
UNIVERSITY
OF BRESCIA

**PHD PROGRAM IN TECHNOLOGY FOR HEALTH**

XXXIII CYCLE

Department of Information Engineering
Via Branze 38, 25123 Brescia, Italy

---

# Glycan-protein interactions in pathological processes: development of new molecular and cellular models for glycobiology-oriented studies

---

Candidate:
Giulia Paiardi

Supervisor:
Prof. Marco Rusnati

Co-supervisor:
Prof. Paolo Bergese

SSD MED/04

*To my family*

# Table of contents

# Riassunto

Uno degli obiettivi principali della ricerca medica è lo studio della relazione tra le biomolecole, le loro interazioni e le patologie correlate. A riguardo, sono stati sviluppati numerosi approcci sperimentali volti a ottenere risposte sempre più specifiche, che tuttavia rimangono deficitari riguardo la comprensione dei meccanismi molecolari alla base di queste interazioni. Negli ultimi decenni, grazie al processo tecnologico e sistemi di calcolo sempre più avanzati, gli studi computazionali ed in particolare la simulazione di modelli molecolari sempre più complessi hanno assunto un ormai insostituibile ruolo indirizzamento alla ricerca medica.

Quando si parla di modelli molecolari si fa riferimento alla descrizione numerica di strutture complesse partendo dalle loro caratteristiche (quali geometria, energia, potenziale elettrico, ionizzazione e proprietà spettroscopiche) che permettono la simulazione del loro comportamento e delle loro interazioni a livello atomico basandosi sulle leggi della fisica classica o quantistica.

In questo ambito, è preoccupante il ritardo nello sviluppo dei modelli molecolari per la glicobiologia, nonostante la dimostrata importanza degli zuccheri nei processi fisiologici e patologici. Tale ritardo è principalmente dovuto alla complessità strutturale, flessibilità e lunghezza di queste macromolecole biologiche.

In quest'ottica, lo scopo di questo lavoro è stata l'implementazione di nuove procedure e modelli molecolari computazionali (simulazioni di "docking" e di dinamica molecolare) nel campo della glicobiologia che permettono la migliore comprensione dell'interazione non covalenti degli zuccheri con diverse proteine e le loro conseguenze in molteplici processi-fisiologici e patologici.

I modelli molecolari dei complessi proteine-zuccheri sviluppati durante questo lavoro hanno permesso di caratterizzare i meccanismi diretti ed allosterici con i quali queste lunghe catene polisaccaridiche agiscono sulle proteine. L'analisi di questi complessi attraverso dinamiche molecolari ha permesso di dimostrare che questi zuccheri agiscono duplicemente: (i) favorendo l'oligomerizzazione delle proteine legate e inducendo importanti cambi conformazionali alla loro struttura tridimensionale; (ii) esponendo le proteine legate sulla superficie delle cellule favorendo così la loro interazione con recettori specifici.

Oltre che attraverso la loro interazioni non covalenti con distinte proteine, gli zuccheri svolgono un ruolo importante quando covalentemente associati alle proteine (principalmente recettori) a formare le cosiddette "glicoproteine". A riguardo, in questo lavoro sono riportati modelli molecolari che hanno permesso di dimostrare l'effetto di stabilizzazione e mascheramento dei glicani (soprattutto acido sialico) legati covalentemente ai recettori studiati.

In conclusione, i risultati riportati in questa tesi dimostrano quanto l'applicazione degli studi computazionali alla glicobiologia sia ancora agli albori. La medicina di precisione, che sta assumendo sempre più importanza in ambito biomedico, non può prescindere dall'approfondita conoscenza delle interazioni biomolecolare. A questo riguardo, l'integrazione tra modelli computazionali e risultati sperimentali adottato in questi lavori si è dimostrato estremamente proficuo.

# Abbreviations

| | |
|---|---|
| ACE2 | Angiotensin-converting enzyme 2 |
| AMBER | Assisted model building and energy refinement |
| aMD | Accelerated molecular dynamic |
| CHARMM | Chemistry at Harvard macromolecular mechanics |
| Cryo-EM | Cryo electron microscopy |
| dPCA | Dihedral principal component analysis |
| ED | Essential dynamics |
| FF | Force field |
| GA | Genetic algorithm |
| GAFF | General AMBER force field |
| GAGs | Glycosaminoglycans |
| Glc | glucosamine |
| GlcNAc | *N*-acetyl-D-glucosamine |
| GlcA | D-glucuronic acid |
| GPU | Graphic processing unit |
| HBD | Heparin binding domain |
| HIV | Human immunodeficiency virus |
| HPC | High performance computing |
| HS | Heparan sulfate |
| HSPGs | Heparan sulfate proteoglycans |
| IdoA | Iduronic acid |
| IFP | Interaction fingerprint analysis |
| LJ | Lennard-Jones |
| MD | Molecular dynamics |
| NEUs | sialidases |
| NMR | Nuclear magnetic resonance |
| NPT | Ensembles with constant number of particles, pressure and temperature |
| NTD | N-terminal domain |
| NVT | Ensembles with constant number of particles, volume and temperature |
| OPLS-AA | Optimized potentials for ligand simulation-all-atom |
| PBC | Periodic boundary condition |
| PBD | Protein data bank |
| PCA | Principal component analysis |
| PME | Particle mesh Ewald method |
| RBD | Receptor binding domain |

| | |
|---|---|
| RBm | receptor binding motif |
| REMD | Replica exchange molecular dynamic |
| RMSD | Root mean square deviation |
| RMSF | Root mean square fluctuation |
| S | Spike glycoprotein |
| SARS-CoV-2 | Severe acute respiratory syndrome coronavirus 2 |
| SASA | Solvent accessible surface area |
| Sia | Sialic acid |
| Ter | Terminus |
| VEGFA | Vascular endothelial growth factor |
| VEGFR | Vascular endothelial growth factor receptor |
| VMD | Visual molecular dynamic |

*“Stay hungry. Stay foolish.”*
*[Stewart Brand. Whole Earth Catalog]*

# Preface. A bittersweet computational journey

It was in 1902 when Hermann Emil Fischer, german chemistry professor, was awarded for the first time with the Nobel prize in chemistry *"in recognition of the extraordinary services he has rendered by his work on sugar synthesis"*. And this was just the beginning of something revolutionary. Along the years, many other Nobel laureates have been awarded with these prizes in fields related to the history of glycobiology as Karl Landsteiner (Physiology or Medicine, 1930), *"for his discovery of human blood groups"* and Luis F. Leloir (Chemistry, 1970), *"for his discovery of sugar nucleotides and their role in the biosynthesis of carbohydrates"*.

Although this long story, there has been relatively little attention paid to the various ways in which proteins are "tweaked" through the attachment of sugars. However, the role of glycans is far from a decorative function. Glycans are not only involved in catalytic reactions but also help determine the three-dimensional structures of proteins, which are inherently linked to their function and their activity. Moreover, in contrast to some of the other chemical tags employed by cells, sugars exhibit a mind-boggling diversity of structures, can confer cell-type specificity, and are crucial components of cell-to-cell signaling. At the same time, glycans make problematic characterizations; they are the most difficult biological molecules to analyze and synthesize.

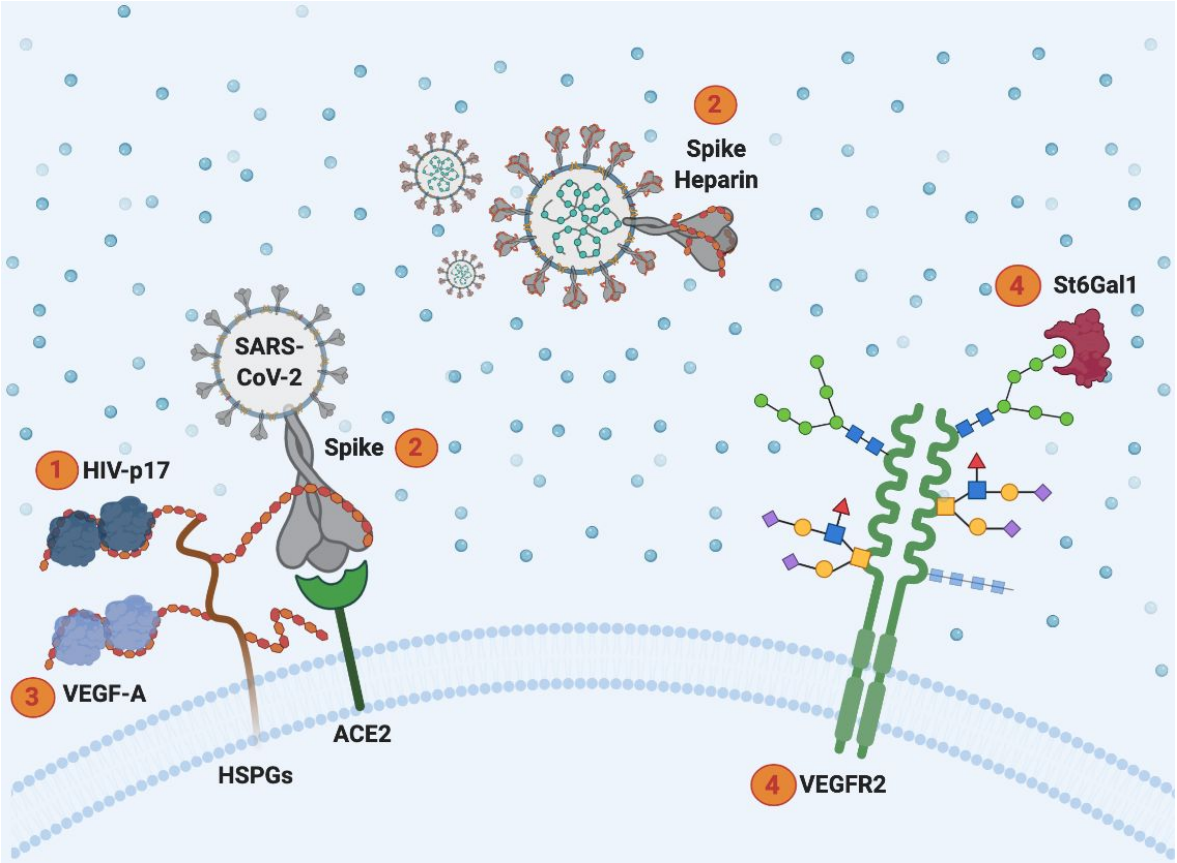
In my Ph.D. years, I have tried to pinpoint the role of glycans from a computational perspective, studying the interactions between sugars and different - but in some cases mechanistically related - sophisticated biomolecules. Molecular docking, classical molecular dynamics simulations and enhanced sampling have been the "vessels" of this exciting bittersweet computational journey from the role of heparan sulfate proteoglycans and heparin to the sialic acid. Importantly, all the data reported in this work are based on molecular models in line with the experimental data reported in the original paper or already released in literature.

