlate with the closest to the RBD surface. Therefore, in this description the MD simulated hexa- RBD complex was considered in 'bound state' if the center of mass (CoM) of both GlcA (D) and GlcNS,6S (C) was within 8 Å of the CoM of the nearest amino acid residue on the RBD. This filtering process was performed using the MDAnalysis [229] library (v2.8.0) in Python (v3.10.16).

5.2.4 3D model validation by reduced matrix (RedMat)

RedMat (http://redmat.iiq.us-csic.es/login) compares the experimental STD$_0$ measured upon interaction between hexa and Omicron RBD with the corresponding theoretical values, calculated on the fraction of merged trajectories of Omicron RBD and hexa, yielded after the NMR-based distance filtering. Several attempts were made to determine the correct parameters that provide satisfactory results. The parameters selected for the determination of R-NOE values were a spectrometer frequency at 600 MHz, a correlation time (τ_c) of the complex at 25 ns, and a cut-off distance of 15 Å.

5.2.5 Cluster analysis

The cluster analysis was performed on merged trajectories, after applying specific filtering criteria. In case of Omi-RBD-hexa-I and Omi-RBD-hexa-II complexes, the NMR-base distance criterion, and the R-NOE value below 0.3 were applied. In contrast, for the WT-RBD-hexa complex, the analysis was conducted using only the bound state condition, as no STD$_0$ data are available for this system. The analysis was carried out using CPPTRAJ (v.18.01) [228], considering all carbon, oxygen, sulfur, and nitrogen atoms of the hexa ligand. Clusters were generated using the hierarchical agglomerative approach, with a 10 Å cut-off distance between clusters, where distances were calculated as the average distance between members of two clusters. A cluster was considered representative if its population exceeded 10%.

5.3 ^{1}H-STD NMR experiments

The receptor binding domain (RBD) region of SARS-CoV-2 Omicron variant spike (S) protein, expressed in HEK293 cells, was purchased from GenScript (Netherlands). The protein solution (c= 1.12 mg/ml in phosphate buffer solution (PBS), pH = 7.4) was buffer-exchanged into 20 mM HEPES-d18 deuterated buffer (pH = 7.2, 200 mM NaCl), using VWR® centrifugal filter (10kDa membrane, load capacity = 0.5 mL).

NMR experiments were conducted on a Bruker Avance III spectrometer operating at 600.13 MHz and equipped with a high-sensitivity 5 mm TCI cryoprobe. The ligand was dissolved in 0.2 mL of purified protein solution, and the final mixture was transferred into a 3 mm NMR tube. Final concentration of the protein and ligand were 5.4 μM and 540 μM, respectively, resulting in a protein-to-ligand ratio of 1:100. All spectra were recorded at 293K. Each spectrum was recorded with 2048 scans, 4 dummy scans, and a relation delay d_1 of 6 seconds. A 10 ms spin-lock pulse was applied to suppress the broad protein signals. On- and off-resonance spectra were acquired at 580 Hz (0.96 ppm) and 20.000 Hz (33 ppm), respectively, using interleaved acquisition. Five experiments were performed using different saturation time (t_{sat} = 0.5, 1, 2, 3, 5 sec). ^{1}H-STD NMR spectra were recorded using an edited version of the std.difff3 pulse sequence, in which the residual water HOD signal was suppressed (via presaturation) during gaussian pulse for protein saturation. The difference spectra were obtained by phase cycling and subtraction of the on- and off-resonance datasets. STD intensities at their initial slope (STD_0) were calculated for the anomeric signals, due to their sharpness and resolution. Experimental build-up curves were fitted using the mono-exponential function equation [133] (see **Equation 3** in section 2.8.1). The resulting STD_0 values were expressed as relative percentage, normalized to the most intense signal.

Appendix

	Heparan sulfate (HS)	Heparin (HP)
Sulfate group per hexosamine	0.8-1.8	1.8-2.4
GlcNS content	40-60%	>85%
IdoA content	30-50%	>70%
Antithrombin binding	0-0.3%	≈30%

Table A. 1 – Principal structural differences between heparan sulfate and heparin

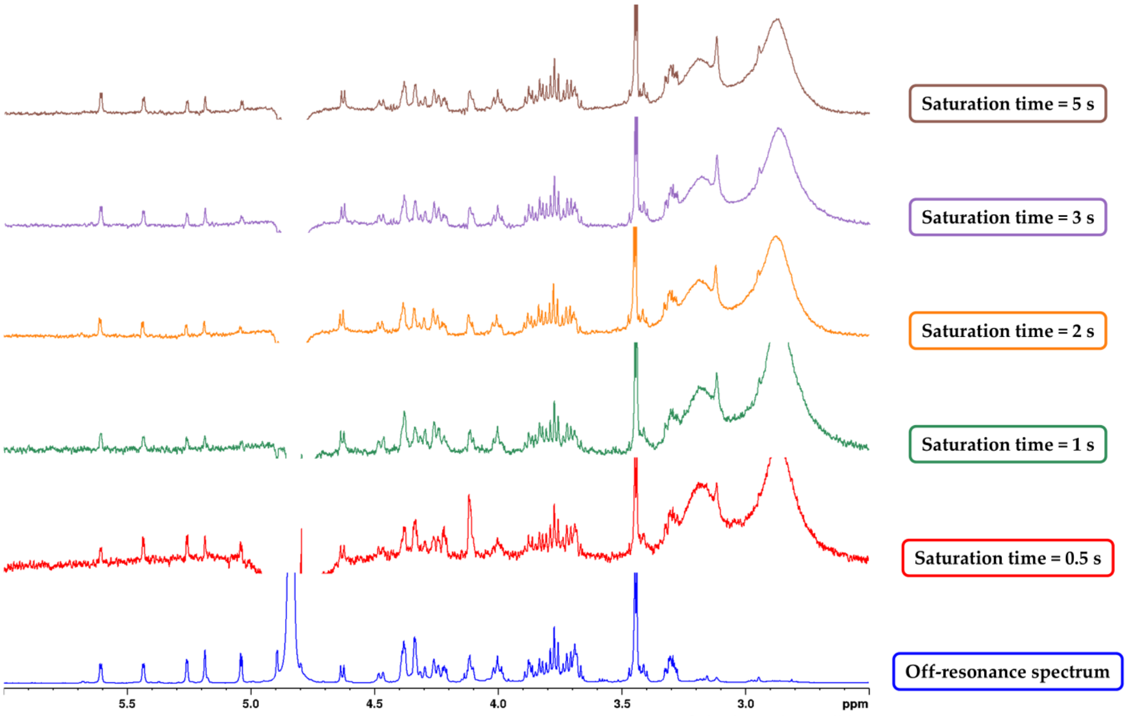


Figure A. 1 - ^{1}H -STD NMR experiments performed on hexa – RBD protein of Omicron variant using different saturation time. In each spectrum, the saturation time is reported on the right.

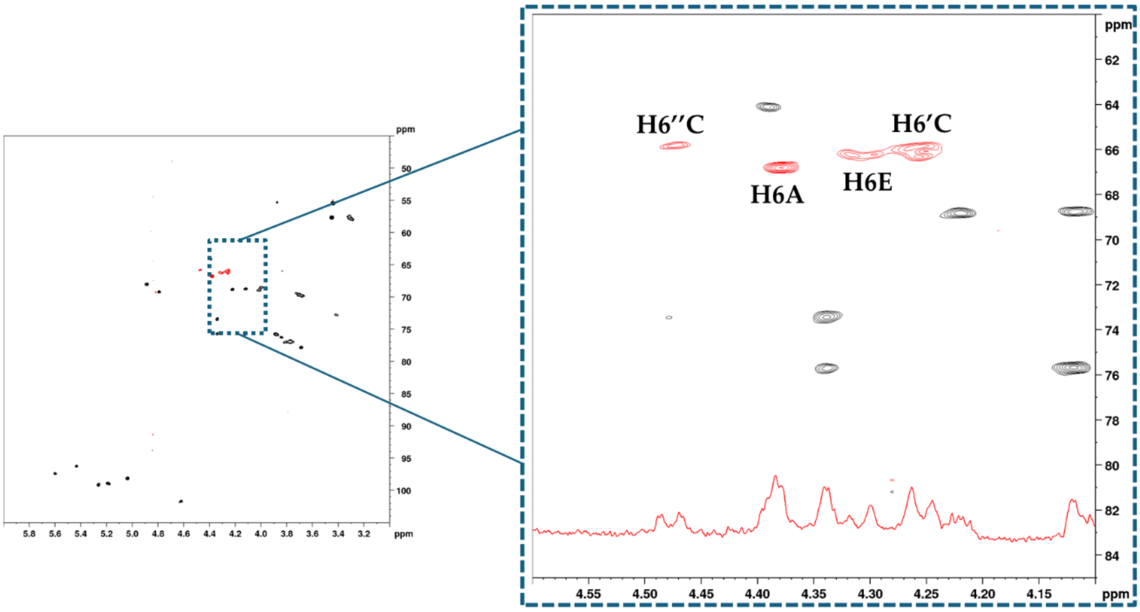


Figure A. 2 - Zoom view of ^{1}H-^{13}C Multiplicity-Edited HSQC of hexa in complex with Omicron RBD protein. The ^{1}H-STD NMR spectrum acquired with a saturation time of 2 s is superimposed on the lower part of the zoomed view. The CH_2.peaks (colored in red) are labelled.

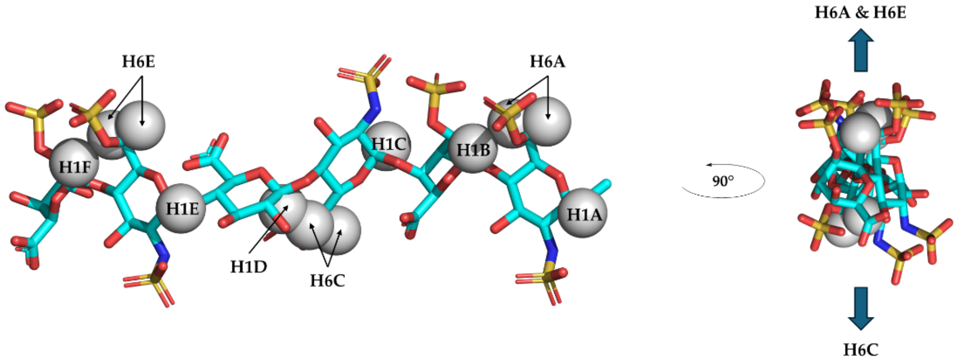


Figure A. 3 - The 3D structure of hexa with the anomeric and C6 protons represented as white spheres.

Table A. 2 - Ring conformers population in the free state of hexa (panel a) and PPS ligand (panel b). Conformations which account for less than 0.5% of the population are omitted, with the resultant percentages then being rounded to the nearest whole number.

(a) **Hexa in free state - Ring conformers (%)**

	$^4C_1$	$^2S_0$	$B_{3,O}$	$^1S_3$	$^1C_4$
GlcNS,6S(A)	100	-	-	-	-
IdoA2S(B)	-	3	-	-	97
GlcNS,6S(C)	100	-	-	-	-
GlcA(D)	94	1	3	1	1

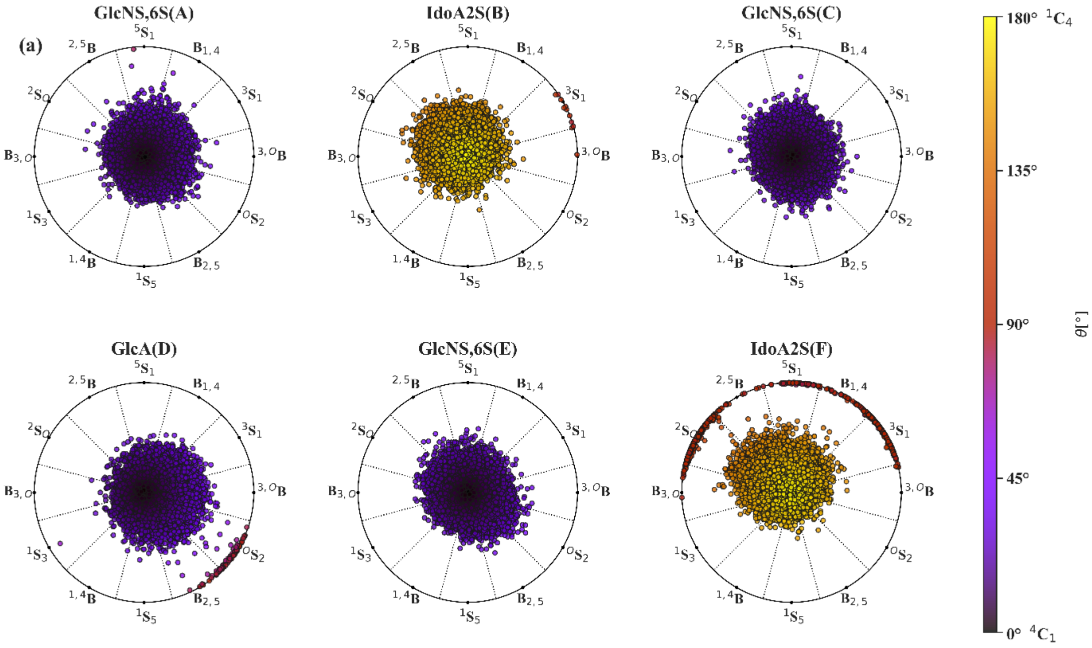
GlcNS,6S(E)	100	-	-	-	-
IdoA2S(F)	-	1	-	-	99

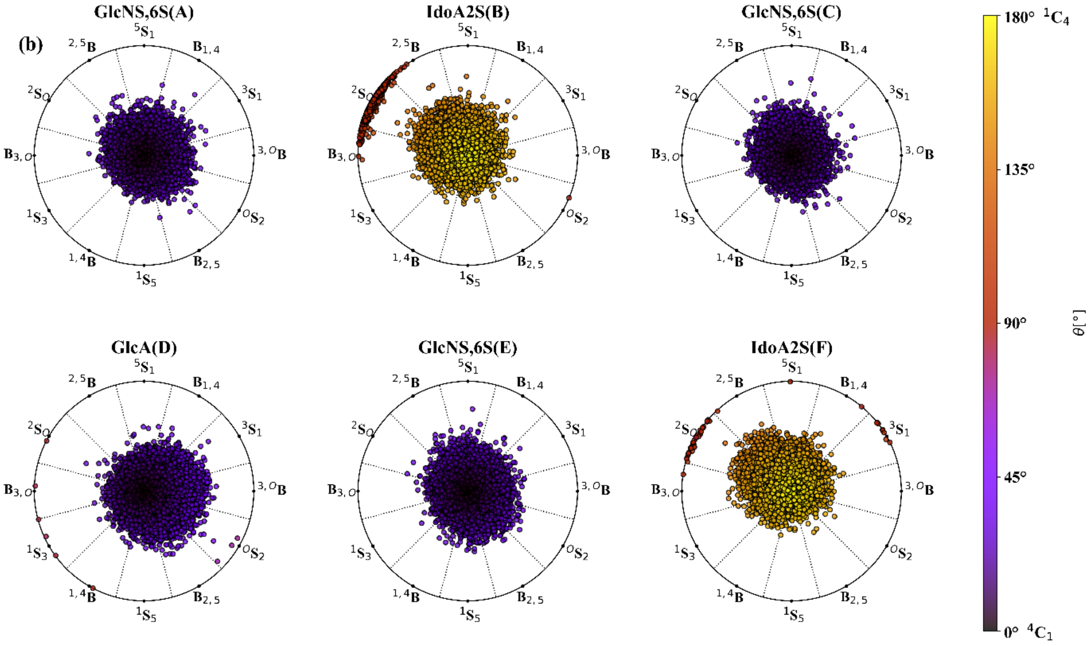
(b) **PPS in free state - Ring conformers (%)**

	${}^{4}C_{1}$	${}^{2}S_{0}$	$B_{3,O}$	${}^{1}S_{3}$	${}^{1}C_{4}$
Xly-1	-	<1	<1	<1	>99
Xly-2	-	<1	<1	<1	>99
Xly-3	-	<1	<1	<1	>99
Xly-4	-	<1	<1	<1	>99
Xly-5	-	<1	<1	<1	>99
Xly-6	-	<1	<1	<1	>99

Table A. 3 – Input and average ϕ_i/ψ_i values of each inter-glycosidic bond pairs (ϕ_i/ψ_i)of the hexa and PPS in the free state.