In **Chapter 1** I introduce the concept of molecular models in glycomics with an overview of the problem related with this field of research. Subsequently, I present the molecular models techniques that I have used in this thesis aimed at reconstructing the ratio of processes.

The **Chapter 2** consists of a brief review about structure, role and activity of heparan sulfate proteoglycans and heparin. Sections 2.1, 2.2 and 2.3 contain the results achieved during my work about the mechanistic effects induced by these sugars interacting with p17-HIV protein, SARS-CoV-2 spike glycoprotein and vascular endothelial growth factor, respectively. The results presented in Section 2.1 are published in 2019 in *Scientific Reports* and consist in the implementation of a computational method to model long sugar interacting with proteins. This project was conducted in collaboration with the National Research Council under the co-supervision of Dott.ssa Pasqualina D'ursi. The results presented in Section 2.2 were conducted thanks to the funding of the PRACE project (partnership for advanced computing in Europe) under the co-supervision of Prof. Rebecca

Wade during my 9 months endorsement at the Heidelberg Institute for Theoretical Studies. The results are collected in a paper submitted to *PNAS*. In particular, secondary to the COVID-19 pandemia, I have pinpointed the role and the mechanistic effect induced by the sugars mentioned above interacting with spike glycoprotein of SARS-CoV-2. Finally, Section 2.3 gathers the preliminary results of the project that have as specific aim to highlight the structural and dynamical interactions between the vascular endothelial growth factor and the heparan sulphate proteoglycans.

**Chapter 3** is focused on the sialic acid as a new target for drug discovery. As in the previous chapter, after an introduction about this sugar, Section 3.1 presents the data obtained in my work. The results offer a pipeline to set up molecular models of ternary complexes simulating the Michaelis complex useful to identify structural frameworks and druggable pockets. Virtual screening of approved small molecules will be subsequently tested to identify compounds able to interfere with the formation of the ternary complex.




***Graphical representation on the projects.*** *(1) Characterization of the heparin/HIV-p17 oligomerization (2) New insight into the mechanistic effect of heparin/HSPGs on SARS-CoV-2 spike glycoprotein. (3) New insight into the role of heparin and HSPGs in the interaction of VEGF-A and its receptor VEGFR2. (4) Modelling of the VEGFR2-St6Gal1 complex.*

# Chapter 1.

# Molecular Models in Glycomics: an Overview

## 1.1. Introduction

One of the main issues of medical research is the study of the relationship between biomolecules and diseases. To this aim, a variety of approaches have been developed over the years, many of which are based on experimental cell-based assays (1), animal models (2) or, less frequently, clinical-based approaches (3). Although reliable and apt to provide the direct connection between biomolecules and diseases, these approaches lack in understanding the molecular mechanisms underlying the pathological processes under study. To fill this gap, always more often a mixed approach is adopted that flanks the experimental models described above to computational studies (4). Among the latter, molecular modelling in particular is gaining a growing interest and wider exploitation.

But, what is molecular modelling?

Imagine a person who for the first time hears about a bicycle and wants to figure out how it works, how to ride it and how to fix the problems when it breaks. Hard to put everything into words, challenging also to explain all with only a picture. Watching a virtual animation or a movie could really help in the aim. The same is for medical research when trying to understand how two biomolecules interact without knowing their structure or dynamicity means facing the same problem. Researchers can identify and characterize biomolecular interactions experimentally but without understanding the mechanisms involved or the mechanistic effects induced by the interactions. Understanding what happens at the molecular and even atomic level during an interaction is extremely helpful since this type of knowledge can give indispensable insight about the functions of biomolecules, and molecular modelling is so far the only way to clarify these mechanisms.

Molecular modelling is the science of taking into account molecular features like geometry, energy, electric and ionization potential and spectroscopic properties and use them to represent molecular structures numerically, simulating their behaviour on the basis of the equations of quantum and classical physics (5). The representation of molecular structures is often combined with a wide range of experimental techniques such as X-ray crystallography, Nuclear Magnetic Resonance (NMR), Cryo Electron Microscopy (Cryo-EM), electron paramagnetic resonance, and Förster Resonance energy transfer, that provide structures of the biomolecules with different degrees of resolution released in free data banks as Protein Data Bank (PBD) (https://www.rcsb.org/) and are the starting point needed and on which are based the simulations.

Among today's mainstream simulation methods, molecular docking and Molecular Dynamics (MD) simulations are the subject of this dissertation (better detailed at section 1.2 below). The former

tries to predict the interactions between biomolecules starting from their known geometry and the structure of the molecules. The latter allows to study the behaviour of biomolecules at a molecular level through the resolution of the equations of classical physics. In this way, MD simulations can provide dynamic information on a wide range of biological phenomena including conformational changes, protein-ligand and protein-protein interactions, peptide folding as well as predictions of perturbation of the environment as the alteration of pH and salt concentration and mutations (6).

Although molecular docking has been used for a long time as a standalone method for drug design, it is now often integrated into workflows with other computational methods, including MD simulations (Fig. 01). The former is often based on the selection of suitable protein conformations for the docking screening, incorporating structural information and physicochemical properties of the chemical scaffold to select the most suitable poses of the ligand resulting from the molecular docking (Fig. 01). In MD simulation, the flexibility of residues in the binding pocket as well as the potential conformational changes are investigated, greatly improving the results of the virtual studies (Fig. 01) (7).

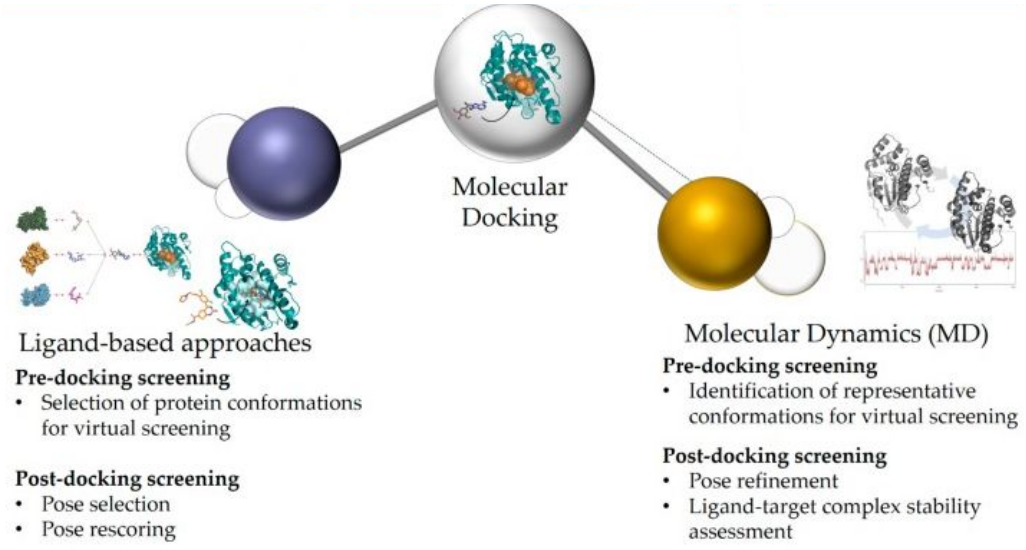



***Fig. 01.*** *Computational methods based on Molecular docking can be divided in ligand-based and structural-based approaches and in particular the MD simulations is the most suitable example for the former. [Adapted from* Pinzi L, et al. (7)*].*

The so-called *omics* branch of science is specifically aimed at characterizing and quantifying large pools of biomolecules translating this information into structures, functions and dynamics. Examples of -*omics* disciplines are genomics, proteomics and glycomics which have as the aim to study the genome, proteins and glycans, respectively. In particular, the latter is the main object of this work. Glycomics, the science which explores the role of carbohydrates in biological processes, is assuming an increasing importance due to the discovery that a large number of biological phenomena are encoded in glycans structures (or glycocode) and interactions (8).

Glycans, also called carbohydrates, consist of a large family of biomolecules which comprises glycosaminoglycans (GAGs), oligosaccharide chains that can exist in a free, soluble form or covalently attached to proteins or lipids (glycoproteins, proteoglycans and glycolipids, respectively). They are widely expressed and involved in a wide range of physiological and pathological processes. Of particular interest are carbohydrate–protein interactions, which regulate biological phenomena such as signal transduction, protein oligomerization, concentration of protein at the cell-surface and viral infection (9). Thus, correct glycosylation patterns are essential for the normal function of organisms while aberrant glycosylation is associated with many human diseases (10).

Clarifying the role of glycans in these processes is a major challenge owing to various factors. First of all, the complexity of glycans biosynthesis. These pathways are not under direct control of genes being rather governed by different enzymes and affected by multiple factors such as the cellular type on which are expressed and the composition of the surrounding environment (11). Again, the structural complexity of glycans is also tricky. These biomolecules can be divided from a structural point of view in non-linear polymers and linear long polyanionic chains, both characterized by distributions of well-defined conformational states rather than robust secondary or tertiary structures, as in case of proteins and nucleic acids. It derives that the main challenge of structural glycoscience is related to the characterization of structural basis and dynamic features of glycans along with the understanding of their behaviour and mechanistic and/or allosteric effect associated with the protein-glycans interactions. Furthermore, some glycans, as in case of the glycoproteins and GAGs, are carriers of information that can be depicted only secondary to their interaction with other biological macromolecules. In this context, it is imperative that spatial and dynamic properties of sugars have to be accurately determined to clarify the relationship between their structure and function.

Unfortunately, experimental structures collected by X-ray crystallography and NMR are limited because of the inherently difficult to crystallize these biomolecules. Indeed, NMR has been used to characterize the dynamics of glycans in solution but with restricted results due to the flexibility of the glycosidic linkage and the limit of molecular weight constraints (12). Also, X-ray crystallography allows to solve only the smaller systems (i.e. the first monosaccharides covalently attached to proteins) due again to the conformational flexibility of glycans that makes their crystallization very difficult. Finally, prediction of glycans sequence and structure based only on computational methods remains a largely unsolved problem.