Dihedral angle	**Input angle hexa (°)**	**Average angle hexa (°)**	**Input angle PPS (°)**	**Average angle PPS (°)**
$\phi_1 - \psi_1$	41 / 14	47 / 8	41 / -14	44 /-24
$\phi_2 - \psi_2$	-40 / -30	-40 / -28	41 / -14	44 / -23
$\phi_3 - \psi_3$	60 / 30	45 / 13	41 / -14	44 / -23
$\phi_4 - \psi_4$	-39 / -33	-35 / -28	41 / -14	44 / -23
$\phi_5 - \psi_5$	41 / 14	45 / -1	41 / -14	44 / -24


(a)
GlcNS,6S(A)
IdoA2S(B)
GlcNS,6S(C)
GlcA(D)
GlcNS,6S(E)
IdoA2S(F)
$^5S_1$
$B_{1,4}$
$^3S_1$
$^{3,O}B$
$^OS_2$
$B_{2,5}$
$^1S_5$
$^{1,4}B$
$^1S_3$
$B_{3,O}$
^2S_O
$^{2,5}B$
180° $^1C_4$
135°
90°
45°
0° $^4C_1$
$\theta[°]$


(b)
GlcNS,6S(A)
IdoA2S(B)
GlcNS,6S(C)
GlcA(D)
GlcNS,6S(E)
IdoA2S(F)
$^5S_1$
$B_{1,4}$
$^3S_1$
$^{3,O}B$
$^OS_2$
$B_{2,5}$
$^1S_5$
$^{1,4}B$
$^1S_3$
$B_{3,O}$
^2S_O
$^{2,5}B$
180° $^1C_4$
135°
90°
45°
0° $^4C_1$
$\theta[°]$

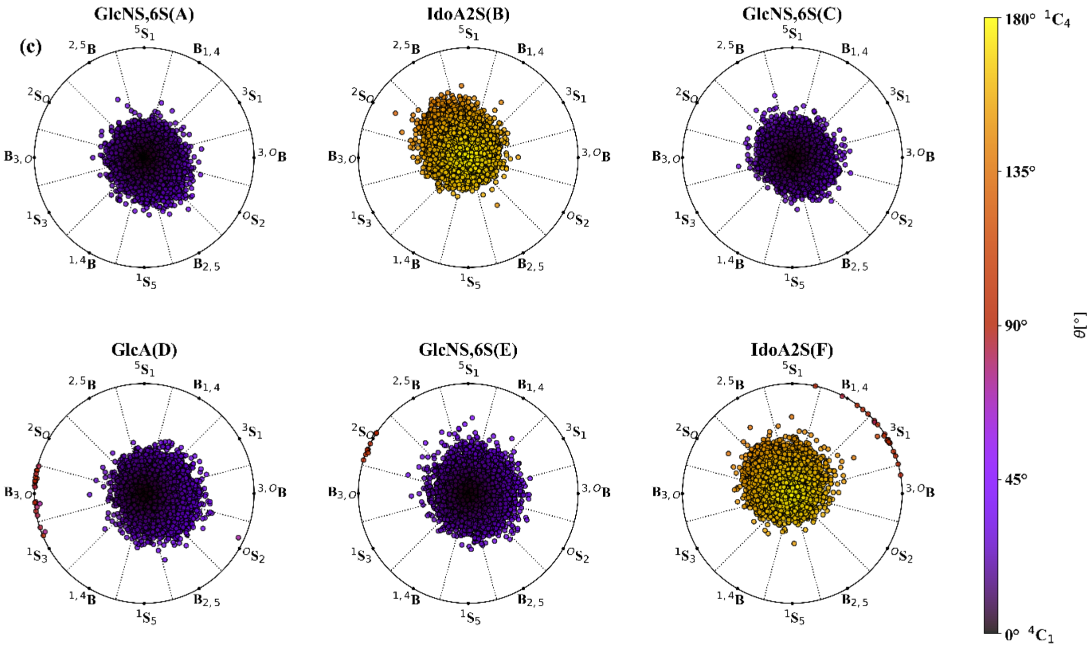


Figure A. 4 – Stoddart diagram of Omi-RBD-hexa-I (panel a), Omi-RBD-hexa-II (panel b) and WT-RBD-hexa (panel c) merged trajectories (600ns), reporting the conformer population of each residue of hexa. The colour code identifies the position of each dot on the corresponding spherical polar presentation: violet $^4C_1$ conformation (θ = 0°), yellow for the $^1C_4$ (θ = 180°) conformation, and red for boat (B) and skew-boat (S) conformation (θ = 90°).

Table A. 4 – Salt bridges (SB) and hydrogen bonds (HB) established in hexa (panel a) and PPS (panel b) in docked poses with the RBD domain of the wild-type, Delta and Omicron variants of SARS-CoV-2. The pair of atoms on which the distance is calculated are bolded and underlined.

(a)

	Protein residue	**Ligand residue**	**Distance**
Wild-type	K444 – $\underline{\mathbf{N}}H_3^+$	IdoA2S(F) – $\underline{\mathbf{C}}OO^-$	4.15Å (SB)
	R346 – $\underline{\mathbf{N}}HCNH_2^+$	IdoA2S(F) – 3-$\underline{\mathbf{O}}H$	2.74Å (HB)
	K356 – $\underline{\mathbf{N}}H_3^+$	GlcNS,6S(A) – 2NH$\underline{\mathbf{S}}O_3^-$	3.81Å (SB)
Delta	R357 – $NH\underline{\mathbf{C}}NH_2^+$	IdoA2S(F) – 2-O$\underline{\mathbf{S}}O_3^-$	4.98Å (SB/HB)
	R466 – $HNC\underline{\mathbf{N}}H_2^+$	IdoA2S(F) – $CO\underline{\mathbf{O}}^-$	3.92Å (SB)
	N354 – $C\underline{\mathbf{N}}H_2$	GlcNS,6S(E) – NH$\underline{\mathbf{S}}O_3^-$	4.4Å (SB)
	R466 – $HN\underline{\mathbf{C}}NH_2^+$	GlcNS,6S(E) – NH$\underline{\mathbf{S}}O_3^-$	3.95Å (SB)
	K356 – $\underline{\mathbf{N}}H_3^+$	GlcA(D) – $CO\underline{\mathbf{O}^-}$	4.98Å (SB)
	R346 – $NH\underline{\mathbf{C}}NH_2^+$	GlcNS,6S(C) – 6-O$\underline{\mathbf{S}}O_3^-$	4.22 Å (SB)
	R346 – $NH\underline{\mathbf{C}}NH_2^+$	IdoA2S(B) - $\mathbf{C}OO^-$	4.49Å (SB/HB)
	T345 – $CH_3C\underline{\mathbf{O}}H$	GlcNS,6S(A) – 3-$\underline{\mathbf{O}}H$	3.10Å (HB)
Omicron Cluster I	R357 – $NH\underline{\mathbf{C}}NH_2^+$	IdoA2S(F) – 2-O$\underline{\mathbf{S}}O_3^-$	4.95Å (SB/HB)
	Y396 – $\underline{\mathbf{O}}H$	IdoA2S(F) – 3-$\underline{\mathbf{O}}H$	2.75Å(HB)
	R357 – $NH\underline{\mathbf{C}}NH_2^+$	GlcNS,6S(E) – 2-NH$\underline{\mathbf{S}}O_3^-$	4.75 Å (SB/HB)
	R466 – $HNC\underline{\mathbf{N}}H_2$	GlcNS,6S(E) – 6-O$\underline{\mathbf{S}}O_3^-$	4.07 Å (SB/HB)

	K356 – $\underline{N}H_3^+$	GlcNS,6S(C) – 6-O$\underline{S}O_3^-$	4.94 Å (SB)
	N354 – C$\underline{N}H_2$	GlcNS,6S(C) – 6-O$\underline{S}O_3^-$	3.41 Å (HB)
	N354 – C$\underline{N}H_2$	IdoA2S(B) - $\underline{C}OO^-$	4.52 Å (HB)
	R346 – NH$\underline{C}NH_2$	GlcNS,6S(A) – 2-NH$\underline{S}O_3^-$	4.72 Å (SB/HB)
Omicron Cluster II	K444-$\underline{N}H_3^+$	IdoA2S(F) - $\underline{C}OO^-$	3.61 Å (SB/HB)
	K444-$\underline{N}H_3^+$	GlcNS,6S(E) – NH$\underline{S}O_3^-$	3.17 Å (HB)
	K444-$\underline{N}H_3^+$	GlcNS,6S(E) – 3-$\underline{O}$H	4.16 Å (SB/HB)
	R346 – NH$\underline{C}NH_2^+$	GlcNS,6S(E) – 2-NH$\underline{S}O_3^-$	5.01 Å (SB/HB)
	R346 – NH$\underline{C}NH_2^+$	GlcNS,6S(C) – 6-O$\underline{S}O_3^-$	4.09 Å (SB)
	R346 – COC$\underline{N}$H	IdoA2S(F) – $\underline{C}OO^-$	4.03 Å (HB)
	K356 – $\underline{N}H_3^+$	GlcNS,6S(A) – 2-NH$\underline{S}O_3^-$	4.06 Å (SB)

(b)

	Protein residue	**Ligand residue**	**Distance**
Wild-type	K462 – $\underline{N}H_3^+$	Xly(A) – 2-O$\underline{S}O_3^-$	3.71Å (HB/SB)
	K462 – $\underline{N}H_3^+$	Xly(A) – 3-O$\underline{S}O_3^-$	3.61Å (HB/SB)
	R355 – NH$\underline{C}NH_2^+$	Xly(B) – 2-O$\underline{S}O_3^-$	4.80Å (HB/SB)
	R466 – HN$\underline{C}NH_2^+$	Xly(C) – 3-O$\underline{S}O_3^-$	4.97Å (HB)
	R466 – HN$\underline{C}NH_2^+$	Xly(D) – 2-O$\underline{S}O_3^-$	4.22Å (SB)
	R357 – $C_\alpha\underline{N}H$	Xly(D) – 3-O$\underline{S}O_3^-$	4.08Å (HB)
	K356 – $\underline{N}H_3^+$	Xly(E) – 2-O$\underline{S}O_3^-$	4.79Å (SB)
	N354 – $\underline{N}H_2$	Xly(F) – 3-O$\underline{S}O_3^-$	3.52Å (HB)
	K356 – $\underline{N}H_3^+$	Xly(F) – 3-O$\underline{S}O_3^-$	5.01Å (SB)
Delta	357 – HN$\underline{C}NH_2^+$	Xly(A) – 2-O$\underline{S}O_3^-$	3.62Å (SB/HB)
	K356 – $\underline{N}H_3^+$	Xly(A) – 4-O$\underline{S}O_3^-$	5.39Å (SB)
	K356 – $\underline{N}H_3^+$	Xly(B) – 2-O$\underline{S}O_3^-$	5.89Å (SB)
	K356 – $\underline{N}H_3^+$	Xly(B) – 3-O$\underline{S}O_3^-$	4.45Å (SB/HB)
	R466 – HN$\underline{C}NH_2^+$	Xly(C) – 2-O$\underline{S}O_3^-$	4.17Å (SB)
	N354 – $\underline{N}H_2$	Xly(C) – 2-O$\underline{S}O_3^-$	3.89Å (HB)
	N354 – $\underline{N}H_2$	Xly(C) – 3-O$\underline{S}O_3^-$	3.39Å (HB)
	R346 – HN$\underline{C}NH_2^+$	Xly(C) – 3-O$\underline{S}O_3^-$	4.14Å (SB/BS)
	R346 – HN$\underline{C}NH_2^+$	Xly(E) – 2-O$\underline{S}O_3^-$	4.72Å (SB)
	R452 – HN$\underline{C}NH_2^+$	Xly(F) – 2-O$\underline{S}O_3^-$	5.13Å (SB)
Omicron	K444-$\underline{N}H_3^+$	Xly(A) – 2-O$\underline{S}O_3^-$	4.04Å (SB/HB)
	K444-$\underline{N}H_3^+$	Xly(A) – 3-O$\underline{S}O_3^-$	3.55Å (SB/HB)
	R346 - HN$\underline{C}NH_2^+$	Xly(A) – 2-O$\underline{S}O_3^-$	4.73Å (SB/HB)
	R346 - HNC$\underline{N}H_2^+$	Xly(A) – 3-O$\underline{S}O_3^-$	4.04Å(SB/HB)
	K444-$\underline{N}H_3^+$	Xly(A) – 4-O$\underline{S}O_3^-$	3.46Å (SB/HB)
	R346 - HN$\underline{C}NH_2^+$	Xly(C) – 3-O$\underline{S}O_3^-$	4.68Å (SB/HB)
	K356-$\underline{N}H_3^+$	Xly(F) – 2-O$\underline{S}O_3^-$	4.42Å (SB)
	R355 - $C_\alpha\underline{N}H$	Xly(F) – 2-O$\underline{S}O_3^-$	4.24Å (HB)
	R466- HN$\underline{C}NH_2^+$	Xly(F) – 3-O$\underline{S}O_3^-$	4.66Å (SB)

Table A. 5 – Ring conformers population percentage of RBD-Omi-hexa-I (panel a), RBD-Omi-hexa-II (panel b), and RBD-WT-hexa (panel c) merged trajectories. Each population was labelled based on the IUPAC nomenclature for ring conformation: the two chairs $^4C_1$ and $^1C_4$ located at the sphere's poles (corresponding to θ = 0° or 180°, respectively), and the 6 boats (B) and skew-boats (S) ring conformation located at the sphere equator (θ=90°). Conformations populated at less than 0.5% are omitted, and the percentages are rounded to the nearest whole number.

(a) RBD-Omi-hexa-I – Ring Conformers (%)

	$^{4}C_{1}$	**$^{2}S_{0}$**	**$^{3}S_{1}$**	**$^{0}S_{2}$**	**$^{1}C_{4}$**
GlcNS,6S(A)	100	-	-	-	-
IdoA2S(B)	-	-	-	-	100
GlcNS,6S(C)	100	-	-	-	-
GlcA(D)	100	-	-	-	-
GlcNS,6S(E)	100	-	-	-	-
IdoA2S(F)	-	1	1	-	98

(b) RBD-Omi-hexa-II – Ring Conformers (%)

	$^{4}C_{1}$	**$^{2}S_{0}$**	**$^{3}S_{1}$**	**$^{0}S_{2}$**	**$^{1}C_{4}$**
GlcNS,6S(A)	100	-	-	-	-
IdoA2S(B)	-	6	-	-	94
GlcNS,6S(C)	100	-	-	-	-
GlcA(D)	100	-	-	-	-
GlcNS,6S(E)	100	-	-	-	-
IdoA2S(F)	-	-	-	-	100

(c) RBD-WT-hexa – Ring Conformers (%)

	$^{4}C_{1}$	**$^{2}S_{0}$**	**$^{3}S_{1}$**	**$^{0}S_{2}$**	**$^{1}C_{4}$**
GlcNS,6S(A)	100	-	-	-	-
IdoA2S(B)	-	-	-	-	100
GlcNS,6S(C)	100	-	-	-	-

GlcA(D)	100	-	-	-	-
GlcNS,6S(E)	100	-	-	-	-
IdoA2S(F)	-	-	-	-	100

Table A. 6 – Most populated ϕ_i/ψ_i state in Omi-RBD-hexa-I, Omi-RBD-hexa-I, and WT-RBD-hexa. Due to the absence of neighboring residue, the flexibility of the terminal glycosidic linkage between IdoA2S(F) and GlcNS,6S(E), and between IdoA2S(B) and GlcNS,6S(A) is greater than the central linkages E-D, D-C, C-B, as shown by wider peaks of ϕ_i/ψ_i distribution and for the presence of alternative weakly populated states.