In conclusion, the progress made in algorithms and computational power allows MD simulations of glycans but the lack in experimental structures of sugar themselves and in interacting with proteins demands strenuous effort to overcome these limits and clarify the biological phenomena.

## 1.2. Molecular models techniques applied to study the glycan-protein interactions

### 1.2.1. Molecular docking

As mentioned above, collecting crystal structures of proteoglycans or protein-glycans complexes is still a challenging task. On one hand, these interactions often occur in the range of microseconds, as in the case of enzymatic reactions, being thus difficult to crystalize the complex in the exact moment in which the reaction occurs; on the other hand, those interactions occur in a specific microenvironment and/or biological settings which are hardly reproducible in x-ray or NMR conditions. It derives that in absence of experimental data, the current molecular docking software, which has not indifferent limitations predicting the protein-glycan interactions, become indispensable to fulfil the tasks proposed in the following works. Hereafter a description of how molecular docking prediction works.

Molecular docking is a computational method that predicts the favoured complex conformation between biomolecules (ligand-receptor) when these are bound to form a stable complex (13). A single biomolecule is described by its dihedral angles, bond distance and bond angles which defines the geometry and its overall structure. Thus, a unique set of these coordinates defines a specific position of each atom within the biomolecule and a subsequent overall of the three-dimensional structure. However, complexes, as in case of protein-glycan systems, are also regulated by additional forces which act both in the catalytic or active site as well as non-specific ones that interact out of the binding site, increasing the complexity of a correct prediction. Another obstacle to this prediction is the flexibility of glycans and of binding sites. The former is related to the already mentioned 1-4 glycosidic linkage between the monosaccharides. The latter is due to possible fitting of the binding site upon its interaction with the sugar. This flexibility, also defined as “conformational change”, is of importance in case of interaction between protein and strong polyanionic chains such as heparin or heparan sulfate (HS) in particular (better detailed in Chapter 2). Finally, although the release of the first docking software specifically dedicated to study glycans interactions such as Vina-Carb (14), HADDOCK (15) or ClusPro (16), improved the quality of the computation of the force field and potential energy by which molecules are described.

To date, the available docking algorithms are basically split into two main groups: the searching algorithms and the scoring algorithms. The formers create an ensemble of possible conformations of the ligand considering the molecule as a dynamic structure and are usually based on a linear combination of many structures or on genetic algorithms (GA) that generate new conformations as they move along the structures or the GA. The latter are based on scoring functions able to evaluate each different conformation of the system and to provide a value that describes the

energy of the system at a given conformation. Low energies indicate better, more stable interactions. Although each docking software is implemented based on the above-mentioned specific algorithms, all of them rank and select the best pose based on its conformation, orientation and translation.

AutoDock software is the docking software used in the following works because implemented with the most suitable force field to compute the glycans flexibility. AutoDock is based on a GA used to find the global minima. This algorithm describes each ligand state as a genotype whereas the corresponding atomic coordinates as a phenotype. Moreover, the fitness is calculated using the energy function and defines the total interaction energy of the complex. Offspring generations of ligand conformations are generated through crossover, by inheritance of genes and additional mutations from either the parents. Obviously, selection of the offspring conformations is based on the fitness value (17).

In a docking scheme, two issues must be considered: the accuracy of the prediction and efficiency of the calculation using reasonable resources and time. AutoDock 4.2 uses two techniques to meet these requirements, namely a rapid grid-based energy evaluation method combined with an efficient search algorithm for torsional degrees of freedom, which gives the specific advantage to predict the glycans conformations (17). In order to evaluate the energies in a rapid way, the affinity potentials for each atom type of the ligand are pre-calculated in AutoDock, where the protein is put in a grid, the size of which is determined by the size of the ligand and can be changed manually according to practical needs. Each grid point stores the affinity potentials for all the atom types of a ligand with the protein. Then, in an AutoDock job, the interaction energy of a particular ligand conformation with the protein target is calculated using the values from the grids.

The search for the ligand conformation related to the torsional degree of freedom is carried out with the search methods implemented in AutoDock. At present, the most efficient search method is the Lamarckian GA, which uses the already described GA in association with the local search algorithm to achieve efficient global phase space coverage and local search optimization. AutoDock 4.2 uses a semi-empirical scoring function to estimate protein-ligand binding free energies that is based on the combination between the classical force fields with empirical data to rapidly rank the candidates (17). In this way, mechanistic approaches are used to compute the enthalpic contributions while empirical approaches are used for solvation free energy and conformational entropy. Finally, the empirical parameters are taken from fitting to known complexes collected through experimental data (17). The protein and ligand molecules start in unbound conformations and form a bound complex after docking, whose binding free energy ($\Delta G$) is calculated as follows:

$$V = W_{vdW}\sum_{i,j}\left(\frac{A_{ij}}{r_{ij}^{12}}-\frac{B_{ij}}{r_{ij}^{6}}\right)+W_{hbond}\sum_{i,j}E(t)\left(\frac{C_{ij}}{r_{ij}^{12}}-\frac{D_{ij}}{r_{ij}^{10}}\right) + W_{elec}\sum_{i,j}\frac{q_i q_j}{\varepsilon(r_{ij})r_{ij}}+W_{sol}\sum_{i,j}(S_i V_j + S_j V_i)\exp(-\frac{r_{ij}^2}{2\sigma^2}).$$

where V stands for the potential energy, P represents the protein and L the ligand, and $\Delta S_{conf}$ is the entropic charge upon ligand binding. Each potential energy term includes the van der Waals, hydrogen-bond, electrostatics, and desolvation energies:

$$\Delta G = (V_{bound}^{L-L} - V_{unbound}^{L-L}) + (V_{bound}^{P-P} - V_{unbound}^{P-P}) + (V_{bound}^{P-L} - V_{unbound}^{P-L} + \Delta S_{conf}),$$

The weights ($W_{vdw}$, $W_{hbond}$, $W_{elec}$, *and* $W_{sol}$) are obtained by fitting experimental binding affinity data. The first term is a classical 12/6 potential for Lennard-Jones interactions. $A_{ij}$ and $B_{ij}$ are parameters for repulsive and dispersion interactions, respectively. The second term is a hydrogen bond term using a 12/10 potential, and $C_{ij}$ and $D_{ij}$ are obtained by fitting to the typical data of hydrogen bonds in the experiment. The term *E(t)* represents the directionality of the hydrogen bond, and *t* is the deviation from ideal hydrogen bond geometry. The third term is the electrostatic potential with an apparent dielectric constant ε(rij). The last term is an empirical desolvation energy calculated using the volume (V) surrounding a specific atom, a weighted solvation parameter (S), and is a distance-weighting factor σ. Finally, the entropy changes upon ligand binding ($\Delta S_{conf}$) can be estimated from the number of active torsions of the ligand ($N_{tors}$) with an empirical parameter $W_{conf}$:

$$\Delta S_{conf} = W_{conf} N_{tors}.$$

Based on all these considerations, AutoDock seems to be really promising in the field of glycomics. Nevertheless, all that glitters are not gold. As already said, AutoDock was designed as an automated docking software for predicting the interactions between flexible ligands and proteins. Thus, the premises are really good but the concept of flexible ligand is underestimated if we consider glycans structure. This is because the algorithm is set to allow at most 32 free torsions, which are good in case of small molecules but if we consider linear tetrameric structures of glycans, they cover 28 of the 32 total free torsions, giving the idea of the limitations of the method and the exigence to implement new methods to overcome the limitations.

### 1.2.2. Molecular Dynamics (MD) simulations

The history of MD simulations started more than 60 years ago, when in 1957 Ander and Wainwright carried out the first simulation in the context of phase transition in a system of hard spheres but we need to wait until 1977 for the first MD of protein (18). And, slowly, the groundwork that allows to carry out the simulations resulted in the achievement recognized by the 2013 Nobel Prize in Chemistry (19). From there on, MD simulations have become more popular and visible. This

remarkable improvement is in part due to the increased number of available structures in the Protein Data Bank in association with the implementation of computer technologies, such as the high performance computing (HPC), the improvement of supercomputer and the introduction of Graphic Processing Units (GPUs) that allow an easier handling of the techniques and decrease the computing time for the simulations.

MD simulation method consists in solving coupled equations of motion numerically for a system in which the molecules move at a constant velocity. The solution results in a trajectory of the biomolecule, from which thermodynamic and dynamic properties of the system can be extracted by statistical mechanical relationship. Importantly, the reliability of the method for predicting the behaviour of a system depends on the assumptions used to describe the interactions of the system.

In this context, the interest in applying MD simulations to the glycans is increasing and associated with the peculiar features of these biomolecules. The exigence of applying MD simulations is due to the lack of crystal structures as well due to the possibility of pinpoint the dynamic behaviour of molecules which are characterized by a cloud of conformations more than a single secondary or tertiary structure. At this regard, Franck and co-workers highlighted the importance of MD simulations in this field carrying out a simulation of Human Immunodeficiency Virus (HIV) -gp120 glycoprotein fully glycosylated, showing for the first time the cloud trend assumed by the N-glycans covalently attached to the protein and demonstrating the shielding capability of the sugars as a mechanisms adopted by viruses to escape the immunity defences (20). More recently, MD simulations have been instrumental in demonstrating the same shielding effect by Casalino and co-workers for the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV2) virus. MD simulations results have also suggested promising molecular targets to design specific vaccines and small molecules (21).

Hereafter are described the main features on which MD simulations are based to describe systems as realistic as possible.

Describing the systems with molecular dynamics. Considering a system of N particles, this is characterized by a set of atomic positions $(R=\{R_1,R_2...R_N\}$ and relative momenta $(P=\{P_1,P_2...P_N\})$. Each atom can be seen and represented as a single point in a 6N multidimensional space, called phase space (Γ). Thus, a single point in the phase space represents a microscopic state of the system, while a collection of points in the Γ defines an ensemble. Using MD simulations we can generate a time sequence of points in the phase space, i.e. a sequence of different positions and momenta of the system belonging to the same ensemble. For each microscopic state of the system in the phase space, it is possible to estimate the observable value of a certain property A as a function of Γ, A(Γ), as the ensemble average or thermodynamic average:

$$A_{obs} = \langle A \rangle_{ens} = \int A(\Gamma)\rho(\Gamma)d\Gamma$$

where ρΓ is the probability distribution function of collection of points Γ, and δΓ = $dR_1...dR_N dP_1...dP_N$.