Dihedral angle	Omi-RBD-hexa-I	Omi-RBD-hexa-II	WT- RBD-hexa
$\phi_1 - \psi_1$	47 / 8	47 / 6	42 / -1
$\phi_2 - \psi_2$	-38 / -30	-40 / -30	-40 / -30
$\phi_3 - \psi_3$	50 / 21	47 / 11	55 / 25
$\phi_4 - \psi_4$	-40 / -28	-35 / -25	-42 / -28
$\phi_5 - \psi_5$	33 / -28	45 / 1	45 / 4

Table A. 7 – Results of contact analysis for Omi-RBD-hexa-I (tables a and b: salt bridges and polar contact), Omi-RBD-hexa-II (tables c and d), and WT-RBD-hexa (tables e and f) filtered trajectories are summarized. Salt bridges interactions were identified based on the proximity between the heavy atoms of the negatively charged group of hexa [C6 in carboxylate group (COO^-) and S1 in the O-sulfate ($-OSO_3^-$) and N-sulfate ($NHSO_3^-$) groups] and the heavy atoms of positively charged residues in the RBD, specifically NZ and NE in lysine and CZ and NE in arginine. A distance cutoff of 6 Å was applied. Hydrogen bond interactions were determined by assessing contacts between the heavy atoms of the negatively charged groups of hexa (C6 in carboxylate group (COO^-) and S1 in the O-sulfate ($-OSO_3^-$) and N-sulfate ($NHSO_3^-$) groups, and O3 in 3-OH groups) and the heavy atoms of RBD residues capable of hydrogen bonding, namely ND2 (Asn), OG and N (Thr), N (Arg, Lys), OH (Tyr), and OG (Ser). A more stringent distance cutoff of 4.5 Å was used. Only interactions with an occurrence frequency of 10% or higher were retained for analysis.

Protein residue	Atom name	Ligand residue	Atom name	Percentage (%)	Distance (Å)	Dev. Std (Å)
R357	CZ	GlcA(D) COO^-	C6	66.5	4.52	0.35
R357	NE	GlcA(D) COO^-	C6	62.2	4.89	0.82
R357	CZ	GlcNS,6S(E) 6 - OSO_3^-	S1	56.2	4.84	0.52
R357	NE	GlcNS,6S(E)	S1	40.6	5.08	0.56

		6 - OSO_3^-				
R357	NE	IdoA2(F) COO^-	C6	40.4	5.11	0.55
R357	CZ	IdoA2(F) 2 - OSO_3^-	S1	39.9	4.88	0.49
R357	CZ	IdoA2(F) COO^-	C6	36.3	4.95	0.63
(a)	NE	GlcNS,6S(E) 2-$NHSO_3^-$	S1	33.7	5.09	0.64
R357	NE	IdoA2(F) 2 - OSO_3^-	S1	33.3	5.48	0.33
K356	NZ	GlcNS,6S(C) 6 - OSO_3^-	S1	32.8	4.50	0.79
R466	CZ	IdoA2S(B) COO^-	C6	26.9	4.33	0.36
R466	NE	IdoA2S(B) COO^-	C6	25.0	5.53	0.19
R346	CZ	GlcNS,6S(A) 6 - OSO_3^-	S1	24.4	4.64	0.47
R357	CZ	GlcNS,6S(E) 2-$NHSO_3^-$	S1	23.8	5.17	0.66
K356	NZ	GlcNS,6S(C) 2-$NHSO_3^-$	S1	20.8	4.45	0.89
R346	NE	GlcNS,6S(A) 2-$NHSO_3^-$	S1	20.0	4.99	0.77
R346	CZ	GlcNS,6S(A) 2-$NHSO_3^-$	S1	19.6	4.75	0.58
R346	NE	GlcNS,6S(A) 6 - OSO_3^-	S1	19.5	4.92	0.83
K356	NZ	IdoA2S(B) - COO^-	C6	19.2	4.66	0.95
K356	NZ	GlcNS,6S(E) 2-$NHSO_3^-$	S1	15.4	4.13	0.73

R357	NE	GlcNS,6S(C) 6 - OSO_3^-	S1	11.8	4.78	0.80

(b)

Protein Residue	**Atom name**	**Ligand residue**	**Atom name**	**Percentage (%)**	**Distance (Å)**	**Dev. Std (Å)**
R357	CZ	IdoA2(F) 3-OH	O3	65.8	4.01	0.63
R357	NE	IdoA2(F) COO^-	O3	63.5	3.63	0.76
N394	ND2	IdoA2(F) COO^-	C6	50.9	3.68	0.41
Y396	OH	IdoA2(F) COO^-	C6	43.7	3.11	0.44
R357	N	GlcNS,6S(C) 6 - OSO_3^-	S1	30.8	3.91	0.27
N394	ND2	IdoA2(F) 2 - OSO_3^-	S1	25.1	4.02	0.23
Y396	OH	GlcNS,6S(E) 6 - OSO_3^-	S1	22.3	3.92	0.55
S359	OG	GlcA(D) COO^-	C6	20.9	3.73	0.33
N354	ND2	IdoA2S(B) COO^-	C6	16.3	3.78	0.28
R357	N	GlcNS,6S(E) 2-$NHSO_3^-$	S1	13.8	4.04	0.17
N394	ND2	GlcNS,6S(E) 6 - OSO_3^-	S1	12.4	3.99	0.26
N360	ND2	GlcNS,6S(C) 2-$NHSO_3^-$	S1	11.8	4.14	0.21
S349	OG	GlcNS,6S(A) 2-$NHSO_3^-$	S1	11.7	3.65	0.22
S349	N	GlcNS,6S(A)	S1	11.5	3.78	0.27

		2-$NHSO_3^-$				
Y396	OH	IdoA2(F) COO^-	C6	10.8	3.70	0.27

(c)

Protein Residue	Atom name	Ligand residue	Atom name	Percentage (%)	Distance (Å)	Dev. Std (Å)
R346	NE	GlcA(D) COO^-	C6	67.7	4.81	0.867
R346	CZ	GlcA(D) COO^-	C6	67.6	4.5	0.652
R346	NE	GlcNS,6S(C) 6 - OSO_3^-	S1	50.6	4.55	0.849
R346	CZ	GlcNS,6S(C) 2 - $NHSO_3^-$	S1	50.3	4.78	0.502
R346	CZ	GlcNS,6S(C) 6 - OSO_3^-	S1	43.7	4.67	0.598
R346	NE	GlcNS,6S(C) 2 - $NHSO_3^-$	S1	41.5	4.96	0.816
K356	NZ	IdoA2S(B) - COO^-	C6	39.6	4.32	0.791
K356	NE	GlcNS,6S(A) 2 - $NHSO_3^-$	S1	29.9	4.1	0.671
K357	CZ	GlcNS,6S(A) 2 - $NHSO_3^-$	S1	26.8	4.47	0.418
K356	NZ	GlcNS,6S(C) 6 - OSO_3^-	S1	25.5	4.47	0.77
K444	NZ	IdoA2(F) 2 - OSO_3^-	S1	21.9	4.21	0.701

R466	CZ	GlcNS,6S(E) 6 - OSO_3^-	S1	20.2	4.51	0.469
K444	NZ	IdoA2(F) COO^-	C6	18	3.94	0.772
R466	NE	GlcNS,6S(E) 6 - OSO_3^-	S1	17.5	5.4	0.472
R355	NE	GlcNS,6S(E) 6 - OSO_3^-	S1	16.1	5.27	0.431
K444	NZ	GlcNS,6S(E) 6 - OSO_3^-	S1	14.6	4.14	0.594
R466	CZ	IdoA2(B) 2 - OSO_3^-	S1	12.1	4.81	0.486
R466	CZ	GlcNS,6S(C) 2 - $NHSO_3^-$	S1	10.6	4.32	0.223
R466	NE	GlcNS,6S(C) 2 - $NHSO_3^-$	S1	10.2	5.57	0.137

(d)

Protein Residue	**Atom name**	**Ligand residue**	**Atom name**	**Percentage (%)**	**Distance (Å)**	**Dev. Std (Å)**
T345	N	GlcNS,6S(C) 6 - OSO_3^-	S1	49.3	3.8	0.3
T345	OG1	GlcNS,6S(C) 6 - OSO_3^-	S1	43.1	3.77	0.29
R346	N	GlcNS,6S(C) 6 - OSO_3^-	S1	39.2	4.15	0.19
A344	N	GlcNS,6S(C) 6 - OSO_3^-	S1	38.2	5.41	0.318
K444	NZ	IdoA2(F)	O3	16.1	4.81	1

		COO^-				
N354	ND2	GlcNS,6S(C) 2 -$NHSO_3^-$	S1	15.5	3.94	0.22
N354	ND2	IdoA2S(B) COO^-	C6	13.8	3.88	0.23

(e)

Protein Residue	**Atom name**	**Ligand residue**	**Atom name**	**Percentage (%)**	**Distance (Å)**	**Dev. Std (Å)**
R346	CZ	GlcA(D) COO^-	C6	84.6	4.63	0.45
R346	NE	GlcA(D) COO^-	C6	79.9	4.87	0.58
K356	NZ	GlcNS,6S(C) 6 - OSO_3^-	S1	72.6	4.30	0.66
R346	NE	GlcNS,6S(E) 2 - $NHSO_3^-$	S1	57.0	5.33	0.54
R357	CZ	IdoA2(B) 2 - OSO_3^-	S1	57.0	4.66	0.28
R357	NE	IdoA2(B) 2 - OSO_3^-	S1	56.9	4.84	0.46
R357	CZ	GlcNS,6S(C) 6 - OSO_3^-	S1	55.8	4.55	0.33
K356	NZ	GlcA(D) COO^-	C6	54.6	3.74	0.46
R357	NE	GlcNS,6S(A) 6 - OSO_3^-	S1	41.7	5.65	0.28
R466	CZ	GlcNS,6S(A) 6 - OSO_3^-	S1	37.5	4.31	0.25

R346	NE	IdoA2(F) COO^-	C6	37.1	5.05	0.64
R346	CZ	IdoA2(F) COO^-	S1	36.3	5.11	0.64
R357	NE	GlcNS,6S(A) 2 - $NHSO_3^-$	S1	35.8	4.08	0.63
R466	NE	GlcNS,6S(A) 6 - OSO_3^-	S1	35.4	5.53	0.27
K356	NZ	GlcNS,6S(E) 2 - $NHSO_3^-$	S1	34.5	4.74	0.69
K356	NZ	IdoA2S(B) COO^-	C6	33.1	4.09	0.68
R357	CZ	GlcNS,6S(A) 2 - $NHSO_3^-$	S1	32.7	4.52	0.41
R355	NE	GlcNS,6S(A) 6 - OSO_3^-	S1	31.1	5.22	0.38
R346	NE	GlcNS,6S(C) 6 - OSO_3^-	S1	30.2	4.10	0.66
R346	CZ	GlcNS,6S(C) 6 - OSO_3^-	S1	29.2	4.47	0.45
R346	CZ	IdoA2(F) COO^-	C6	27.6	5.13	0.63
R355	CZ	GlcNS,6S(A) 6 - OSO_3^-	S1	19.6	5.43	0.44
R346	CZ	GlcNS,6S(E) 6 - OSO_3^-	S1	18.5	4.89	0.56
R466	CZ	IdoA2(B) 2 - OSO_3^-	S1	18.4	4.98	0.49
R466	NE	IdoA2(B) 2 - OSO_3^-	S1	14.0	5.21	0.41

R346	NE	GlcNS,6S(E) 6 - OSO_3^-	S1	10.2	5.38	0.58

(f)

Protein Residue	**Atom name**	**Ligand residue**	**Atom name**	**Percentage (%)**	**Distance (Å)**	**Dev. Std (Å)**
R346	CZ	IdoA2(F) COO^-	O3	58.1	4.24	0.74
R346	NE	IdoA2(F) COO^-	O3	55.0	4.27	0.76
R357	N	GlcNS,6S(C) 2 -$NHSO_3^-$	S1	54.7	4.02	0.13
T345	N	GlcNS,6S(E) 2 - $NHSO_3^-$	S1	48.0	3.81	0.27
R346	N	GlcNS,6S(E) 2- $NHSO_3^-$	S1	42.0	4.09	0.20
T345	OG1	GlcNS,6S(E) 2 - $NHSO_3^-$	S1	32.9	3.84	0.33
N354	ND2	IdoA2S(B) COO^-	C6	27.2	3.80	0.16
T345	OG1	IdoA2(F) COO^-	C6	24.4	3.67	0.43
N360	ND2	GlcNS,6S(A) 6 - OSO_3^-	S1	18.5	3.99	0.30

Table A. 8 – Results of the cluster analysis conducted for Omi-RBD-hexa-I (table a) and Omi-RBD-hexa-II (table b) and WT-RBD-hexa (table c), reporting the identified cluster members and their relative population percentage (perc. %) for the five most populated families For the Omicron RBD – hexa systems, the analysis was restricted to trajectory frames exhibiting an R-NOE factor below 0.3, whereas for the WT-RBD system, only frames from the distance-based filtered trajectory were considered.

(a)

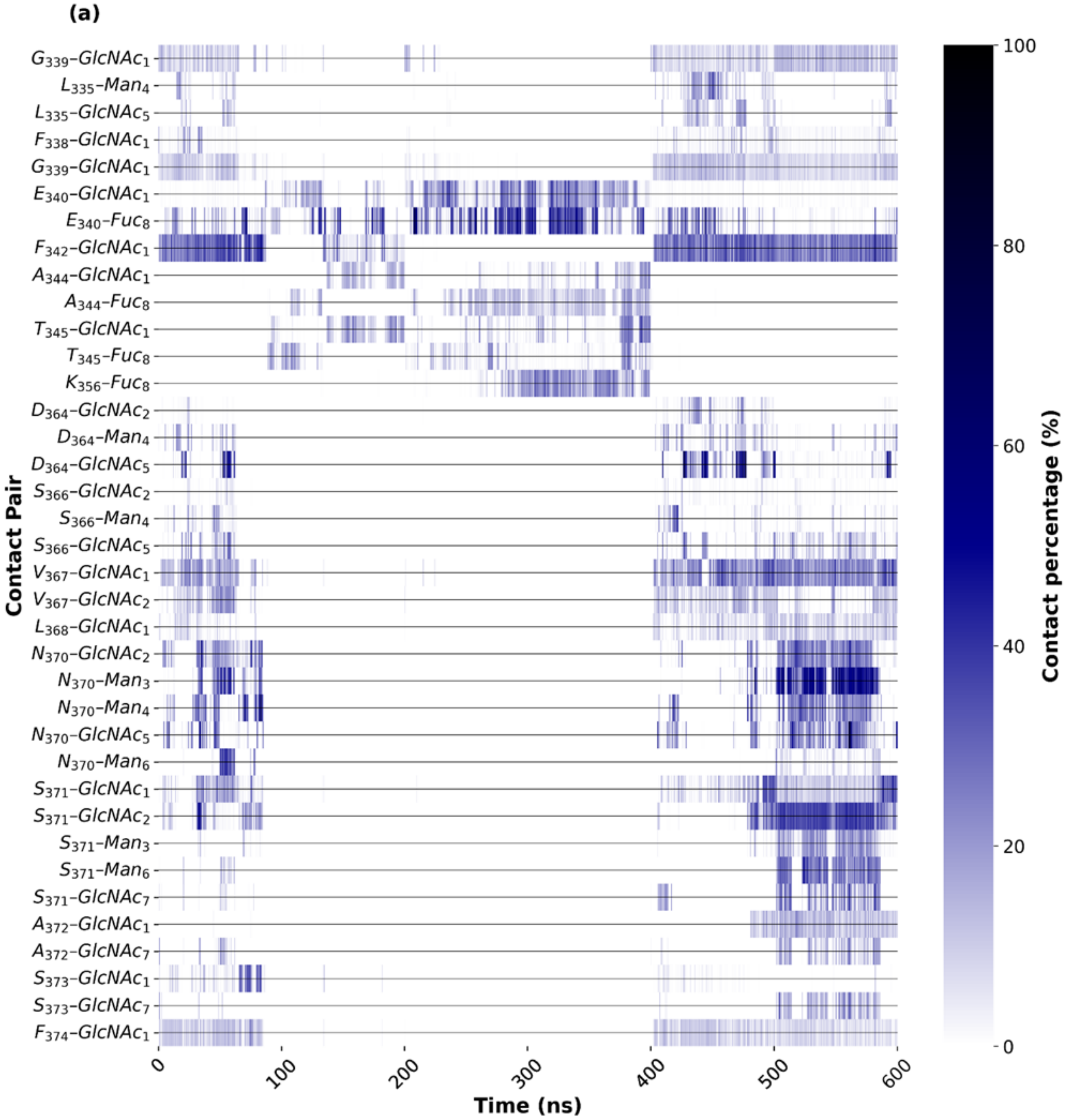
Cluster	**Frames**	**Perc. (%)**
1	2366	43.3
2	2129	39.0
3	439	8.0
4	172	3.2
5	146	2.7

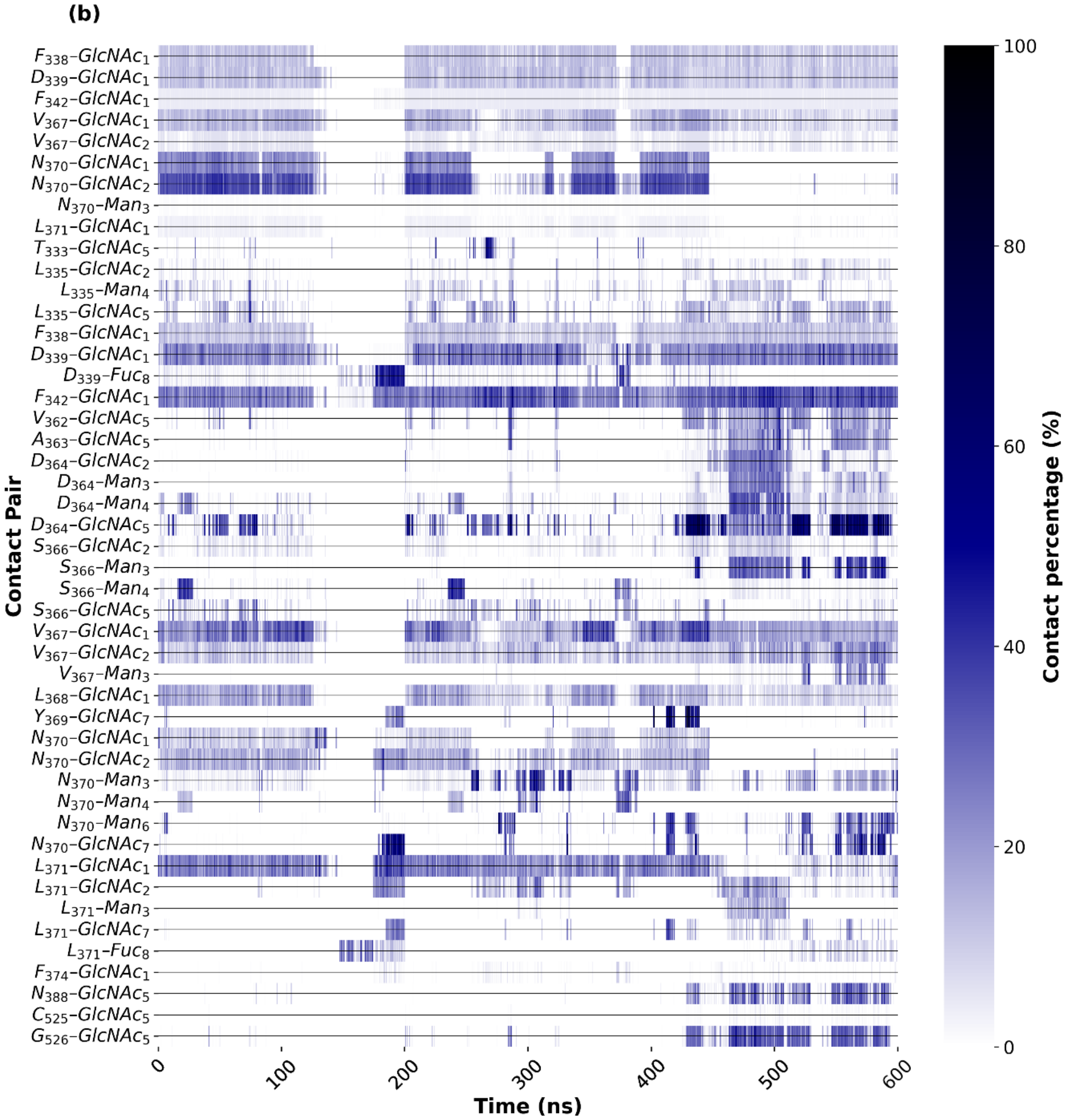
(b)

Cluster	**Frames**	**Perc. (%)**
1	2130	73.0
2	290	9.9
3	137	4.7
4	129	4.4
5	61	2.1

(c)

Cluster	**Frames**	**Perc. (%)**
1	5692	48.2
2	2398	20.3
3	921	7.8
4	574	4.9
5	456	3.9


(a)
G_{339}-$GlcNAc_1$
L_{335}-Man_4
L_{335}-$GlcNAc_5$
F_{338}-$GlcNAc_1$
G_{339}-$GlcNAc_1$
E_{340}-$GlcNAc_1$
E_{340}-Fuc_8
F_{342}-$GlcNAc_1$
A_{344}-$GlcNAc_1$
A_{344}-Fuc_8
T_{345}-$GlcNAc_1$
T_{345}-Fuc_8
K_{356}-Fuc_8
D_{364}-$GlcNAc_2$
D_{364}-Man_4
D_{364}-$GlcNAc_5$
S_{366}-$GlcNAc_2$
S_{366}-Man_4
S_{366}-$GlcNAc_5$
V_{367}-$GlcNAc_1$
V_{367}-$GlcNAc_2$
L_{368}-$GlcNAc_1$
N_{370}-$GlcNAc_2$
N_{370}-Man_3
N_{370}-Man_4
N_{370}-$GlcNAc_5$
N_{370}-Man_6
S_{371}-$GlcNAc_1$
S_{371}-$GlcNAc_2$
S_{371}-Man_3
S_{371}-Man_6
S_{371}-$GlcNAc_7$
A_{372}-$GlcNAc_1$
A_{372}-$GlcNAc_7$
S_{373}-$GlcNAc_1$
S_{373}-$GlcNAc_7$
F_{374}-$GlcNAc_1$
Contact Pair
0
100
200
300
400
500
600
Time (ns)
100
80
60
40
20
0
Contact percentage (%)


(b)
F$_{338}$-GlcNAc$_{1}$
D$_{339}$-GlcNAc$_{1}$
F$_{342}$-GlcNAc$_{1}$
V$_{367}$-GlcNAc$_{1}$
V$_{367}$-GlcNAc$_{2}$
N$_{370}$-GlcNAc$_{1}$
N$_{370}$-GlcNAc$_{2}$
N$_{370}$-Man$_{3}$
L$_{371}$-GlcNAc$_{1}$
T$_{333}$-GlcNAc$_{5}$
L$_{335}$-GlcNAc$_{2}$
L$_{335}$-Man$_{4}$
L$_{335}$-GlcNAc$_{5}$
F$_{338}$-GlcNAc$_{1}$
D$_{339}$-GlcNAc$_{1}$
D$_{339}$-Fuc$_{8}$
F$_{342}$-GlcNAc$_{1}$
V$_{362}$-GlcNAc$_{5}$
A$_{363}$-GlcNAc$_{5}$
D$_{364}$-GlcNAc$_{2}$
D$_{364}$-Man$_{3}$
D$_{364}$-Man$_{4}$
D$_{364}$-GlcNAc$_{5}$
S$_{366}$-GlcNAc$_{2}$
S$_{366}$-Man$_{3}$
S$_{366}$-Man$_{4}$
S$_{366}$-GlcNAc$_{5}$
V$_{367}$-GlcNAc$_{1}$
V$_{367}$-GlcNAc$_{2}$
V$_{367}$-Man$_{3}$
L$_{368}$-GlcNAc$_{1}$
Y$_{369}$-GlcNAc$_{7}$
N$_{370}$-GlcNAc$_{1}$
N$_{370}$-GlcNAc$_{2}$
N$_{370}$-Man$_{3}$
N$_{370}$-Man$_{4}$
N$_{370}$-Man$_{6}$
N$_{370}$-GlcNAc$_{7}$
L$_{371}$-GlcNAc$_{1}$
L$_{371}$-GlcNAc$_{2}$
L$_{371}$-Man$_{3}$
L$_{371}$-GlcNAc$_{7}$
L$_{371}$-Fuc$_{8}$
F$_{374}$-GlcNAc$_{1}$
N$_{388}$-GlcNAc$_{5}$
C$_{525}$-GlcNAc$_{5}$
G$_{526}$-GlcNAc$_{5}$
Contact Pair
0
100
200
300
400
500
600
Time (ns)
100
80
60
40
20
0
Contact percentage (%)

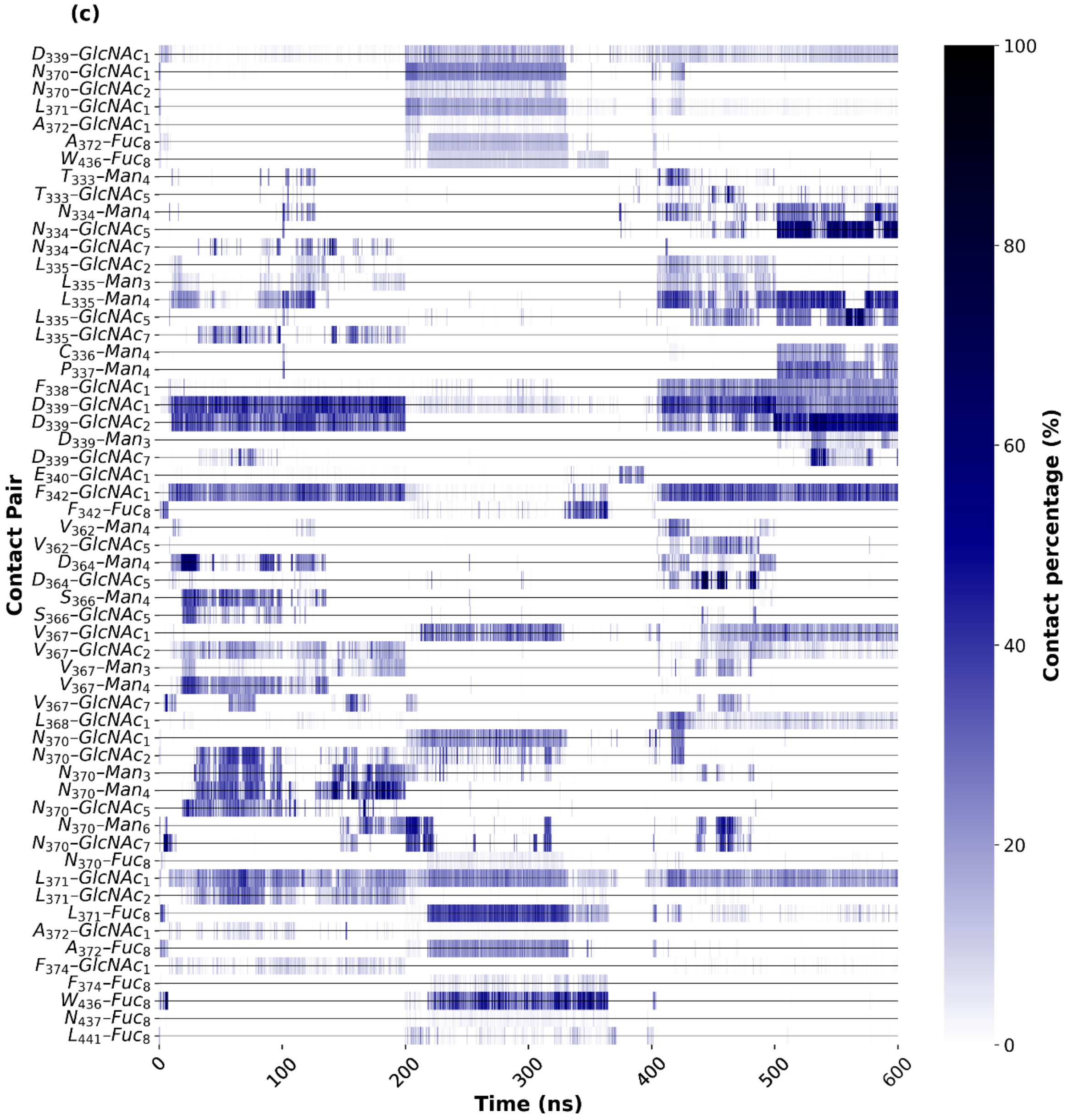


Figure A. 5 – Heatmaps depicting the contact frequencies between FA2G2 and RBD in WT-RBD-hexa (panel a), in the Omi-RBD-hexa-I (panel b) and RBD-Omi-hexa-II (panel c) merged trajectories. For clarity, only contacts with a global occupancy greater than 10%—calculated as the fraction of simulation frames in which a given contact is maintained over the 600 ns trajectory—are displayed in the heatmaps.

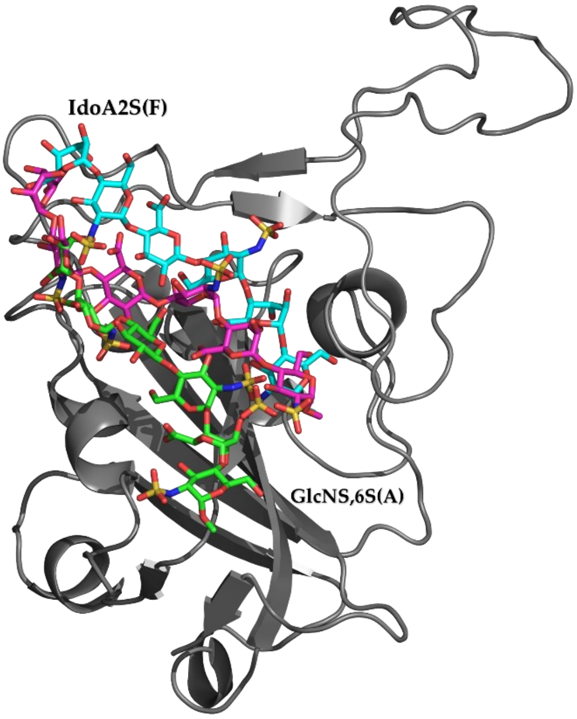


Figure A. 6 – Superimposition of the representative structures from the main clusters of the wild-type RBD and hexa systems (shown in green and cyan, respectively) alongside the WT RBD–hexa complex with the lowest R-NOE factor (0.24; shown in magenta). Oxygen, sulfur, and nitrogen atoms are colored red, yellow, and blue, respectively. Both the non-reducing and reducing ends are indicated. For clarity, the FA2G2 octasaccharide linked to N343 and all hydrogen atoms have been omitted.

Bibliography

[1] S. Perez *et al.*, "Glycosaminoglycans: What Remains To Be Deciphered?," *JACS Au,* vol. 3, no. 3, pp. 628–656, Mar 27 2023, doi: 10.1021/jacsau.2c00569.

[2] E. Jd, K. K, and L. U, "Proteoglycans and Sulfated Glycosaminoglycans," *Essentials of Glycobiology,* pp. 1–14, 2009. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/20301236/.