The thermodynamic state of the system is defined by parameters as the number of particles N, volume V, temperature T and pressure P. Based on these parameters, the probability distribution function has the form of the Boltzmann distribution function:

$$\rho_{NVT} = \frac{e^{-\frac{H(\Gamma)}{K_B T}}}{Z}$$

where H(Γ) is the classical Hamiltonian of the system defined as:

$$H(\Gamma) = H(\{\vec{R}_I\}, \{\vec{P}_I\}) = \sum_{I=1}^{N} \frac{\vec{P}_I^{\,2}}{2M_I} + U(\{\vec{R}_I\})$$

where $R_I$, $P_I$ *and* $M_I$ are the position, momentum and mass of the particle *I, U* is the potential energy, $K_B$ is the Boltzmann constant and *Z* is the canonical partition function.

On these basis, through the integration of the Newton's equation of motion, we can estimate the evolution of the system as a function of time starting from its microstate at time 0 until its microstate at time τ. Along the MD simulations, a set of microstates of the system, i.e. a trajectory of points in the phase space Γ(t) is collected. According to the "ergodic hypothesis" and starting from a trajectory, we can calculate and connect the time average value of an observable $\langle A \rangle_\tau$ to $\langle A \rangle_{obs}$ (22). Basically, if the system will be able to evolve for an infinite time, it would be it would be able to sample all the possible states and conformations and thus its behaviour averaged over time and over the phase space become the same:

$$\lim_{\tau\to\infty} \langle A(\Gamma) \rangle_\tau = \lim_{\tau\to\infty} \frac{1}{\tau} \int_0^\tau A[\Gamma(t)]dt = \langle A(\Gamma) \rangle_\Gamma.$$

So, the longer is the simulation the better is the sampling of the states. Thus, the average behaviour of molecular biological systems can be predicted by estimating macroscopic observables using MD simulations. In this way, the Lagrangian equation of motion can be used to write the Newton's equations of motion taking in consideration the variation of time:

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{q}_k}\right) - \frac{\partial \mathcal{L}}{\partial q_k} = 0$$

where $q_k$ are the generalized coordinates and $\dot{q}_k$ are their time derivatives. In addition, the Lagrangian function $\mathcal{L}(q,\dot{q},t)$ can be defined in terms of kinetic $K(q\dot{q},t)$ and potential $\Gamma(q\dot{q},t)$ energies:

$$\mathcal{L}(\mathbf{q}\dot{\mathbf{q}}, t) \equiv K(\mathbf{q}\dot{\mathbf{q}}, t) + \Gamma(\mathbf{q}\dot{\mathbf{q}}, t)$$

For a system of atoms with Cartesian coordinates $R_i$ the equation of motion described below yields:

$$\mathbf{F}_i = m_i \mathbf{a}_i = m_i \mathbf{R}_i = \frac{\partial \mathcal{L}}{\partial \mathbf{R}_i} = \frac{\partial \Gamma}{\partial \mathbf{R}_i}$$

where $m_i$ and $a_i$ are the mass and acceleration of particle $i$ and $F_i$ is the force of that particle. Thus, starting from a set of initial coordinates, as PDB crystal structures, NMR data, cryo-EM or homology modelling coordinates, and velocities randomly generated from the Maxwell-Boltzmann probability distribution at a given temperature T.

*Force Fields in MD simulations*. How atoms and molecules interact with each other is a central part for MD simulations. The potential energy of molecules is described as an empirical force field (FF) parametrized to reproduce the experimental data, which represent our knowledge of the potential energy surface of the system and are used to calculate the forces for propagating the dynamical systems. In classical MD simulations, the electronic degrees of freedom are not taken into account and atoms are considered as point particles simply approximated by using parametric functions of nuclear coordinates, the force field. Many force fields are available in the literature, such as CHARMM (Chemistry at Harvard Macromolecular Mechanics), OPLS-AA (Optimized Potentials for Liquid Simulations, All-Atom), AMBER (Assisted Model Building and Energy Refinement), and GAFF (General AMBER Force Field). Notably, in all the projects proposed this dissertation, I have employed AMBER force fields. Each of these force fields is a package that includes multiple parametric functions specifically set up for different biomolecules: protein, ions, water, nucleic acids, sugars. These force fields have similar basic functional form, representing bond stretching, angle bending, dihedral torsion, van der Waals, and Coulombic interactions for each molecule, and are described as:

$$U_{tot} = \sum_{bonds} \frac{k_l}{2}(l - l_{eq})^2 + \sum_{angles} \frac{k_\theta}{2}(\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{k_\chi}{2}[1 + \cos(n\chi - \delta_\chi)] + \sum_{i=1}^{N-1} \sum_{i<j}^{N} \left\{ 4\epsilon_{ij} \left[ \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right\}$$

where $U_{tot}$ is the total energy; $k_l$, $k_\theta$ and $k_\chi$ are the bond, angle, and dihedral angle force constants, respectively; $l_{eq}$, $\theta_{eq}$, and $\chi$ the bond length, bond angle, and dihedral angle, respectively, and the subscript *eq* represents the equilibrium values for the individual terms. Coulomb and Lennard-Jones (LJ) terms contribute to non-bonded interactions where $\varepsilon_{ij}$ is the LJ well depth and $\sigma_{ij}$ is the finite distance at which the inter-particle potential is zero, $q_i$ is the partial atomic charge, and $r_{ij}$ is the

distance between atoms *i* and *j*. Due to this fact, LJ potential is considered a short-range potential and calculated using a cut-off of 10 Å (23).

MD simulations in explicit solvent consider the solvent explicitly, that is characterized by atoms connections between the individual water molecules. The water molecules are described by appropriate force fields. Actually, the most used family of water-models are the so-called fixed charge force fields, in which the fixed charges are assigned on the four fixed sites of the molecule: two charges are positive and simulate the hydrogen atoms, whilst the other two are negative and simulate the lone pairs. The four charges are arranged tetrahedrally about the oxygen (24). Within this family there are some of the most popular FF, for example TIP3P and TIP4P (25), which are widely used in the simulations of biological systems. In conclusion, if the essential physical interactions of molecules and/or particles are represented faithfully, explicit methods can reliably predict properties.

*Glycans Force Field in MD simulations*. Being the glycans the subject of my work, it is important to describe the force fields specifically developed for sugars, that are the main features characterizing glycans in MD simulations. Here a description of the GLYCAM_06 force field (26), used in this dissertation. Before the release of the GLYCAM_06 force field, the performance of several parameter sets, which included the first versions of GLYCAM (27), has been quantitatively evaluated against data from quantum mechanical calculations as well as on a relative basis. The conclusion was that all the parameters and force field evaluated perform similar results and no one was top ranked compared to the others. On a low note, the early version of GLYCAM used to simulate sugars in explicit solvent worked poorly in reproducing diffusion rate and differ substantially compared to the other force fields in predicting the putative radial pair distribution function (RPD) between hydroxyl groups and water model (TIP3P) (26).

The low results compared to the other parameters prompted the optimization and release of the GLYCAM_06 force field. At first, the structures of the monosaccharides were collected using experimental neutron diffraction with particular attention at the main features of the glycans such as the conformation or glycosidic linkage. In addition, the new force field package comprises new parameters such as bond and valence angle deformation force constants, dihedral angle rotational barriers, and electrostatic properties, which are collected with quantum mechanical (QM) calculations because of the difficulty or impossibility to calculate them with experimental data. Finally, the partial atomic charges were derived by fitting the QM molecular electrostatic potentials. In particular, this method starts from the general practice of assigning partial atomic charges to every atom in the molecule adopted (26) by not fitting partial charges to aliphatic hydrogen atoms. In addition the special treatment for the 1–4 non-bonded interaction was removed because it was redundant (26).

*Periodic boundary condition*. Consider a molecule arranged in a 10 × 10 × 10 water box. Nearly half the water molecules are on the outer faces of the box and would lie at the edge with the vacuum, generating artifacts due to finite-size effects of the system, and these will have a large effect on the measured properties. To avoid this problem, periodic boundary condition (PBC) can be applied. This method allows to ideally surround the central cube, the only one that will be explicitly treated, with infinite water boxes with the result of simulating an infinite solution system around the box (28). Finally, it's important to bear in mind the periodicity of the system when considering properties which are influenced by long-range correlations. Fortunately, many algorithms have been implemented to overcome this problem and among them particle mesh Ewald method (PME) is the one adopted in the MD simulations carried out in this dissertation (29). The basic idea behind this technique is to split the relevant part of the potential into a short-range part, ordinarily treated within a cut-off, and a long-range term, in which the remaining interactions are Fourier transformed (30).

*Temperature and Pressure*. To carry out MD simulations as realistic as possible, the systems are described as microcanonical ensembles in which the number of particles (N), volume (V) and total energy (E) are kept constant. However, to obtain more accurate information that can be related to the experiments, MD simulations can be coupled with thermostat and barostat. The former allows to simulate canonical ensembles (NVT) in which the volume (V) and temperature (T) are kept constant. The latter maintains pressure (P) and temperature (T) stable (NPT ensemble). During this time, many algorithms have been implemented but I will briefly describe the two used for the simulations presented in this dissertation (30).

*Langevin thermostat.* In the Langevin thermostat the equation of motion is modified to maintain a stable temperature as following:

$$m\ddot{\mathbf{r}}_i = -\nabla_i U - m\Gamma\dot{\mathbf{r}}_i + \mathbf{W}_i(t)$$

where Γ is a friction coefficient and $W_i$ is a random force unrelated to time and across particles. The result is that the smaller particles create a damping force to the momenta when the large particles push the smallest out of the way. The smaller (thermal) particles also move with kinetic energy and give random kicks to the large particles, resulting in a random force that acts on all the particles and decreases their velocity using a constant friction. The Langevin thermostat is mostly used in the NVT-MD simulations carried out in my work (http://www.math.ucsd.edu/).

*Berendsen thermostat.* In the NPT ensemble, the pressure is controlled by coupling the system with a barostat. One of the main problems of the velocity-rescaling method is that it does not allow temperature fluctuations which are present in the canonical ensemble. To overcome this problem, Berendsen and co-workers introduce a first-order coupling barostat in conjunction with the

temperature control method. With this algorithm, the pressure of the simulated system is relaxed using a time constant toward the reference pressure (31). One advantage of the Berendsen thermostat is that it allows the temperature fluctuations, thereby not fixing it to a constant value. The motivation for the Berendsen thermostat is the minimization of local disturbances of a stochastic thermostat while keeping the global effects unchanged (http://www.math.ucsd.edu/).