[3] J. R. Couchman and C. A. Pataki, "An introduction to proteoglycans and their localization," *J Histochem Cytochem,* vol. 60, no. 12, pp. 885–97, Dec 2012, doi: 10.1369/0022155412464638.

[4] T. Inubushi, P. Nag, J. I. Sasaki, Y. Shiraishi, and T. Yamashiro, "The significant role of glycosaminoglycans in tooth development," *Glycobiology,* vol. 34, no. 5, Apr 19 2024, doi: 10.1093/glycob/cwae024.

[5] D. Shi, A. Sheng, and L. Chi, "Glycosaminoglycan-Protein Interactions and Their Roles in Human Disease," *Front Mol Biosci,* vol. 8, p. 639666, 2021, doi: 10.3389/fmolb.2021.639666.

[6] Q. Wang and L. Chi, "The Alterations and Roles of Glycosaminoglycans in Human Diseases," *Polymers (Basel),* vol. 14, no. 22, Nov 18 2022, doi: 10.3390/polym14225014.

[7] S. Yamada, K. Sugahara, and S. Ozbek, "Evolution of glycosaminoglycans: Comparative biochemical study," *Commun Integr Biol,* vol. 4, no. 2, pp. 150–8, Mar 2011, doi: 10.4161/cib.4.2.14547.

[8] B. Mulloy, J. Hogwood, E. Gray, R. Lever, and C. P. Page, "Pharmacology of Heparin and Related Drugs," *Pharmacol Rev,* vol. 68, no. 1, pp. 76–141, Jan 2016, doi: 10.1124/pr.115.011247.

[9] L. Mauri *et al.*, "1D and 2D-HSQC NMR: Two Methods to Distinguish and Characterize Heparin From Different Animal and Tissue Sources," *Front Med (Lausanne),* vol. 6, p. 142, 2019, doi: 10.3389/fmed.2019.00142.

[10] U. Lindahl, G. Backstrom, L. Thunberg, and I. G. Leder, "Evidence for a 3-O-sulfated D-glucosamine residue in the antithrombin-binding sequence of heparin," *Proc Natl Acad Sci U S A,* vol. 77, no. 11, pp. 6551–5, Nov 1980, doi: 10.1073/pnas.77.11.6551.

[11] M. Warda *et al.*, "Isolation and characterization of heparan sulfate from various murine tissues," *Glycoconj J,* vol. 23, no. 7-8, pp. 555–63, Nov 2006, doi: 10.1007/s10719-006-7668-1.

[12] K. J. Murphy, C. L. Merry, M. Lyon, J. E. Thompson, I. S. Roberts, and J. T. Gallagher, "A new model for the domain structure of heparan sulfate based on the novel specificity of K5 lyase," *J Biol Chem,* vol. 279, no. 26, pp. 27239–45, Jun 25 2004, doi: 10.1074/jbc.M401774200.

[13] Z. Shriver, I. Capila, G. Venkataraman, and R. Sasisekharan, "Heparin and heparan sulfate: analyzing structure and microheterogeneity," *Handb Exp Pharmacol,* no. 207, pp. 159–76, 2012, doi: 10.1007/978-3-642-23056-1_8.

[14] J. D. Esko and U. Lindahl, "Molecular diversity of heparan sulfate," *J Clin Invest,* vol. 108, no. 2, pp. 169–73, Jul 2001, doi: 10.1172/JCI13530.

[15] U. Lindahl, M. Kusche, K. Lidholt, and L. G. Oscarsson, "Biosynthesis of heparin and heparan sulfate," *Ann N Y Acad Sci,* vol. 556, no. 1, pp. 36–50, 1989, doi: 10.1111/j.1749-6632.1989.tb22488.x.

[16] J. C. Munoz-Garcia, F. Corzana, J. L. de Paz, J. Angulo, and P. M. Nieto, "Conformations of the iduronate ring in short heparin fragments described by time-averaged distance restrained molecular dynamics," *Glycobiology,* vol. 23, no. 11, pp. 1220–9, Nov 2013, doi: 10.1093/glycob/cwt058.

[17] A. S. Perlin and G. R. Sanderson, "l-iduronic acid, a major constituent of heparin," *Carbohydrate Research,* vol. 12, no. 2, pp. 183–192, 1970, doi: 10.1016/s0008-6215(00)80095-6.

[18] B. Mulloy and M. J. Forster, "Conformation and dynamics of heparin and heparan sulfate," *Glycobiology,* vol. 10, no. 11, pp. 1147–56, Nov 2000, doi: 10.1093/glycob/10.11.1147.

[19] B. Mulloy, M. J. Forster, C. Jones, and D. B. Daviest, "and molecular-modelling studies of the solution conformation of heparin," in "Biochem. J," 1993, vol. 293.

[20] D. R. Ferro *et al.*, "Conformer populations of L-iduronic acid residues in glycosaminoglycan sequences," *Carbohydr Res,* vol. 195, no. 2, pp. 157–67, Jan 15 1990, doi: 10.1016/0008-6215(90)84164-p.

[21] R. El Masri, A. Seffouh, H. Lortat-Jacob, and R. R. Vives, "The "in and out" of glucosamine 6-O-sulfation: the 6th sense of heparan sulfate," *Glycoconj J,* vol. 34, no. 3, pp. 285–298, Jun 2017, doi: 10.1007/s10719-016-9736-5.

[22] S. Sarrazin, W. C. Lamanna, and J. D. Esko, "Heparan sulfate proteoglycans," *Cold Spring Harb Perspect Biol,* vol. 3, no. 7, pp. 1–33, Jul 1 2011, doi: 10.1101/cshperspect.a004952.

[23] V. De Pasquale and L. M. Pavone, "Heparan sulfate proteoglycans: The sweet side of development turns sour in mucopolysaccharidoses," *Biochim Biophys Acta Mol Basis Dis,* vol. 1865, no. 11, p. 165539, Nov 1 2019, doi: 10.1016/j.bbadis.2019.165539.

[24] H. Lortat-Jacob, A. Grosdidier, and A. Imberty, "Structural diversity of heparan sulfate binding domains in chemokines," *Proc Natl Acad Sci U S A,* vol. 99, no. 3, pp. 1229–34, Feb 5 2002, doi: 10.1073/pnas.032497699.

[25] R. S. Aquino, K. Hayashida, A. Hayashida, and P. W. Park, "Role of HSPGs in Systemic Bacterial Infections," *Methods Mol Biol,* vol. 2303, pp. 605–625, 2022, doi: 10.1007/978-1-0716-1398-6_46.

[26] M. A. Gubbiotti, T. Neill, and R. V. Iozzo, "A current view of perlecan in physiology and pathology: A mosaic of functions," *Matrix Biol,* vol. 57-58, pp. 285–298, Jan 2017, doi: 10.1016/j.matbio.2016.09.003.

[27] D. Shukla *et al.*, "A novel role for 3-O-sulfated heparan sulfate in herpes simplex virus 1 entry," *Cell,* vol. 99, no. 1, pp. 13–22, Oct 1 1999, doi: 10.1016/s0092-8674(00)80058-6.

[28] C. Artpradit, L. N. Robinson, B. K. Gavrilov, T. T. Rurak, M. Ruchirawat, and R. Sasisekharan, "Recognition of heparan sulfate by clinical strains of dengue virus serotype 1 using recombinant subviral particles," *Virus Res,* vol. 176, no. 1-2, pp. 69–77, Sep 2013, doi: 10.1016/j.virusres.2013.04.017.

[29] V. Cagno *et al.*, "Highly sulfated K5 Escherichia coli polysaccharide derivatives inhibit respiratory syncytial virus infectivity in cell lines and human tracheal-bronchial histocultures," *Antimicrob Agents Chemother,* vol. 58, no. 8, pp. 4782–94, Aug 2014, doi: 10.1128/AAC.02594-14.

[30] M. Donalisio *et al.*, "Inhibition of human respiratory syncytial virus infectivity by a dendrimeric heparan sulfate-binding peptide," *Antimicrob*

Agents Chemother, vol. 56, no. 10, pp. 5278–88, Oct 2012, doi: 10.1128/AAC.00771-12.

[31] S. M. Johnson *et al.*, "Respiratory Syncytial Virus Uses CX3CR1 as a Receptor on Primary Human Airway Epithelial Cultures," *PLoS Pathog,* vol. 11, no. 12, p. e1005318, Dec 2015, doi: 10.1371/journal.ppat.1005318.

[32] J. Yue *et al.*, "Heparan Sulfate Facilitates Spike Protein-Mediated SARS-CoV-2 Host Cell Invasion and Contributes to Increased Infection of SARS-CoV-2 G614 Mutant and in Lung Cancer," *Front Mol Biosci,* vol. 8, p. 649575, 2021/6// 2021, doi: 10.3389/fmolb.2021.649575.

[33] A. D. Cardin and H. J. Weintraub, "Molecular modeling of protein-glycosaminoglycan interactions," *Arteriosclerosis,* vol. 9, no. 1, pp. 21–32, Jan–Feb 1989, doi: 10.1161/01.atv.9.1.21.

[34] P. Carlsson and L. Kjellen, "Heparin biosynthesis," *Handb Exp Pharmacol,* vol. 207, no. 207, pp. 23–41, 2012, doi: 10.1007/978-3-642-23056-1_2.

[35] D. M. H. Beurskens, J. P. Huckriede, R. Schrijver, H. C. Hemker, C. P. Reutelingsperger, and G. A. F. Nicolaes, "The Anticoagulant and Nonanticoagulant Properties of Heparin," *Thromb Haemost,* vol. 120, no. 10, pp. 1371–1383, Oct 2020, doi: 10.1055/s-0040-1715460.

[36] R. Lever and C. P. Page, "Non-anticoagulant effects of heparin: an overview," *Handb Exp Pharmacol,* vol. 207, no. 207, pp. 281–305, 2012, doi: 10.1007/978-3-642-23056-1_12.

[37] N. M. Kuderer, A. A. Khorana, G. H. Lyman, and C. W. Francis, "A meta-analysis and systematic review of the efficacy and safety of anticoagulants as cancer treatment: impact on survival and bleeding complications," *Cancer,* vol. 110, no. 5, pp. 1149–61, Sep 1 2007, doi: 10.1002/cncr.22892.

[38] D. Sanford, A. Naidu, N. Alizadeh, and A. Lazo-Langner, "The effect of low molecular weight heparin on survival in cancer patients: an updated systematic review and meta-analysis of randomized trials," *J Thromb Haemost,* vol. 12, no. 7, pp. 1076–85, Jul 2014, doi: 10.1111/jth.12595.

[39] J. M. Motta *et al.*, "A low-anticoagulant heparin suppresses metastatic dissemination through the inhibition of tumor cell-platelets association," *Biomed Pharmacother,* vol. 171, p. 116108, Feb 2024, doi: 10.1016/j.biopha.2023.116108.

[40] S. N. Ma *et al.*, "The anti-cancer properties of heparin and its derivatives: a review and prospect," *Cell Adh Migr,* vol. 14, no. 1, pp. 118–128, Dec 2020, doi: 10.1080/19336918.2020.1767489.

[41] L. Borsig, R. Wong, J. Feramisco, D. R. Nadeau, N. M. Varki, and A. Varki, "Heparin and cancer revisited: mechanistic connections involving platelets, P-selectin, carcinoma mucins, and tumor metastasis," *Proc Natl Acad Sci U S A,* vol. 98, no. 6, pp. 3352–7, Mar 13 2001, doi: 10.1073/pnas.061615598.

[42] L. Wang, J. R. Brown, A. Varki, and J. D. Esko, "Heparin's anti-inflammatory effects require glucosamine 6-O-sulfation and are mediated by blockade of L- and P-selectins," *J Clin Invest,* vol. 110, no. 1, pp. 127–36, Jul 2002, doi: 10.1172/JCI14996.

[43] R. P. McEver, "Selectins: initiators of leucocyte adhesion and signalling at the vascular wall," *Cardiovasc Res,* vol. 107, no. 3, pp. 331–9, Aug 1 2015, doi: 10.1093/cvr/cvv154.

[44] L. Litov *et al.*, "Heparin as an Anti-Inflammatory Agent," *bioRxiv,* 2020, doi: 10.1101/2020.07.29.223859.

[45] L. Litov *et al.*, "Molecular Mechanism of the Anti-Inflammatory Action of Heparin," *Int J Mol Sci,* vol. 22, no. 19, Oct 3 2021, doi: 10.3390/ijms221910730.

[46] C. J. Mycroft-West *et al.*, "Heparin Inhibits Cellular Invasion by SARS-CoV-2: Structural Dependence of the Interaction of the Spike S1 Receptor-Binding Domain with Heparin," *Thromb Haemost,* vol. 120, no. 12, pp. 1700–1715, Dec 2020, doi: 10.1055/s-0040-1721319.

[47] S. Y. Kim *et al.*, "Characterization of heparin and severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) spike glycoprotein binding interactions," *Antiviral Res,* vol. 181, p. 104873, Sep 2020, doi: 10.1016/j.antiviral.2020.104873.

[48] C. Mycroft-West *et al.*, "Mycroft-West et al. (2020) Running title: SARS-CoV-2 surface S1 Receptor Binding Domain binds heparin The 2019 coronavirus (SARS-CoV-2) surface protein (Spike) S1 Receptor Binding Domain undergoes conformational change upon heparin binding," 2020, doi: 10.1101/2020.02.29.971093.

[49] Y. Wang, Y. Zhang, P. Wang, T. Jing, Y. Hu, and X. Chen, "Research Progress on Antiviral Activity of Heparin," *Curr Med Chem,* vol. 31, no. 1, pp. 7–24, 2024, doi: 10.2174/0929867330666230203124032.

[50] Y. L. Lin, H. Y. Lei, Y. S. Lin, T. M. Yeh, S. H. Chen, and H. S. Liu, "Heparin inhibits dengue-2 virus infection of five human liver cell lines," *Antiviral Res,* vol. 56, no. 1, pp. 93–6, Oct 2002, doi: 10.1016/s0166-3542(02)00095-5.

[51] C. M. Su, C. L. Liao, Y. L. Lee, and Y. L. Lin, "Highly sulfated forms of heparin sulfate are involved in japanese encephalitis virus infection," *Virology,* vol. 286, no. 1, pp. 206–15, Jul 20 2001, doi: 10.1006/viro.2001.0986.

[52] A. J. Nahmias and S. Kibrick, "Inhibitory effect of heparin on herpes simplex virus," *J Bacteriol,* vol. 87, no. 5, pp. 1060–6, May 1964, doi: 10.1128/jb.87.5.1060-1066.1964.

[53] D. Carstanjen, P. Dutt, and T. Moritz, "Heparin inhibits retrovirus binding to fibronectin as well as retrovirus gene transfer on fibronectin fragments," *J Virol,* vol. 75, no. 13, pp. 6218–22, Jul 2001, doi: 10.1128/JVI.75.13.6218-6222.2001.

[54] Q. Zhang *et al.*, "Heparan sulfate assists SARS-CoV-2 in cell entry and can be targeted by approved drugs in vitro," *Cell Discovery,* vol. 6, no. 1, pp. 1–14, 2020/12// 2020, doi: 10.1038/S41421-020-00222-5;SUBJMETA=1461,221,313,631,80;KWRD=ENDOCYTOSIS,GLYCOBIOLOGY.

[55] L. Ayerbe, C. Risco, and S. Ayis, "The association between treatment with heparin and survival in patients with Covid-19," *J Thromb Thrombolysis,* vol. 50, no. 2, pp. 298–301, Aug 2020, doi: 10.1007/s11239-020-02162-z.

[56] A. L. Gelbach *et al.*, "Interactions between heparin and SARS-CoV-2 spike glycoprotein RBD from omicron and other variants," *Front Mol Biosci,* vol. 9, p. 912887, 2022/8// 2022, doi: 10.3389/fmolb.2022.912887.

[57] T. M. Clausen *et al.*, "SARS-CoV-2 Infection Depends on Cellular Heparan Sulfate and ACE2," *Cell,* vol. 183, no. 4, pp. 1043–1057 e15, Nov 12 2020, doi: 10.1016/j.cell.2020.09.033.

[58] N. Zhu *et al.*, "A Novel Coronavirus from Patients with Pneumonia in China, 2019," *N Engl J Med,* vol. 382, no. 8, pp. 727–733, Feb 20 2020, doi: 10.1056/NEJMoa2001017.

[59] C. Huang *et al.*, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *Lancet,* vol. 395, no. 10223, pp. 497–506, Feb 15 2020, doi: 10.1016/S0140-6736(20)30183-5.

[60] D. Cucinotta and M. Vanelli, "WHO Declares COVID-19 a Pandemic," *Acta Biomed,* vol. 91, no. 1, pp. 157–160, Mar 19 2020, doi: 10.23750/abm.v91i1.9397.

[61] Y. Li and J. Z. Li, "SARS-CoV-2 Virology," *Infect Dis Clin North Am,* vol. 36, no. 2, pp. 251–265, Jun 2022, doi: 10.1016/j.idc.2022.01.004.

[62] "COVID-19 epidemiological update – 24 December 2024." https://www.who.int/publications/m/item/covid-19-epidemiological-update---24-december-2024 (accessed.

[63] Y. Alimohamadi, H. H. Tola, A. Abbasi-Ghahramanloo, M. Janani, and M. Sepandi, "Case fatality rate of COVID-19: a systematic review and meta-analysis," *J Prev Med Hyg,* vol. 62, no. 2, pp. E311–E320, Jun 2021, doi: 10.15167/2421-4248/jpmh2021.62.2.1627.

[64] K. S. Chan *et al.*, "SARS: prognosis, outcome and sequelae," *Respirology,* vol. 8 Suppl, no. Suppl 1, pp. S36–40, Nov 2003, doi: 10.1046/j.1440-1843.2003.00522.x.

[65] F.-O.-W. M. T. W. Group, "MERS: Progress on the global response, remaining challenges and the way forward," *Antiviral Res,* vol. 159, pp. 35–44, Nov 2018, doi: 10.1016/j.antiviral.2018.09.002.

[66] Y. Qu, C. Y. Lee, and K. F. Lam, "A novel method to monitor COVID-19 fatality rate in real-time, a key metric to guide public health policy," *Sci Rep,* vol. 12, no. 1, p. 18277, Oct 31 2022, doi: 10.1038/s41598-022-23138-4.

[67] J. Shang *et al.*, "Structural basis of receptor recognition by SARS-CoV-2," *Nature,* vol. 581, no. 7807, pp. 221–224, May 2020, doi: 10.1038/s41586-020-2179-y.

[68] A. C. Walls, Y. J. Park, M. A. Tortorici, A. Wall, A. T. McGuire, and D. Veesler, "Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein," *Cell,* vol. 181, no. 2, pp. 281–292 e6, Apr 16 2020, doi: 10.1016/j.cell.2020.02.058.

[69] J. Lan *et al.*, "Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor," *Nature,* vol. 581, no. 7807, pp. 215–220, May 2020, doi: 10.1038/s41586-020-2180-5.

[70] J. Zhang, T. Xiao, Y. Cai, and B. Chen, "Structure of SARS-CoV-2 spike protein," *Curr Opin Virol,* vol. 50, pp. 173–182, Oct 2021, doi: 10.1016/j.coviro.2021.08.010.

[71] M. Hoffmann, H. Kleine-Weber, and S. Pohlmann, "A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells," *Mol Cell,* vol. 78, no. 4, pp. 779–784 e5, May 21 2020, doi: 10.1016/j.molcel.2020.04.022.

[72] K. Le, S. Kannappan, T. Kim, J. H. Lee, H. R. Lee, and K. K. Kim, "Structural understanding of SARS-CoV-2 virus entry to host cells," *Front Mol Biosci,* vol. 10, p. 1288686, 2023, doi: 10.3389/fmolb.2023.1288686.

[73] C. B. Jackson, M. Farzan, B. Chen, and H. Choe, "Mechanisms of SARS-CoV-2 entry into cells," *Nat Rev Mol Cell Biol,* vol. 23, no. 1, pp. 3–20, Jan 2022, doi: 10.1038/s41580-021-00418-x.

[74] A. G. Harrison, T. Lin, and P. Wang, "Mechanisms of SARS-CoV-2 Transmission and Pathogenesis," *Trends Immunol,* vol. 41, no. 12, pp. 1100–1115, Dec 2020, doi: 10.1016/j.it.2020.10.004.

[75] R. Sanjuan, M. R. Nebot, N. Chirico, L. M. Mansky, and R. Belshaw, "Viral mutation rates," *J Virol,* vol. 84, no. 19, pp. 9733–48, Oct 2010, doi: 10.1128/JVI.00694-10.

[76] P. Supasa *et al.*, "Reduced neutralization of SARS-CoV-2 B.1.1.7 variant by convalescent and vaccine sera," *Cell,* vol. 184, no. 8, pp. 2201–2211 e7, Apr 15 2021, doi: 10.1016/j.cell.2021.02.033.

[77] D. Zhou *et al.*, "Evidence of escape of SARS-CoV-2 variant B.1.351 from natural and vaccine-induced sera," *Cell,* vol. 184, no. 9, pp. 2348–2361 e6, Apr 29 2021, doi: 10.1016/j.cell.2021.02.037.

[78] W. Dejnirattisai *et al.*, "Antibody evasion by the P.1 strain of SARS-CoV-2," *Cell,* vol. 184, no. 11, pp. 2939–2954 e9, May 27 2021, doi: 10.1016/j.cell.2021.03.055.

[79] C. Liu *et al.*, "Reduced neutralization of SARS-CoV-2 B.1.617 by vaccine and convalescent serum," *Cell,* vol. 184, no. 16, pp. 4220–4236 e13, Aug 5 2021, doi: 10.1016/j.cell.2021.06.020.

[80] W. Dejnirattisai *et al.*, "SARS-CoV-2 Omicron-B.1.1.529 leads to widespread escape from neutralizing antibody responses," *Cell,* vol. 185, no. 3, pp. 467–484 e15, Feb 3 2022, doi: 10.1016/j.cell.2021.12.046.

[81] S. Xue, Y. Han, F. Wu, and Q. Wang, "Mutations in the SARS-CoV-2 spike receptor binding domain and their delicate balance between ACE2 affinity and antibody evasion," *Protein Cell,* vol. 15, no. 6, pp. 403–418, May 28 2024, doi: 10.1093/procel/pwae007.

[82] A. Asif, I. Ilyas, M. Abdullah, S. Sarfraz, M. Mustafa, and A. Mahmood, "The Comparison of Mutational Progression in SARS-CoV-2: A Short Updated Overview," *Journal of Molecular Pathology,* vol. 3, no. 4, pp. 201–218, 2022, doi: 10.3390/jmp3040018.

[83] S. Chakraborty, "E484K and N501Y SARS-CoV 2 spike mutants Increase ACE2 recognition but reduce affinity for neutralizing antibody," *Int Immunopharmacol,* vol. 102, p. 108424, Jan 2022, doi: 10.1016/j.intimp.2021.108424.

[84] B. Luan and T. Huynh, "Insights into SARS-CoV-2's Mutations for Evading Human Antibodies: Sacrifice and Survival," *J Med Chem,* vol. 65, no. 4, pp. 2820–2826, Feb 24 2022, doi: 10.1021/acs.jmedchem.1c00311.

[85] R. Kumar *et al.*, "Understanding Mutations in Human SARS-CoV-2 Spike Glycoprotein: A Systematic Review & Meta-Analysis," *Viruses,* vol. 15, no. 4, Mar 27 2023, doi: 10.3390/v15040856.

[86] S. Cherian *et al.*, "SARS-CoV-2 Spike Mutations, L452R, T478K, E484Q and P681R, in the Second Wave of COVID-19 in Maharashtra, India," *Microorganisms,* vol. 9, no. 7, Jul 20 2021, doi: 10.3390/microorganisms9071542.

[87] S. H. Kim *et al.*, "SARS-CoV-2 evolved variants optimize binding to cellular glycocalyx," *Cell Rep Phys Sci,* vol. 4, no. 4, p. 101346, Apr 19 2023, doi: 10.1016/j.xcrp.2023.101346.

[88] A. L. Alaofi and M. Shahid, "Mutations of SARS-CoV-2 RBD May Alter Its Molecular Structure to Improve Its Infection Efficiency," *Biomolecules,* vol. 11, no. 9, Aug 25 2021, doi: 10.3390/biom11091273.

[89] G. Rucker, H. Qin, and L. Zhang, "Structure, dynamics and free energy studies on the effect of point mutations on SARS-CoV-2 spike protein binding with ACE2 receptor," *PLoS One,* vol. 18, no. 10, p. e0289432, 2023, doi: 10.1371/journal.pone.0289432.

[90] K. Solanki *et al.*, "Structural analysis of spike proteins from SARS-CoV-2 variants of concern highlighting their functional alterations," *Future Virol,* vol. 17, no. 10, Jul 2022, doi: 10.2217/fvl-2022-0003.

[91] V. Durmaz *et al.*, "Structural bioinformatics analysis of SARS-CoV-2 variants reveals higher hACE2 receptor binding affinity for Omicron B.1.1.529 spike

RBD compared to wild type reference," *Sci Rep,* vol. 12, no. 1, p. 14534, Aug 25 2022, doi: 10.1038/s41598-022-18507-y.

[92] C. H. S. da Costa, C. A. B. de Freitas, C. N. Alves, and J. Lameira, "Assessment of mutations on RBD in the Spike protein of SARS-CoV-2 Alpha, Delta and Omicron variants," *Sci Rep,* vol. 12, no. 1, p. 8540, May 20 2022, doi: 10.1038/s41598-022-12479-9.

[93] T. Koley, A. Goswami, M. Kumar, N. Upadhyay, and G. Hariprasad, "Comparative Structural Analysis of Human ACE2 Receptor with Spike Protein of SARS-CoV-2 Variants: Implications to Understand Infectivity of the Virus," *Adv Appl Bioinform Chem,* vol. 15, pp. 21–27, 2022, doi: 10.2147/AABC.S360787.

[94] K. C. Chan, Y. Song, Z. Xu, C. Shang, and R. Zhou, "SARS-CoV-2 Delta Variant: Interplay between Individual Mutations and Their Allosteric Synergy," *Biomolecules,* vol. 12, no. 12, Nov 23 2022, doi: 10.3390/biom12121742.

[95] A. Khan, S. A. Khan, K. Zia, M. S. Altowyan, A. Barakat, and Z. Ul-Haq, "Deciphering the Impact of Mutations on the Binding Efficacy of SARS-CoV-2 Omicron and Delta Variants With Human ACE2 Receptor," *Front Chem,* vol. 10, p. 892093, 2022, doi: 10.3389/fchem.2022.892093.

[96] S. P. Ju, Y. C. Yang, and H. Y. Chen, "Unraveling the binding mechanisms of SARS-CoV-2 variants through molecular simulations," *Heliyon,* vol. 10, no. 5, p. e27193, Mar 15 2024, doi: 10.1016/j.heliyon.2024.e27193.

[97] M. Valerio, L. Borges-Araujo, M. N. Melo, D. Lousa, and C. M. Soares, "SARS-CoV-2 variants impact RBD conformational dynamics and ACE2 accessibility," *Front Med Technol,* vol. 4, p. 1009451, 2022, doi: 10.3389/fmedt.2022.1009451.

[98] H. Hever, Z. Darula, and K. F. Medzihradszky, "Characterization of Site-Specific N-Glycosylation," *Methods Mol Biol,* vol. 1934, pp. 93–125, 2019, doi: 10.1007/978-1-4939-9055-9_8.

[99] A. Shajahan, L. Pepi, B. Kumar, N. Murray, and P. Azadi, "Site Specific N- and O-glycosylation mapping of the Spike Proteins of SARS-CoV-2 Variants of Concern," *Res Sq,* Nov 16 2022, doi: 10.21203/rs.3.rs-2188138/v1.

[100] A. Varki *et al.*, "Symbol Nomenclature for Graphical Representations of Glycans," *Glycobiology,* vol. 25, no. 12, pp. 1323–4, Dec 2015, doi: 10.1093/glycob/cwv091.

[101] S. Neelamegham *et al.*, "Updates to the Symbol Nomenclature for Glycans guidelines," *Glycobiology,* vol. 29, no. 9, pp. 620–624, Aug 20 2019, doi: 10.1093/glycob/cwz045.

[102] C. Steentoft *et al.*, "Mining the O-glycoproteome using zinc-finger nuclease-glycoengineered SimpleCell lines," *Nat Methods,* vol. 8, no. 11, pp. 977–82, Oct 9 2011, doi: 10.1038/nmeth.1731.

[103] M. He, X. Zhou, and X. Wang, "Glycosylation: mechanisms, biological functions and clinical implications," *Signal Transduct Target Ther,* vol. 9, no. 1, p. 194, Aug 5 2024, doi: 10.1038/s41392-024-01886-1.

[104] S. S. Pinho and C. A. Reis, "Glycosylation in cancer: mechanisms and clinical implications," *Nat Rev Cancer,* vol. 15, no. 9, pp. 540–55, Sep 2015, doi: 10.1038/nrc3982.

[105] X. Pang, H. Li, F. Guan, and X. Li, "Multiple Roles of Glycans in Hematological Malignancies," *Front Oncol,* vol. 8, p. 364, 2018, doi: 10.3389/fonc.2018.00364.

[106] M. Zabczynska, P. Link-Lenczowski, and E. Pochec, "Glycosylation in Autoimmune Diseases," *Adv Exp Med Biol,* vol. 1325, pp. 205–218, 2021, doi: 10.1007/978-3-030-70115-4_10.

[107] L. Casalino *et al.*, "Beyond shielding: The roles of glycans in the SARS-CoV-2 spike protein," *ACS Central Science,* vol. 6, no. 10, pp. 1722–1734, 2020/10// 2020, doi: 10.1021/ACSCENTSCI.0C01056/SUPPL_FILE/OC0C01056_SI_006.ZIP.

[108] Y. Watanabe, J. D. Allen, D. Wrapp, J. S. McLellan, and M. Crispin, "Site-specific glycan analysis of the SARS-CoV-2 spike," *Science,* vol. 369, no. 6501, pp. 330–333, Jul 17 2020, doi: 10.1126/science.abb9983.

[109] Y. Watanabe *et al.*, "Vulnerabilities in coronavirus glycan shields despite extensive glycosylation," *Nat Commun,* vol. 11, no. 1, p. 2688, May 27 2020, doi: 10.1038/s41467-020-16567-0.

[110] C. M. Ives *et al.*, "Role of N343 glycosylation on the SARS-CoV-2 S RBD structure and co-receptor binding across variants of concern," *Elife,* vol. 13, Jun 12 2024, doi: 10.7554/eLife.95708.

[111] M. Mandalari *et al.*, "Interaction Between Heparan Sulfate Oligosaccharide and the Receptor-Binding Domain of the Wild-Type and Omicron Variant of the SARS-CoV-2 Spike Protein," *Biomolecules,* vol. 15, no. 9, Sep 19 2025, doi: 10.3390/biom15091343.

[112] H. Wilkinson and R. Saldova, "Current Methods for the Characterization of O-Glycans," *J Proteome Res,* vol. 19, no. 10, pp. 3890–3905, Oct 2 2020, doi: 10.1021/acs.jproteome.0c00435.

[113] J. Kim *et al.*, "Structural and Quantitative Characterization of Mucin-Type O-Glycans and the Identification of O-Glycosylation Sites in Bovine Submaxillary Mucin," *Biomolecules,* vol. 10, no. 4, Apr 20 2020, doi: 10.3390/biom10040636.