### 1.2.3. Enhanced sampling techniques

Classical MD simulations have been shown as powerful tools to understand and predict the evolution of complex biological systems, but this method also has limitations. Using this technique is hard to sample conformational variations of small peptides or complex systems because it is hard to overcome high-energy-barriers to jump between the states with a reasonable amount of replicas and simulation time. Thus, enhanced sampling methods have been developed in recent years to accelerate rare events with notable success. These methods allow us to overcome the local-minimum states by enhancing the sampling as a function of one or a few predefined collective variables relevant for the event, increasing the probability of these events. Among these methods, replica exchange MD (REMD) simulation and the accelerate MD (aMD) were used in my work and are briefly presented hereafter.

*REMD*. This technique allows to overcome high energy-barrier and to sample the conformation of peptides and proteins by combining MD simulations with the Monte Carlo algorithm (32). With this algorithm, several replicas of the starting system are simulated in parallel at different temperatures or at the same temperature but using different Hamiltonians at which the replica can easily surmount the energy barriers. Periodically, the configurations of neighbouring replicas are exchanged and this exchange is accepted by a Metropolis acceptance criterion that guarantees the detailed balance (Fig. 02).

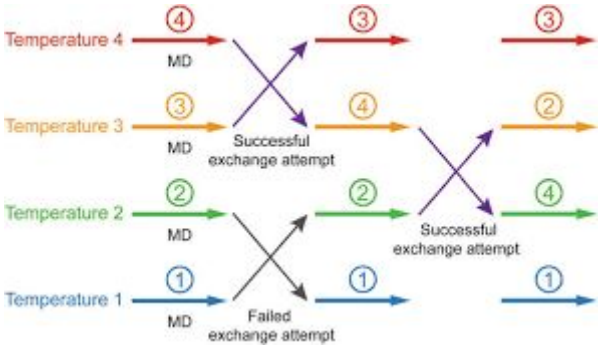



***Fig 02.*** *Illustration of replica exchange molecular dynamics (REMD) method.*

The prerequisite for an efficient sampling is an adequate overlap of potential energy distributions between replicas. The overlap is large enough for satisfying acceptance ratios (10-50 %)

in case of systems with a small number of degrees of freedom and small temperature differences between the replicas. In this way, REMD is able to overcome the local minimum and successfully sample the conformational space through the exploration of the free energy landscape (33). REMD simulations are therefore restricted to systems comprising not more than a few thousand atoms and/or small peptides, typically in the context of peptide folding.

*aMD*. Hamelberg and co-workers proposed an aMD algorithm which is based on a bias potential function that can be used to overcome the local minimum without knowing other information about the potential energy neither the saddle point (28). With this algorithm, it is possible to sample the potential energy landscape using a bias potential added to the pre-existing one which allows to accelerate the time scale of the simulation. The representation of the potential energy landscape on the modified surface derives and is an expression of the bias potential, permitting it to successfully sample the potential energy minima. Using this bias potential it is possible to sample the conformational space more exhaustively than in the case of classical MD simulations, thus reaching the convergence of the system (34). aMD simulations demonstrate to be useful also for the sampling of systems composed of small peptides up to complete protein structures.

Based on all the considerations and information collected in this introductory chapter, we can conclude that glycomics studies are still in infancy beside the huge potential of the molecular model to sustain the experimental research. For this reason, the overarching goal of the studies reported in this work is the development of proper bioinformatics tools to allow the computational characterization of sugar-protein interaction and the prediction of their biological consequences, with the specific aim of merge the predicted results with experimental data to fulfil the specific objectives of the glycobiology.

# Chapter 2.

# Demystifying Heparan Sulfate-Proteins interactions

Proteoglycans are the most abundant and heterogeneous family of co-receptors expressed on the cell surface and in the extracellular matrix. Characterized by one or more polysaccharide chains (GAGs) covalently attached to the core protein, proteoglycans are involved in a variety of physiological and pathological processes, so that the comprehension of their mechanistic effects at cellular levels are required for the comprehension of many body functions and diseases.

Heparan sulfate proteoglycans (HSPGs) is a subset of proteoglycans composed of a core protein on which are covalently attached long linear heparan sulfate (HS) polyanionic chains. Whether the structure of the core protein can affect the function of the HSPGs is not clear but is known that the core can directly interact with both intra- and extra-cellular protein. HSPGs are classified based on their localization and structure of the core protein. The main families are represented by HSPGs whose core is a trans-membrane protein with a short intracellular tail (Syndecans) or are anchored to the membrane through a glycophosphatidylinositol residue (Glypicans). Other HSPGs are instead secreted by cells and deposited in the extracellular matrix (collagen XVII, agrin and perlecan). Finally, serglycin are HSPGs found in the cytoplasmic secretory granules of endothelial, endocrine, and hematopoietic cells.

The long linear HS are composed by repetitive disaccharide units of β1–4-linked D-glucuronic acid (GlcA) or and α1–4-linked *N*-acetyl-D-glucosamine (GlcNAc) that undergo different modifications including the sulfation at specific positions to confer highly negative charge and binding-specificity. In particular, during the biosynthesis, HS can be subjected to a series of reactions processes: (i) deacetylation or sulfation of the GlcNAc (the most frequent at position C2 and C6 to GlcNS6S); (ii) epimerization of adjacent C5-GlcA to L-iduronic acid (IdoA); (iii) ester-linked sulfate groups at C2-IdoA (IdoA2S) or at C6-GlcA or more rarely at C3-GlcNAc. Interestingly, these modifications do not occur along all the HS at the same rate giving a rise of specific domain based on the sulfation: sections with variably sulfated domains (so-called NS domains), other with no modifications (the so-called NAc domains), other with a mix of NAc/NS domain located in the transition zone. Finally, at variance with proteins, HS are not synthesized under the direct control of genes, thus the synthesis of HS is not template driven and can greatly change among the different species and even cell types. What dictates the size and composition of HS in different cells at different times in development remains one of the great enigmas in modern cell glycobiology (9).

Relevantly, the terms HS and heparin are often used interchangeably but this is a simplification that needs an explanation. HS are synthesized by almost all eukaryotic cells, covalently attached to proteoglycans and their length can span from 50 to 250 disaccharides units (20-100kDa). Differently, heparin is mainly produced by mast cells, contains larger NS domains and fully sulfated disaccharides

sequences (Fig. 03). Heparin can be found in two main states: unfractionated heparin (spanning from 12 to 14 kDa, ≈30 disaccharides) and low molecular weight heparin (spanning from 4 to 6 kDa, ≈15 disaccharides). Interestingly, the latter are extensively used as anticoagulant because the high pattern of sulfation also comprises the rare sulfation at 3-O-GlcNAc which specifically acts inhibiting the antithrombin III (35). Finally, due to the high degree of heterogeneity that characterizes the HS structure depending on their source, in both experimental and computational studies very often heparin is used as a "simplification" of HS.

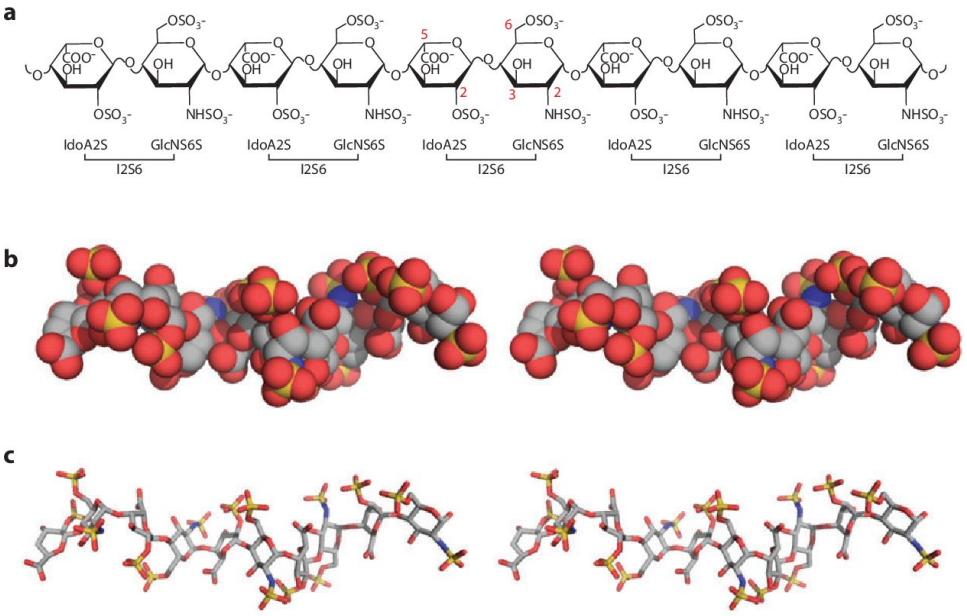



***Fig. 03***. ***(a)*** *Heparin chemical structure. I2S6 is the most abundant disaccharide unit identified in heparin and is characterized by the sulfation at C2-IdoA and at N- and C6-GlcNac. The red numbers associated with the carbon atoms highlight the possible substitutions. (**b** and **c**) Heparin crystal structures (PDBid 1E0O) are represented in spheres and sticks, respectively. Both the structures are colored by elements: carbon, oxygen, nitrogen and sulfate atoms are colored in grey, red, blue and yellow, respectively. In both the structures it is possible to notice the helical nature of the heparin structure. Abbreviations: GlcNS6S, N-sulfo 6-sulfo Glc; IdoA2S, 2-sulfo L-iduronic acid. [Adapted from Xu D et al (9)].*

HSPGs are involved in multiple biological phenomena through their capability to interact with a wide range of protein-partners, the so-called HS- or heparin-binding proteins. These proteins are characterized by the presence of basic domains, corresponding to sequences of basic residues that primarily interact with the negatively charged sulfated groups of HS/heparin. These basic motifs were identified for the first time by Cardin and Weintraub in 1989 and for this reason called Cardin-Weintraub motifs. The Cardin-Weintraub motifs, also known as heparin-binding domains (HBDs) because of the similarity between HSPGs and heparin, consist of specific sequences 'XBBXBX' and 'XBBBXXBX', where X represents an hydropathic residue and B a basic residue (R or K), which are responsible for the interactions with the sulfate groups of the HS (36). Through the

interactions with the heparin-binding proteins, HSPGs can regulate a plethora of biological processes including protein oligomerization (37), ligand/receptor interaction (38) and protein internalization (39) that can affect cell proliferation, adhesion and migration (40), angiogenesis (41) and viral infection (42).

Finally, HSPGs and heparin are able to induce conformational drifts secondary to the interaction with the heparin-binding proteins. For this reason, they carry information that can be depicted only pinpointing and characterizing the systems at molecular level. Unfortunately, a few number of crystal structures of short heparin/HSPGs and no crystal structure of long heparin/HS in complex with protein have ever been released. In addition, publications in which the dynamic behaviour of these systems is characterized are so far limited despite the importance of issue.

Thus, the detailed understanding of the HS and heparin-dependent processes is required to clarify both physiological and pathological processes at molecular level. For this reason, the first aim of my research, described here below in Sections from 2.1 to 2.3, has been directed to the implementation of proper bioinformatics tools to allow the computational characterization of protein-HSPGs/heparin molecular models and the characterization of the mechanistic effects related to their biological consequences.

# 2.1. Implementation of an incremental docking method to characterize the heparin/HIV-1 p17 protein interaction

*The results from this work has been collected in a published paper: Bugatti A*, **Paiardi G***, Urbinati C, Chiodelli P, Orro A, Uggeri M, Milanesi L, Caruso A, Caccuri F, D'Ursi P and Rusnati M. Heparin and heparan sulfate proteoglycans promote HIV-1 p17 matrix protein oligomerization: computational, biochemical and biological implications. Scientific Reports. *Co-authorship. (2019) 9:15768. Supplementary information is available for this paper at https://doi.org/10.1038/s41598-019-52201-w.*

## 2.1.2. Introduction

Matrix protein p17 contributes to structural integrity of HIV virions and regulates viral replication (43). Also, p17 is released by HIV-infected cells, being detectable in the nanomolar range in plasma, brain and lymph nodes of patients treated with HAART (44). In its extracellular form, p17 deregulates the functions of B-lymphocytes (45). Importantly, p17 also induces endothelial cell activation and angiogenesis (46). With these multiple effects, p17 contributes to AIDS progression and to the pathogenesis of AIDS-associated diseases (47). HSPGs consist of a core protein and of heparin-like GAG composed of repeated disaccharides units of 2-O-sulfated L-Iduronic acid (IdoA) and N, 6-O-sulfated D-glucosamine (Glc) occasionally interrupted by non-sulfated uronic acids and under-sulfated hexosamines (48). At the surface of leukocytes, they act as receptors for extracellular p17. In the extracellular environment, free heparin released by mast cells sequesters p17, modulating its bioavailability (49). Besides p17, HSPGs act as receptors also for HIV-1 gp120 (50) and Tat (51) while free heparin promotes Tat oligomerization and biological activity (52). HSPGs act as co-receptor also for many other viral proteins and cytokines, promoting their oligomerization required for receptors clustering and activation (53).

p17 spontaneously oligomerizes forming trimers and even hexamers (54). Two different domains have been identified in p17 that specifically mediate its self-assembling: the C-terminus (ter) that self-interacts with the same region of other p17 molecules (55) and amino acids $E_{42}$-$N_{47}$, $Q_{59}$, $Q_{63}$, that interact with the $Q_{69}$-$E_{74}$ region of another p17 molecule (56). On the other hand, due to the presence of "coiled coil" sequences, p17 tends to misfold, behaving as an "amyloidogenic" protein that forms toxic assemblies in the brain that have been demonstrated to contribute to AIDS-associated neurodegeneration (44).

Taken together, the capacity of p17 to oligomerize and to bind to heparin/HSPGs, along with the involvement of the GAG in the process of oligomerization of many cytokines prompted us to study the effect of heparin/HSPGs on p17 oligomerization and biological consequences by adopting a multidisciplinary approach including bioinformatics, biochemical and cell-based models.

For what concerns bioinformatics, the docking of protein to heparin (used as a structural

analogue of HSPGs) and MD simulations have been limited only to short (di-, tetra- or hexa-) oligosaccharides, mainly due to heparin conformational flexibility and high charge density, the weak surface complementarity of heparin/ protein interactions, the absence of well-defined binding pockets, the difficulty to define the impact of solvation/desolvation, the large electrostatic interactions involved and the large number of torsional angles between glycosidic bonds (57). Only in selected cases, 14-mer heparin were used with molecular docking to study FGF, VEGF and CXCL8 dimerization (58-60). Relevant to this point, longer heparins or HS are found in nature that are responsible for cytokine oligomerization. With these premises, to corroborate our experimental data, we have here modelled heparin up to 24-mer and performed docking and MD simulations with p17 dimer.

## 2.1.2. Material and methods

*Models*: p17 and the heparin tetrasaccharide (4-mer heparin) were modelled as described (54). A 4-mer heparin probe with the deletion of the H atom of the hydroxyl groups at position 4 in the first IdoA and of the hydroxyl group at position 1 of the Glc (4-mer heparin modified probe) was prepared and used in docking simulations to promote the 1→4 glycosidic linkage. Two p17 mutants were modelled by the Alanine Scanning Method in which the lysine residues of both the N- and C-ter basic domains were replaced with alanine (N-ter K→A p17 and NC-ter K→A p17).

*Heparin path identification:* Blind docking simulations were performed by ClusPro web-server (16) using 4-mer heparin and N-ter K→A or NC-ter K→A p17 dimers to identify HBDs on p17 other than those in the N- and C-ter basic domains. The identified heparin probe sites were filtered by best score, cluster size, visual inspection and positioned in a dimer of wild type (wt) p17 to finally achieve an alignment and hence a traced heparin path.

*Incremental docking and heparin modelling:* The 4-mer heparin modified probe was used in local docking simulation along the traced heparin path in the wt p17 dimer by AutoDock 4.2 (17). The "sliding window method" was set up to create a sequence of overlapping sliding grids, each covering a whole 4-mer heparin and the last saccharide unit of the previous one. Local docking poses were filtered for free energy of binding, clusters size and correct orientation. The aligned 4-mer heparin modified probes were joined by 1→4 glycosidic linkages using Pymol. Gasteiger-Hückel charges were assigned to the sugar and then minimized using steepest descendent and conjugate gradient methods by Chimera (61), obtaining heparins of increasing length.

*MD simulations:* Amber14 package (62) was used for MD simulations of p17 dimer in complex with 15- and 24-mer heparins. MD simulations were carried out using f99SB force field parameters for protein and GLYCAM_06 for heparins. Each complex was neutralized by adding $Ca^{2+}$ restrained away from the protein (63) and solvated with TIP3P water model. Energy minimization was carried out with the non-bonded cut-off of 8Å through the following steps: (i) p17/heparin complexes and

counter ions were restrained by a harmonic potential of 5 kcal/mol×Å$^2$, while water molecules were relaxed using 2,500 cycles of steepest descent and conjugate gradient methods; (ii) counter ions and hydrogens were relaxed using 5,000 cycles of steepest descent and conjugate gradient methods and restrained by a harmonic potentials of 3 kcal/mol×Å$^2$ and then of 1 kcal/mol×Å$^2$; (iii) the system was relaxed using 5,000 cycles of steepest descent and conjugate gradient methods without any restraint; iv) the system was heated from 0.1K to 100.0K in NVT (constant volume) and from 100.0 to 300.0K in NPT (1.0atm constant pressure). The two generated complexes were simulated in periodic boundary conditions using the Langevin algorithm at 300.0K. During heating and simulations, the $Ca^{2+}$ ions were restrained at 500 kcal/mol×Å$^2$. Equilibration (3 ns) and simulation were validated using the physical observables parameters of the system confirming that the complexes obeyed the NPT ensemble. Electrostatic interactions were calculated using the Particle Mesh Ewald method. A cut-off of 8.0Å was applied to van der Waals forces. Integration time step was set to 1.0 fs during equilibration and 2.0 fs during simulation. MD simulations were performed over 50ns using the pmemd CUDA program of the Amber14 package and a server Tesla K20 GPU.

### 2.1.3. Results

As already mentioned, computational docking of heparin (mostly used as a structural analogue of HSGPs) to its binding-protein still represents a challenge. Accordingly, the p17/heparin complex has been resolved only with a protein monomer and a 6-mer heparin, prompting us to conceive a new computational approach to study the interaction of longer heparins with p17 dimers (Fig. 04). The first step of this workflow consists in a blind docking simulation using ClusPro web server. The results of the simulations with 4-mer heparin probes and wt p17 dimer show that the tetra-saccharides always take position in the N-ter HBDs of the two monomers. On the other hand, using a N-ter K→A p17 dimer, the 4-mer heparin probes dock mainly in the C-ter basic domain.

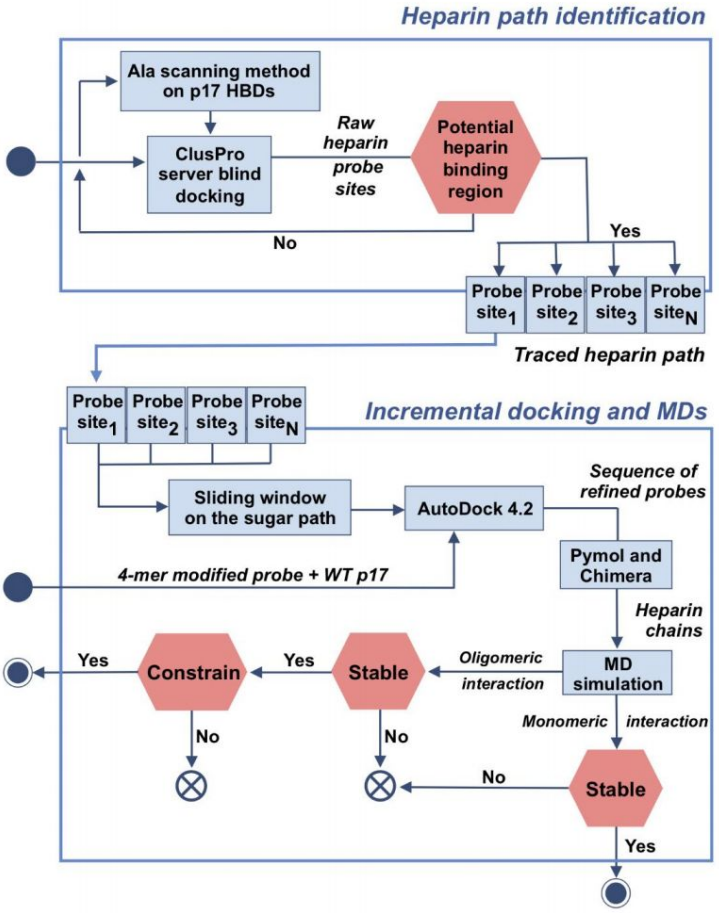



***Fig.04***. *Workflow of computational studies.*

Thus, we used a NC-ter K→A p17 dimer to predict alternative heparin-binding pose, allowing the definition of two alternative shorter and longer heparin paths. In the shorter path, the succession of 4-mer heparin probes starts from the N-ter HBD of p17 monomer A and goes directly to the N-ter HBD of p17 monomer B (Fig. 05A). Also in the longer path the succession of 4-mer heparin probes starts from the N-ter HBD of monomer A but it makes then contact with H4, C-ter basic domain and H3 of monomer A and with H3, C-ter basic domain, H4 and finally N-ter HBD of monomer B (Fig. 05B). Subsequently, the sliding window and incremental docking methods were used to identify and join the best binding poses of each 4-mer heparin in the traced paths (Fig. 05C), predicting two heparin models: the 15-mer heparin, modelled on the shorter path, and the 24-mer heparin modelled on the longer path (Fig. 06). In the complexes formed by p17 dimer with the 15- and 24-mer heparins, H-bonds are mainly formed with $SO_3^-$ or $COO^-$. In the 15-mer heparin complex H-bonds occur mainly within the N-ter HBD domain and α-helix 2 of both monomers (Fig. 06A and Tab. 01), while in the 24-mer heparin complex they occur in all the helices of both monomers and in the C-HBD (Fig. 06B and Tab. 02).