[114] S. Onigbinde *et al.*, "Variations in O-Glycosylation Patterns Influence Viral Pathogenicity, Infectivity, and Transmissibility in SARS-CoV-2 Variants," *Biomolecules,* vol. 13, no. 10, Sep 29 2023, doi: 10.3390/biom13101467.

[115] A. Shajahan, N. T. Supekar, A. S. Gleinich, and P. Azadi, "Deducing the N- and O-glycosylation profile of the spike protein of novel coronavirus SARS-

CoV-2," *Glycobiology,* vol. 30, no. 12, pp. 981–988, Dec 9 2020, doi: 10.1093/glycob/cwaa042.

[116] C. F. S. Eldrid, J. D. Allen, M. L. Newby, and M. Crispin, "Suppression of O-Linked Glycosylation of the SARS-CoV-2 Spike by Quaternary Structural Restraints," *Anal Chem,* vol. 93, no. 43, pp. 14392–14400, Nov 2 2021, doi: 10.1021/acs.analchem.1c01772.

[117] T. Sztain *et al.*, "A glycan gate controls opening of the SARS-CoV-2 spike protein," *Nature Chemistry 2021 13:10,* vol. 13, no. 10, 2021–08–19, doi: 10.1038/s41557-021-00758-3.

[118] F. Zhang *et al.*, "SARS-CoV-2 spike glycosylation affects function and neutralization sensitivity," *mBio,* vol. 15, no. 2, p. e0167223, Feb 14 2024, doi: 10.1128/mbio.01672-23.

[119] M. Sanda, L. Morrison, and R. Goldman, "N- and O-Glycosylation of the SARS-CoV-2 Spike Protein," *Anal Chem,* vol. 93, no. 4, pp. 2003–2009, Feb 2 2021, doi: 10.1021/acs.analchem.0c03173.

[120] G. Paiardi, S. Richter, P. Oreste, C. Urbinati, M. Rusnati, and R. C. Wade, "The binding of heparin to spike glycoprotein inhibits SARS-CoV-2 infection by three mechanisms," *J Biol Chem,* vol. 298, no. 2, p. 101507, Feb 2022, doi: 10.1016/j.jbc.2021.101507.

[121] E. Vicenzi *et al.*, "Coronaviridae and SARS-associated coronavirus strain HSR1," *Emerg Infect Dis,* vol. 10, no. 3, pp. 413–8, Mar 2004, doi: 10.3201/eid1003.030683.

[122] E. Vicenzi *et al.*, "Subverting the mechanisms of cell death: flavivirus manipulation of host cell responses to infection," *Biochem Soc Trans,* vol. 46, no. 3, pp. 609–617, Jun 19 2018, doi: 10.1042/BST20170399.

[123] S. Ghezzi *et al.*, "Heparin prevents Zika virus induced-cytopathic effects in human neural progenitor cells," *Antiviral Res,* vol. 140, pp. 13–17, Apr 2017, doi: 10.1016/j.antiviral.2016.12.023.

[124] D. WuDunn and P. G. Spear, "Initial interaction of herpes simplex virus with cells is binding to heparan sulfate," *J Virol,* vol. 63, no. 1, pp. 52–8, Jan 1989, doi: 10.1128/JVI.63.1.52-58.1989.

[125] V. De Pasquale, M. S. Quiccione, S. Tafuri, L. Avallone, and L. M. Pavone, "Heparan Sulfate Proteoglycans in Viral Infection and Treatment: A Special Focus on SARS-CoV-2," *Int J Mol Sci,* vol. 22, no. 12, Jun 18 2021, doi: 10.3390/ijms22126574.

[126] L. Liu *et al.*, "Heparan Sulfate Proteoglycans as Attachment Factor for SARS-CoV-2," *ACS Central Science,* vol. 7, no. 6, pp. 1009–1018, 2021/6// 2021, doi: 10.1021/ACSCENTSCI.1C00010/SUPPL_FILE/OC1C00010_SI_001.PDF.

[127] F. L. Kearns *et al.*, "Spike-heparan sulfate interactions in SARS-CoV-2 infection," *Curr Opin Struct Biol,* vol. 76, p. 102439, Oct 2022, doi: 10.1016/j.sbi.2022.102439.

[128] X. Wang, L. Bie, and J. Gao, "Structural Insights into the Cofactor Role of Heparin/Heparan Sulfate in Binding between the SARS-CoV-2 Spike Protein and Host Angiotensin-Converting Enzyme II," *J Chem Inf Model,* vol. 62, no. 3, pp. 656–667, Feb 14 2022, doi: 10.1021/acs.jcim.1c01484.

[129] A. Alekseeva *et al.*, "In-depth structural characterization of pentosan polysulfate sodium complex drug using orthogonal analytical tools," *Carbohydr Polym,* vol. 234, p. 115913, Apr 15 2020, doi: 10.1016/j.carbpol.2020.115913.

[130] S. Bertini *et al.*, "Pentosan Polysulfate Inhibits Attachment and Infection by SARS-CoV-2 In Vitro: Insights into Structural Requirements for Binding," *Thromb Haemost,* vol. 122, no. 6, pp. 984–997, Jun 2022, doi: 10.1055/a-1807-0168.

[131] R. Nepravishta *et al.*, "Fast Quantitative Validation of 3D Models of Low-Affinity Protein-Ligand Complexes by STD NMR Spectroscopy," *J Med Chem,* vol. 67, no. 12, pp. 10025–10034, Jun 27 2024, doi: 10.1021/acs.jmedchem.4c00204.

[132] A. Viegas, J. Manso, F. L. Nobrega, and E. J. Cabrita, "Saturation-Transfer Difference (STD) NMR: A Simple and Fast Method for Ligand Screening and Characterization of Protein Binding," April 18, 2011, doi: 10.1021/ed101169.

[133] B. Meyer and T. Peters, "NMR Spectroscopy of Proteins NMR Spectroscopy Techniques for Screening and Identifying Ligand Binding to Protein Receptors Angewandte Chemie Keywords: drug design · ligand-protein interactions · NMR spectroscopy · proteins · screening."

[134] M. Mayer and T. L. James, "NMR-Based Characterization of Phenothiazines as a RNA Binding Scaffoldt," *Journal of the American Chemical Society,* vol. 126, no. 13, pp. 4453–4460, 2004/4// 2004, doi: 10.1021/JA0398870/ASSET/IMAGES/LARGE/JA0398870F00008.JPEG.

[135] A. Frankish *et al.*, "Gencode 2021," *Nucleic Acids Res,* vol. 49, no. D1, pp. D916–D923, Jan 8 2021, doi: 10.1093/nar/gkaa1087.

[136] T. I. Oprea *et al.*, "Unexplored therapeutic opportunities in the human genome," *Nature Reviews Drug Discovery,* vol. 17, no. 5, pp. 317–332, 2018/4// 2018, doi: 10.1038/NRD.2018.14;SUBJMETA=153,154,308,555,556,631,692;KWRD=DRUG+DEVELOPMENT,TARGET+IDENTIFICATION,TARGET+VALIDATION.

[137] R. Santos *et al.*, "A comprehensive map of molecular drug targets," *Nat Rev Drug Discov,* vol. 16, no. 1, pp. 19–34, Jan 2017, doi: 10.1038/nrd.2016.230.

[138] C. Finan *et al.*, "The druggable genome and support for target identification and validation in drug development," *Sci Transl Med,* vol. 9, no. 383, Mar 29 2017, doi: 10.1126/scitranslmed.aag1166.

[139] K. M. Yip, N. Fischer, E. Paknia, A. Chari, and H. Stark, "Atomic-resolution protein structure determination by cryo-EM," *Nature,* vol. 587, no. 7832, pp. 157–161, Nov 2020, doi: 10.1038/s41586-020-2833-4.

[140] J. J. Ziarek, D. Baptista, and G. Wagner, "Recent developments in solution nuclear magnetic resonance (NMR)-based molecular biology," *J Mol Med (Berl),* vol. 96, no. 1, pp. 1–8, Jan 2018, doi: 10.1007/s00109-017-1560-2.

[141] T. Sugiki, N. Kobayashi, and T. Fujiwara, "Modern Technologies of Solution Nuclear Magnetic Resonance Spectroscopy for Three-dimensional Structure Determination of Proteins Open Avenues for Life Scientists," *Comput Struct Biotechnol J,* vol. 15, pp. 328–339, 2017/1// 2017, doi: 10.1016/j.csbj.2017.04.001.

[142] O. Degtjarik, G. Demo, M. Wimmerova, and I. K. Smatanova, "X-Ray crystallography," *Plant Structural Biology: Hormonal Regulations,* pp. 203–221, 2018/8// 2018, doi: 10.1007/978-3-319-91352-0_11.

[143] J. Drenth and J. Mesters, "Principles of protein X-ray crystallography: Third edition," *Principles of Protein X-Ray Crystallography: Third Edition,* vol. 9780387333342, pp. 1–332, 2007, doi: 10.1007/0-387-33746-6/COVER.

[144] J. Jumper *et al.*, "Highly accurate protein structure prediction with AlphaFold," *Nature,* vol. 596, no. 7873, pp. 583–589, Aug 2021, doi: 10.1038/s41586-021-03819-2.

[145] S. Z. Grinter and X. Zou, "Challenges, applications, and recent advances of protein-ligand docking in structure-based drug design," *Molecules,* vol. 19, no. 7, pp. 10150–76, Jul 11 2014, doi: 10.3390/molecules190710150.

[146] X. Y. Meng, H. X. Zhang, M. Mezei, and M. Cui, "Molecular docking: a powerful approach for structure-based drug discovery," *Curr Comput Aided Drug Des,* vol. 7, no. 2, pp. 146–57, Jun 2011, doi: 10.2174/157340911795677602.

[147] A. T. Brint and P. Willett, "Algorithms for the identification of three-dimensional maximal common substructures," May 1, 2002, doi: 10.1021/ci00056.

[148] D. Fischer, R. Norel, H. Wolfson, and R. Nussinov, "Surface motifs by a computer vision technique: searches, detection, and implications for protein-ligand recognition," *Proteins,* vol. 16, no. 3, pp. 278–92, Jul 1993, doi: 10.1002/prot.340160306.

[149] R. Norel, D. Fischer, H. J. Wolfson, and R. Nussinov, "Molecular surface recognition by a computer vision-based technique," *Protein Eng,* vol. 7, no. 1, pp. 39–46, Jan 1994, doi: 10.1093/protein/7.1.39.

[150] M. Rarey, B. Kramer, T. Lengauer, and G. Klebe, "A fast flexible docking method using an incremental construction algorithm," *J Mol Biol,* vol. 261, no. 3, pp. 470–89, Aug 23 1996, doi: 10.1006/jmbi.1996.0477.

[151] R. L. DesJarlais, R. P. Sheridan, J. S. Dixon, I. D. Kuntz, and R. Venkataraghavan, "Docking flexible ligands to macromolecular receptors by molecular shape," May 1, 2002, doi: 10.1021/jm00161.

[152] A. R. Leach and I. D. Kuntz, "Conformational analysis of flexible ligands in macromolecular receptor sites," *Journal of Computational Chemistry,* vol. 13, no. 6, pp. 730–748, 2004, doi: 10.1002/jcc.540130608.

[153] A. Miranker and M. Karplus, "Functionality maps of binding sites: a multiple copy simultaneous search method," *Proteins,* vol. 11, no. 1, pp. 29–34, 1991, doi: 10.1002/prot.340110104.

[154] M. B. Eisen, D. C. Wiley, M. Karplus, and R. E. Hubbard, "HOOK: a program for finding novel molecular architectures that satisfy the chemical and steric requirements of a macromolecule binding site," *Proteins,* vol. 19, no. 3, pp. 199–221, Jul 1994, doi: 10.1002/prot.340190305.

[155] D. S. Goodsell, H. Lauble, C. D. Stout, and A. J. Olson, "Automated docking in crystallography: analysis of the substrates of aconitase," *Proteins,* vol. 17, no. 1, pp. 1–10, Sep 1993, doi: 10.1002/prot.340170104.

[156] T. N. Hart and R. J. Read, "A multiple-start Monte Carlo docking method," *Proteins,* vol. 13, no. 3, pp. 206–22, Jul 1992, doi: 10.1002/prot.340130304.

[157] C. M. Oshiro, I. D. Kuntz, and J. S. Dixon, "Flexible ligand docking using a genetic algorithm," *J Comput Aided Mol Des,* vol. 9, no. 2, pp. 113–30, Apr 1995, doi: 10.1007/BF00124402.

[158] G. M. Morris *et al.*, "Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function," *Journal of Computational Chemistry,* vol. 19, no. 14, pp. 1639–1662, 1998, doi: 10.1002/(sici)1096-987x(19981115)19:14<1639::Aid-jcc10>3.0.Co;2-b.

[159] G. Jones, P. Willett, R. C. Glen, A. R. Leach, and R. Taylor, "Development and validation of a genetic algorithm for flexible docking," *J Mol Biol,* vol. 267, no. 3, pp. 727–48, Apr 4 1997, doi: 10.1006/jmbi.1996.0897.

[160] N. Moitessier, P. Englebienne, D. Lee, J. Lawandi, and C. R. Corbeil, "Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go," *Br J Pharmacol,* vol. 153 Suppl 1, no. Suppl 1, pp. S7–26, Mar 2008, doi: 10.1038/sj.bjp.0707515.

[161] R. A. Friesner *et al.*, "Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy," *J Med Chem,* vol. 47, no. 7, pp. 1739–49, Mar 25 2004, doi: 10.1021/jm0306430.

[162] T. A. Halgren *et al.*, "Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening," *J Med Chem,* vol. 47, no. 7, pp. 1750–9, Mar 25 2004, doi: 10.1021/jm030644s.

[163] R. A. Friesner *et al.*, "Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes," *J Med Chem,* vol. 49, no. 21, pp. 6177–96, Oct 19 2006, doi: 10.1021/jm051256o.

[164] K. Roos *et al.*, "OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules," *J Chem Theory Comput,* vol. 15, no. 3, pp. 1863–1874, Mar 12 2019, doi: 10.1021/acs.jctc.8b01026.

[165] M. D. Eldridge, C. W. Murray, T. R. Auton, G. V. Paolini, and R. P. Mee, "Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes," *J Comput Aided Mol Des,* vol. 11, no. 5, pp. 425–45, Sep 1997, doi: 10.1023/a:1007996124545.

[166] D. Ramirez, "Computational Methods Applied to Rational Drug Design," *Open Med Chem J,* vol. 10, no. 1, pp. 7–20, 2016–04–26 2016, doi: 10.2174/1874104501610010007.

[167] F. D. Prieto-Martínez, E. López-López, K. E. Juárez-Mercado, and J. L. Medina-Franco, "Computational Drug Design Methods—Current and Future Perspectives," *In Silico Drug Design,* 2019/01/01, doi: 10.1016/B978-0-12-816125-8.00002-X.

[168] P. Sledz and A. Caflisch, "Protein structure-based drug design: from docking to molecular dynamics," *Curr Opin Struct Biol,* vol. 48, pp. 93–102, Feb 2018, doi: 10.1016/j.sbi.2017.10.010.

[169] E. Wang *et al.*, "End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design," *Chem Rev,* vol. 119, no. 16, pp. 9478–9508, Aug 28 2019, doi: 10.1021/acs.chemrev.9b00055.

[170] S. Genheden and U. Ryde, "The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities," *Expert Opin Drug Discov,* vol. 10, no. 5, pp. 449–61, May 2015, doi: 10.1517/17460441.2015.1032936.

[171] J. A. McCammon, B. R. Gelin, and M. Karplus, "Dynamics of folded proteins," *Nature,* vol. 267, no. 5612, pp. 585–90, Jun 16 1977, doi: 10.1038/267585a0.

[172] S. A. Adcock and J. A. McCammon, "Molecular dynamics: survey of methods for simulating the activity of proteins," *Chem Rev,* vol. 106, no. 5, pp. 1589–615, May 2006, doi: 10.1021/cr040426m.

[173] A. D. Mackerell, Jr., "Empirical force fields for biological macromolecules: overview and issues," *J Comput Chem,* vol. 25, no. 13, pp. 1584–604, Oct 2004, doi: 10.1002/jcc.20082.

[174] H. Sun, S. J. Mumby, J. R. Maple, and A. T. Hagler, "An ab Initio CFF93 All-Atom Force Field for Polycarbonates," May 1, 2002, doi: 10.1021/ja00086.

[175] I. Yungerman, I. Starodumov, A. Fulati, K. Uto, M. Ebara, and Y. Moskovitz, "Full-Atomistic Optimized Potentials for Liquid Simulations and Polymer Consistent Force Field Models for Biocompatible Shape-Memory Poly(epsilon-caprolactone)," *J Phys Chem B,* vol. 126, no. 21, pp. 3961–3972, Jun 2 2022, doi: 10.1021/acs.jpcb.2c01973.

[176] C. Zhang *et al.*, "AMOEBA Polarizable Atomic Multipole Force Field for Nucleic Acids," *J Chem Theory Comput,* vol. 14, no. 4, pp. 2084–2108, Apr 10 2018, doi: 10.1021/acs.jctc.7b01169.

[177] K. N. Kirschner *et al.*, "GLYCAM06: a generalizable biomolecular force field. Carbohydrates," *J Comput Chem,* vol. 29, no. 4, pp. 622–55, Mar 2008, doi: 10.1002/jcc.20820.

[178] P. Mark and L. Nilsson, "Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K," *The Journal of Physical Chemistry A,* vol. 105, no. 43, pp. 9954–9960, 2001/11// 2001, doi: 10.1021/jp003020w.

[179] M. W. Mahoney and W. L. Jorgensen, "A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions," *The Journal of Chemical Physics,* vol. 112, no. 20, pp. 8910–8922, 2000, doi: 10.1063/1.481505.

[180] S. W. Rick, "A reoptimization of the five-site water potential (TIP5P) for use with Ewald sums," *J Chem Phys,* vol. 120, no. 13, pp. 6085–93, Apr 1 2004, doi: 10.1063/1.1652434.

[181] S. Izadi, R. Anandakrishnan, and A. V. Onufriev, "Building Water Models: A Different Approach," *J Phys Chem Lett,* vol. 5, no. 21, pp. 3863–3871, Nov 6 2014, doi: 10.1021/jz501780a.

[182] A. C. Simmonett and B. R. Brooks, "A compression strategy for particle mesh Ewald theory," *J Chem Phys,* vol. 154, no. 5, p. 054112, Feb 7 2021, doi: 10.1063/5.0040966.

[183] R. J. Loncharich, B. R. Brooks, and R. W. Pastor, "Langevin dynamics of peptides: the frictional dependence of isomerization rates of N-acetylalanyl-N'-methylamide," *Biopolymers,* vol. 32, no. 5, pp. 523–35, May 1992, doi: 10.1002/bip.360320508.

[184] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, "Molecular dynamics with coupling to an external bath," *The Journal of Chemical Physics,* vol. 81, no. 8, pp. 3684–3690, 1984, doi: 10.1063/1.448118.

[185] Q. Ke, X. Gong, S. Liao, C. Duan, and L. Li, "Effects of thermostats/barostats on physical properties of liquids by molecular dynamics simulations," *Journal of Molecular Liquids,* vol. 365, 2022, doi: 10.1016/j.molliq.2022.120116.

[186] G. Bussi, T. Zykova-Timan, and M. Parrinello, "Isothermal-isobaric molecular dynamics using stochastic velocity rescaling," *J Chem Phys,* vol. 130, no. 7, p. 074101, Feb 21 2009, doi: 10.1063/1.3073889.

[187] J. Åqvist, P. Wennerström, M. Nervall, S. Bjelic, and B. O. Brandsdal, "Molecular dynamics simulations of water and biomolecules with a Monte Carlo constant pressure algorithm," *Chemical Physics Letters,* vol. 384, no. 4-6, pp. 288–294, 2004, doi: 10.1016/j.cplett.2003.12.039.

[188] J. Magalhaes *et al.*, "Integration of Enhanced Sampling Methods with Saturation Transfer Difference Experiments to Identify Protein Druggable Pockets," *J Chem Inf Model,* vol. 58, no. 3, pp. 710–723, Mar 26 2018, doi: 10.1021/acs.jcim.7b00733.

[189] O. Korb, H. M. Moller, and T. E. Exner, "NMR-guided molecular docking of a protein-peptide complex based on ant colony optimization," *ChemMedChem,* vol. 5, no. 7, pp. 1001–6, Jul 5 2010, doi: 10.1002/cmdc.201000090.

[190] V. Jayalakshmi and N. R. Krishna, "Complete relaxation and conformational exchange matrix (CORCEMA) analysis of intermolecular saturation transfer effects in reversibly forming ligand-receptor complexes," *J Magn Reson,* vol. 155, no. 1, pp. 106–18, Mar 2002, doi: 10.1006/jmre.2001.2499.

[191] Y. Yuan *et al.*, "Investigation of binding of UDP-Galf and UDP-[3-F]Galf to UDP-galactopyranose mutase by STD-NMR spectroscopy, molecular dynamics, and CORCEMA-ST calculations," *J Am Chem Soc,* vol. 130, no. 10, pp. 3157–68, Mar 12 2008, doi: 10.1021/ja7104152.

[192] M. Parafioriti *et al.*, "Evidence for Multiple Binding Modes in the Initial Contact Between SARS-CoV-2 Spike S1 Protein and Cell Surface Glycans," *Chemistry,* vol. 29, no. 1, p. e202202599, Jan 2 2023, doi: 10.1002/chem.202202599.

[193] (2022). *INN - Pentosan Polysufate sodium*. [Online] Available: https://ec.europa.eu/health/documents/community-register/2022/20220111154170/anx_154170_it.pdf

[194] M. G. Arikan and B. Cakiroglu, "Efficacy of Pentosan Polysulfate Treatment in Patients with Interstitial Cystitis/Bladder Pain Syndrome," *Bladder (San Franc),* vol. 10, p. e21200007, 2023, doi: 10.14440/bladder.2023.866.

[195] D. Cremer and J. A. Pople, "A General Definition of Ring Puckering Coordinates," *Journal of the American Chemical Society,* vol. 97, no. 6, pp. 1354–1358, 1975/3// 1975, doi: 10.1021/JA00839A011/ASSET/JA00839A011.FP.PNG_V03.

[196] M. Ragazzi, D. R. Ferro, B. Perly, P. Sinay, M. Petitou, and J. Choay, "Conformation of the pentasaccharide corresponding to the binding site of heparin for antithrombin III," *Carbohydr Res,* vol. 195, no. 2, pp. 169–85, Jan 15 1990, doi: 10.1016/0008-6215(90)84165-q.

[197] A. G. Gerbst *et al.*, "Ring distortion in pyranosides caused by per-O-sulfation," *Carbohydr Res,* vol. 436, pp. 20–24, Dec 21 2016, doi: 10.1016/j.carres.2016.10.011.

[198] Y. Vo *et al.*, "A Rapid and Mild Sulfation Strategy Reveals Conformational Preferences in Therapeutically Relevant Sulfated Xylooligosaccharides," *Chemistry,* vol. 27, no. 38, pp. 9830–9838, Jul 7 2021, doi: 10.1002/chem.202100527.

[199] T. Nagatsuka, H. Uzawa, and Y. Nishida, "Library assembly of mono-, di- and tri-O-sulfated beta-D-xylopyranosides; effect of O-sulfation on pyranose ring conformation," *Chem Commun (Camb),* no. 27, pp. 4109–11, Jul 21 2009, doi: 10.1039/b907059a.

[200] B. M. Sattelle and A. Almond, "Less is more when simulating unsulfated glycosaminoglycan 3D-structure: comparison of GLYCAM06/TIP3P, PM3-CARB1/TIP3P, and SCC-DFTB-D/TIP3P predictions with experiment," *J Comput Chem,* vol. 31, no. 16, pp. 2932–47, Dec 2010, doi: 10.1002/jcc.21589.

[201] M. P. Jacobson, R. A. Friesner, Z. Xiang, and B. Honig, "On the role of the crystal environment in determining protein side-chain conformations," *J Mol Biol,* vol. 320, no. 3, pp. 597–608, Jul 12 2002, doi: 10.1016/s0022-2836(02)00470-9.

[202] M. P. Jacobson *et al.*, "A hierarchical approach to all-atom protein loop prediction," *Proteins,* vol. 55, no. 2, pp. 351–67, May 1 2004, doi: 10.1002/prot.10613.

[203] G. Eisele *et al.*, "Further advances in identification of pentosan polysulfate monosaccharide composition by NMR," *J Pharm Biomed Anal,* vol. 235, p. 115672, Oct 25 2023, doi: 10.1016/j.jpba.2023.115672.

[204] J. Froese *et al.*, "Evolution of SARS-CoV-2 spike trimers towards optimized heparan sulfate cross-linking and inter-chain mobility," *Sci Rep,* vol. 14, no. 1, p. 32174, Dec 31 2024, doi: 10.1038/s41598-024-84276-5.

[205] G. Paiardi, M. Ferraz, M. Rusnati, and R. C. Wade, "The accomplices: Heparan sulfates and N-glycans foster SARS-CoV-2 spike:ACE2 receptor binding and virus priming," *Proc Natl Acad Sci U S A,* vol. 121, no. 43, p. e2404892121, Oct 22 2024, doi: 10.1073/pnas.2404892121.

[206] C. S. Lupala, Y. Ye, H. Chen, X. D. Su, and H. Liu, "Mutations on RBD of SARS-CoV-2 Omicron variant result in stronger binding to human ACE2 receptor," *Biochem Biophys Res Commun,* vol. 590, pp. 34–41, Jan 29 2022, doi: 10.1016/j.bbrc.2021.12.079.

[207] S. L. Rath, A. K. Padhi, and N. Mandal, "Scanning the RBD-ACE2 molecular interactions in Omicron variant," *Biochem Biophys Res Commun,* vol. 592, pp. 18–23, Feb 12 2022, doi: 10.1016/j.bbrc.2022.01.006.

[208] Z. Zhao *et al.*, "Omicron SARS-CoV-2 mutations stabilize spike up-RBD conformation and lead to a non-RBM-binding monoclonal antibody escape," *Nat Commun,* vol. 13, no. 1, p. 4958, Aug 24 2022, doi: 10.1038/s41467-022-32665-7.

[209] M. H. Peters, "Mutations in the Receptor Binding Domain of Severe Acute Respiratory Coronavirus-2 Omicron Variant Spike Protein Significantly Stabilizes Its Conformation," *Viruses,* vol. 16, no. 6, Jun 4 2024, doi: 10.3390/v16060912.

[210] B. Zheng *et al.*, "S373P Mutation Stabilizes the Receptor-Binding Domain of the Spike Protein in Omicron and Promotes Binding," *JACS Au,* vol. 3, no. 7, pp. 1902–1910, Jul 24 2023, doi: 10.1021/jacsau.3c00142.

[211] R. Nepravishta *et al.*, "Fast Quantitative Validation of 3D Models of Low-Affinity Protein-Ligand Complexes by STD NMR Spectroscopy," 2023/7// 2023, doi: 10.26434/chemrxiv-2022-b7s0x-v2.

[212] H. M. Berman *et al.*, "The Protein Data Bank," *Nucleic Acids Res,* vol. 28, no. 1, pp. 235–42, Jan 1 2000, doi: 10.1093/nar/28.1.235.

[213] P. Han *et al.*, "Receptor binding and complex structures of human ACE2 to spike RBD from omicron and delta SARS-CoV-2," *Cell,* vol. 185, no. 4, pp. 630–640 e10, Feb 17 2022, doi: 10.1016/j.cell.2022.01.001.

[214] S. Jo, T. Kim, V. G. Iyer, and W. Im, "CHARMM-GUI: a web-based graphical user interface for CHARMM," *J Comput Chem,* vol. 29, no. 11, pp. 1859–65, Aug 2008, doi: 10.1002/jcc.20945.

[215] CHARMM-GUI, "CHARMM-GUI," ed, 2019.

[216] H. Woo *et al.*, "Developing a Fully Glycosylated Full-Length SARS-CoV-2 Spike Protein Model in a Viral Membrane," *J Phys Chem B,* vol. 124, no. 33, pp. 7128–7137, Aug 20 2020, doi: 10.1021/acs.jpcb.0c04553.

[217] G. Madhavi Sastry, M. Adzhigirey, T. Day, R. Annabhimoju, and W. Sherman, "Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments," *Journal of Computer-Aided Molecular Design,* vol. 27, no. 3, pp. 221–234, 2013/4// 2013, doi: 10.1007/S10822-013-9644-8/TABLES/9.

[218] M. H. M. Olsson, C. R. SØndergaard, M. Rostkowski, and J. H. Jensen, "PROPKA3: Consistent treatment of internal and surface residues in empirical p K a predictions," *Journal of Chemical Theory and Computation,* vol. 7, no. 2, pp. 525–537, 2011/2// 2011, doi: 10.1021/CT100578Z/SUPPL_FILE/CT100578Z_SI_001.PDF.

[219] C. R. Søndergaard, M. H. M. Olsson, M. Rostkowski, and J. H. Jensen, "Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values," June 9, 2011, doi: 10.1021/ct200133.

[220] J. L. Banks *et al.*, "Integrated Modeling Program, Applied Chemical Theory (IMPACT)," *J Comput Chem,* vol. 26, no. 16, pp. 1752–80, Dec 2005, doi: 10.1002/jcc.20292.

[221] B. Mulloy, M. J. Forster, C. Jones, and D. B. Davies, "N.m.r. and molecular-modelling studies of the solution conformation of heparin," *Biochem J,* vol. 293 (Pt 3), no. Pt 3, pp. 849–58, Aug 1 1993, doi: 10.1042/bj2930849.

[222] J. Singh, Z. Deng, G. Narale, and C. Chuaqui, "Structural interaction fingerprints: a new approach to organizing, mining, analyzing, and designing protein-small molecule complexes," *Chem Biol Drug Des,* vol. 67, no. 1, pp. 5–12, Jan 2006, doi: 10.1111/j.1747-0285.2005.00323.x.

[223] D. A. Case *et al.*, "Amber18 (University of San Francisco)," 2017.

[224] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling, "ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB," *J Chem Theory Comput,* vol. 11, no. 8, pp. 3696–713, Aug 11 2015, doi: 10.1021/acs.jctc.5b00255.

[225] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, "A smooth particle mesh Ewald method," *The Journal of Chemical Physics,* vol. 103, no. 19, pp. 8577–8593, 1995, doi: 10.1063/1.470117.

[226] C. Sagui and T. A. Darden, "Molecular dynamics simulations of biomolecules: long-range electrostatic effects," *Annu Rev Biophys Biomol Struct,* vol. 28, no. 1, pp. 155–79, 1999, doi: 10.1146/annurev.biophys.28.1.155.

[227] J.-P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes," *Journal of Computational Physics,* vol. 23, no. 3, pp. 327–341, 1977/3// 1977, doi: 10.1016/0021-9991(77)90098-5.

[228] D. R. Roe and T. E. Cheatham, 3rd, "PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data," *J Chem Theory Comput,* vol. 9, no. 7, pp. 3084–95, Jul 9 2013, doi: 10.1021/ct400341p.

[229] R. J. Gowers *et al.*, "MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations," *scipy,* pp. 98–105, 2016, doi: 10.25080/MAJORA-629E541A-00E.

[230] J. D. Hunter, "Matplotlib: A 2D Graphics Environment," *Computing in Science & Engineering,* vol. 9, no. 3, pp. 90–95, 2007, doi: 10.1109/mcse.2007.55.

[231] M. Waskom, "seaborn: statistical data visualization," *Journal of Open Source Software,* vol. 6, no. 60, pp. 3021–3021, 2021/4// 2021, doi: 10.21105/joss.03021.