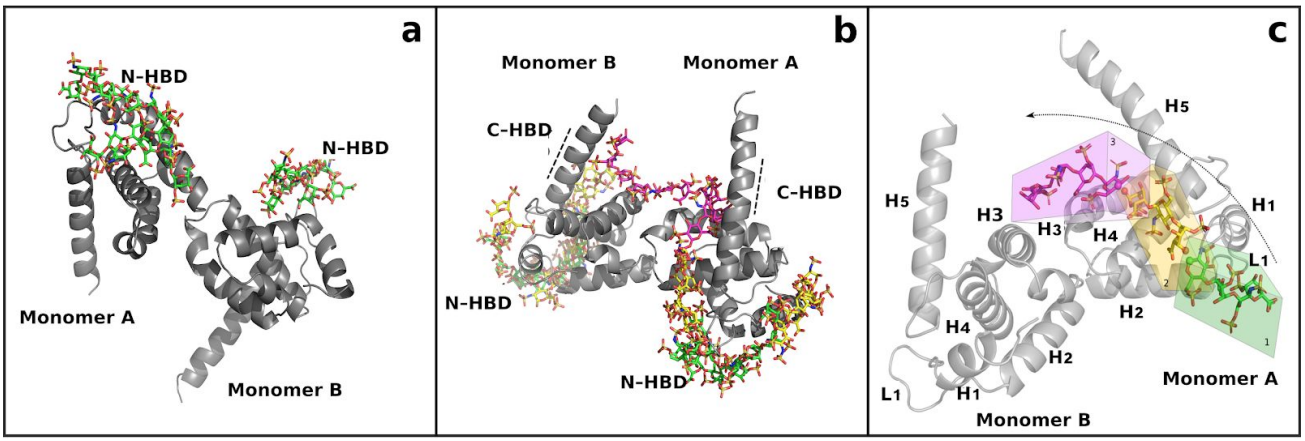



***Fig.05.*** *Shorter (**a**) and longer (**b**) heparin path identification by ClusPro web server. Position of the 4-mer heparin probes (in sticks) obtained by blind docking on the wt p17 dimer are in green, those obtained on N-ter K→A or NC-ter K→A p17 dimers are in magenta and yellow. p17 monomers are represented in grey cartoons. (**c**) Schematic representation of the sliding window method. Different regions composing the p17 protein are shown (H, α-helices; L, loops). The grid box of the first docked probe is in green, the second and third boxes, covering a whole 4-mer heparin and the last monomer of the precedent probe are in yellow and magenta, according to the heparin probes color. p17 monomers are represented as cartoons. For each heparin probe, in sticks, the atoms involved in 1→4 glycosidic linkages are represented as spheres.*

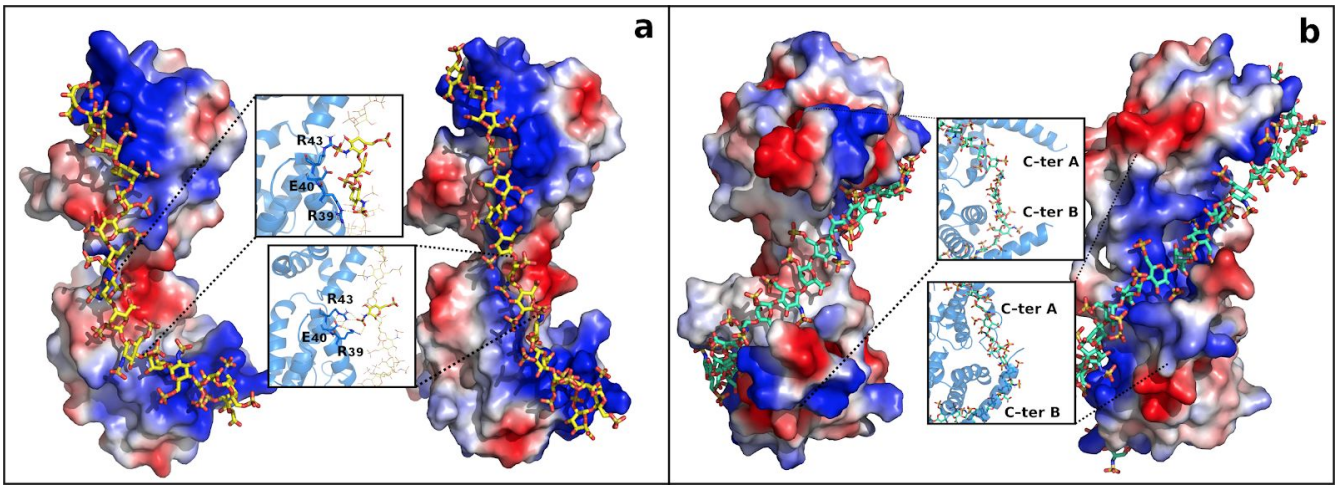



***Fig.06.*** *Electrostatic interactions. Docking prediction (lef) and complex stability after MD simulations (right) of 15-mer **(a)** or 24-mer **(b)** heparin (yellow and green sticks for 15- and 24-mer) in complex with p17 dimer (represented as electrostatic surface). Zoom insets: interactions before and after MD simulations are reported. P17 amino acids and monosaccharides involved in the networks are in sticks, blue and yellow.*

| | | docking | | MDs | |
|---|---|---|---|---|---|
| **heparin unit** | **p17 monomer** | **heparin group** | **p17 amino acid** | **heparin group** | **p17 amino acid** |
| **IdoA 1** | **A** | 2-O-$SO_3^-$<br>5-$O_{hc}$<br>6-$COO^-$ | $R_{22}$<br>$R_{22}$<br>$R_{22}$ | n.d. | n.d. |
| **Glc 2** | | N-$SO_3^-$<br>6-O-$SO_3^-$ | $K_{27}$<br>$R_{22}$ | 6-O-$SO_3^-$ | $R_{22}$, $K_{26}$* |
| **IdoA 3** | | 2-O-$SO_3^-$<br>3-OH | $K_{30}$<br>$K_{32}$ | 2-O-$SO_3^-$<br>3-OH | $K_{27}$<br>$K_{30}$, $K_{30}$ |
| **Glc 4** | | 5-$O_{hc}$<br>6-O-$SO_3^-$ | $K_{32}$<br>$K_{32}$ | 6-O-$SO_3^-$ | $K_{27}$, $K_{30}$, $K_{32}$, $H_{33}$ |
| **IdoA 5** | | 2-O-$SO_3^-$ | $R_{39}$ | 2-O-$SO_3^-$<br>3-OH | $R_{39}$<br>$K_{32}$ |
| **Glc 6** | | 6-O-$SO_3^-$ | $R_{39}$ | N-$SO_3^-$<br>3-OH | $R_{43}$<br>$R_{39}$ |
| **IdoA 7** | | n.d. | n.d. | 5-$O_{hc}$<br>6- $COO^-$ | $R_{43}$<br>$R_{39}$ |
| **Glc 8** | **B** | 3-OH<br>N-$SO_3^-$ | $R_{43}$<br>$R_{43}$ | N-$SO_3^-$ | $R_{39}$, $R_{43}$ |
| **IdoA 9** | | n.d. | n.d. | n.d. | n.d. |
| **Glc 10** | | N-$SO_3^-$<br>6-O-$SO_3^-$ | $R_{39}$<br>$W_{36}$ | 6-O-$SO_3^-$ | $K_{32}$ |
| **IdoA 11** | | 6- $COO^-$ | $K_{32}$ | 6- $COO^-$ | $K_{32}$ |
| **Glc 12** | | n.d. | n.d. | 6-O-$SO_3^-$ | $K_{30}$ |
| **IdoA 13** | | 3-OH | $K_{32}$ | 2-O-$SO_3^-$ | $K_{27}$, $K_{30}$, $K_{32}$ |
| **Glc 14** | | N-$SO_3^-$ | $K_{27}$ | N-$SO_3^-$ | $K_{27}$ |
| **IdoA 15** | | 2-O-$SO_3^-$ | $K_{30}$ | 2-O-$SO_3^-$ | $K_{26}$, $K_{28}$, $K_{30}$ |

***Tab. 01.*** *H-bond analysis in the 15-mer heparin/p17 dimer complex from docking or MD simulations. The 15-mer heparin makes stable contact with amino-lysine and guanidine-arginine groups of the residues $R_{22}$ (87%), $K_{26}$ (99%), $K_{27}$ (99%), $K_{32}$ (95%), $R_{39}$ (97%), $R_{43}$ (93%) of p17 monomer A and with $K_{26}$ (98%), $K_{27}$ (98%), $K_{28}$ (87%), $K_{30}$ (86%), $K_{32}$ (92%) of p17 monomer B. In brackets are the average persistency of the different H-bonds. P17 amino acids in loop 1 are in red, those in the N-ter HBD in blue and those in the -helix 2 in green. IdoA: 2-O-sulfated L-iduronic acid. Glc: N, 6-O-sulfateD-Glc. Hc: heterocyclic. n.d.: not defined. *H-bond with heparin is mediated by the N atom of the peptide bond. In all the other cases, N atoms of the side chain are instead involved.*

| | | docking | | MDs | |
|---|---|---|---|---|---|
| **heparin unit** | **p17 monomer** | **heparin group** | **p17 amino acid** | **heparin group** | **p17 amino acid** |
| **IdoA 1** | **A** | 2-O-$SO_3^-$ | $K_{28}$**, $K_{30}$ | 2-O-$SO_3^-$ | $K_{28}$** |
| **Glc 2** | | N-$SO_3^-$<br>6-O-$SO_3^-$ | $K_{26}$<br>$K_{27}$ | n.d. | n.d. |
| **IdoA 3** | | 2-O-$SO_3^-$ | $K_{26}$ | 3-OH<br>6-$COO^-$ | $K_{26}$<br>$K_{27}$ |
| **Glc 4** | | 6-O-$SO_3^-$ | $R_{22}$ | 6-O-$SO_3^-$ | $R_{22}$ |
| **IdoA 5** | | n.d. | n.d. | 2-O-$SO_3^-$ | $R_{22}$ |
| **Glc 6** | | 6-O-$SO_3^-$ | $K_{98}$ | N-$SO_3^-$ | $K_{98}$ |
| **IdoA 7** | | 2-O-$SO_3^-$ | $R_{76}$<br>$N_{80}$ | 5-$O_{hc}$<br>6-$COO^-$ | $K_{98}$<br>$K_{98}$ |
| **Glc 8** | | 6-O-$SO_3^-$ | $Q_{69}$ | 6-O-$SO_3^-$ | $N_{80}$ |
| **IdoA 9** | | 3-OH | $Q_{65}$ | 2-O-$SO_3^-$<br>5-$O_{hc}$<br>6-$COO^-$ | $R_{76}$<br>$K_{112}$<br>$K_{112}$ |
| **Glc 10** | | N-$SO_3^-$ | $Q_{69}$ | 1-OH<br>N-$SO_3^-$<br>6-O-$SO_3^-$ | $K_{112}$<br>$R_{76}$<br>$Q_{69}$ |
| **IdoA 11** | **A and B** | 2-O-$SO_3^-$ | $Q_{65}$ (monomer B) | 5-$O_{hc}$<br>6-$COO^-$ | $K_{114}$*<br>$K_{114}$ (monomers A and B) |
| **Glc 12** | | N-$SO_3^-$ | $R_{58}$ (monomer B) | N-$SO_3^-$<br>6-O-$SO_3^-$ | $K_{114}$ (monomer B)<br>$A_{115}$*, $Q_{116}$* (monomer A) |
| **IdoA 13** | **B** | 6-$COO^-$ | $Q_{108}$ | 5-$O_{hc}$<br>6-$COO^-$ | $K_{112}$<br>$Q_{65}$, $Q_{69}$, $K_{112}$ |
| **Glc 14** | | N-$SO_3^-$ | $K_{112}$ | 1-OH<br>N-$SO_3^-$<br>4-OH | $K_{112}$<br>$K_{112}$<br>$Q_{69}$ |
| **IdoA 15** | | 2-O-$SO_3^-$<br>3-OH | $E_{105}$*<br>$Q_{65}$ | 2-O-$SO_3^-$<br>3-OH | $Q_{65}$, $Q_{69}$<br>$Q_{69}$ |
| **Glc 16** | | N-$SO_3^-$<br>6-O-$SO_3^-$ | $N_{80}$<br>$Q_{69}$, $R_{76}$ | N-$SO_3^-$<br>3-OH | $R_{76}$, $K_{98}$<br>$K_{98}$ |
| **IdoA 17** | | n.d. | n.d. | 2-O-$SO_3^-$<br>6-$COO^-$ | $R_{22}$, $N_{80}$<br>$K_{98}$ |

| | | | | | |
|---|---|---|---|---|---|
| **Glc 18** | | N-$SO_3^-$<br>6-O-$SO_3^-$ | $R_{76}$<br>$R_{22}$ | 1-OH | $R_{22}$ |
| **IdoA 19** | | 2-O-$SO_3^-$<br>3-OH<br>5-$O_{hc}$ | $N_{80}$<br>$E_{73}$**<br>$R_{22}$ | 2-O-$SO_3^-$ | $R_{22}$, $S_{77}$**, $N_{80}$ |
| **Glc 20** | | N-$SO_3^-$<br>6-O-$SO_3^-$ | $K_{27}$<br>$H_{33}$ | N-$SO_3^-$ | $K_{27}$ |
| **IdoA 21** | | 2-O-$SO_3^-$<br>5-$O_{hc}$<br>6-O-$SO_3^-$ | $K_{30}$<br>$K_{27}$<br>$K_{27}$ | 2-O-$SO_3^-$<br>5-$O_{hc}$<br>6-O-$SO_3^-$ | $K_{30}$, $K_{32}$<br>$K_{27}$<br>$K_{27}$ |
| **Glc 22** | | 1-OH<br>2-O-$SO_3^-$<br>6-O-$SO_3^-$ | $K_{30}$<br>$K_{32}$<br>$K_{27}$ | 3-OH | $K_{30}$ |
| **IdoA 23** | | 6-$COO^-$ | $K_{30}$ | 2-O-$SO_3^-$<br>6-$COO^-$ | $K_{26}$<br>$K_{28}$ |
| **Glc 24** | | N-$SO_3^-$<br>6-O-$SO_3^-$ | $K_{28}$<br>$K_{26}$ | 6-O-$SO_3^-$ | $R_{22}$, $K_{28}$ |

***Tab. 02.*** *H-bonds analysis of the 24-mer heparin/p17 dimer complex from docking or MD simulations. The negatively charged $SO_3^-$ of the 24-mer heparin create a stable network of interactions with monomer A: [$R_{22}$ (99%), $K_{26}$ (98%), $K_{27}$ (99%), $K_{28}$ (80%) of the N-ter HBD, $R_{22}$ (99%), $K_{27}$ (99%), $Q_{65}$ (75%), $Q_{69}$ (80%), $R_{76}$ (75%) of the globular domain, $Q_{65}$ (75%), $K_{98}$ (89%), $K_{112}$ (96%), $K_{114}$ (80%) of the C-ter] and with monomer B [$R_{22}$ (>99%), $K_{27}$ (>99%) , $K_{28}$ (70%), $K_{30}$ (95%), $K_{32}$ (75%) of N-ter HBD, $Q_{69}$ (65%), $R_{76}$ (85%), $S_{77}$ (85%), $N_{80}$ (99%) of the globular domain and $K_{98}$ (96%), $K_{112}$ (98%), $K_{114}$ (90%) of the C-ter]. In brackets are the average persistency of the different H-bonds. P17 amino acids in loop 1 are in red, those in the N- or C-ter HBDs in blue, those in the -helix 2, 3, 4 and 5 in green, violet, cyan and brown, respectively. IdoA: 2-O-sulfated L-iduronic acid. Glc: N, 6-O-sulfateD-Glc. Hc: heterocyclic. n.d.: not defined. H-bond with heparin is mediated by the N (*) or O (**) atom of the peptide bond. In all the other cases, N atoms of the side chain are instead involved.*

MD simulations were performed on both the systems to evaluate stability and conformational drift and to refine the two heparin/p17 dimer complexes.

Root mean square fluctuation (RMSF) analysis of the 15-mer heparin/p17 dimer complex shows that the C-termini of both monomers are the most variable regions (fluctuation up to 5Å), in agreement with relaxation and amide proton exchange studies (64) (Fig.07A). Three other flexible regions span from 7–14, 20–30 and 60–75 (Fig. 07A). The 15-mer heparin backbone is stable, while its $SO_3^-$ are endowed with a higher fluctuation due to solvent exposure. In particular, monosaccharides from 10 to 14 exhibit the highest fluctuation (up to 5Å), due to their interaction with N-ter HBD of p17 monomer B (Fig. 07B). Root mean square deviation (RMSD) of the p17 C-α globular domain is stable during the simulations (average fluctuation=0.5Å) while the C-ter of both the monomers, due to solvent exposure, reach stability only after 25 and 45ns. Heparin remains stable during the simulation (average fluctuation=2Å). It also reaches a second stability plateau after 28ns due to the fitting upon binding of the 8$^{th}$ monosaccharide (Fig. 07C). Electrostatic surface analysis of docking model shows that 15-mer heparin sets up an articulate network of interactions inside the cationic N-ter HBD of monomer A and B, between which is interposed a negative electrostatic surface potential generated by the exposure of $COO^-$ of $E_{40}$, by the N-$SO_3^-$ and 3-$OH^-$ of the 10$^{th}$ and 8$^{th}$ Glc which are engaged in H-bonds with $R_{39}$ and $R_{43}$ of α-helix 2 of monomer B (Fig. 07A and Tab. 01). However, MD simulations show that, after 20 ns, these two H-bonds are lost due to solvent exposure of N-$SO_3^-$− and 3-$OH^-$, while a new H-bond network is formed between the $COO^-$ of $E_{40}$ and the

guanidinium groups of $R_{39}$ and $R_{43}$ and between the N-$SO_3^-$ of the 10th Glc and the guanidinium group of $R_{39}$ (Fig. 06A and Tab. 01). This new arrangement induced by the heparin conformational drift leads to the formation of a continuous positive channel in the p17 dimer structure (Fig. 06A). H-bond analysis shows that the 15-mer heparin makes an articulate network of H-bonds with the p17 dimer (Tab. 01). Worth of mention, $K_{30}$ interacts with 2-O-$SO_3^-$ and 3-$OH^-$ of the 3rd IdoA and with 6-O-$SO_3^-$ of the 4th Glc with an average persistency of 65%, while H33 interacts with 6-O-$SO_3^-$ of the 4th Glc with an average persistency of 50%. Interestingly, the H-bond persistence of the last two residues together is higher than 80%, indicating a synergic effect mainly due to the groups of the 3rd IdoA. Furthermore, $R_{39}$ and $R_{43}$ of monomer B interact simultaneously with $E_{40}$ and heparin (persistency=50% for both), further stabilizing the interaction network between the protein and the sugar.

RMSF of the 24-mer heparin/p17 dimer complex shows that the C-ter of p17 monomer B displays a lower fluctuation in respect to that in the 15-mer heparin/p17 dimer complex, due to a better H-bond network and a lower exposure to solvent, in particular in the second half of heparin (monomers 12–24) (Fig. 07D and Tab. 02). As already highlighted in the 15-mer heparin/p17 dimer complex (Fig. 07E), a higher fluctuation can be instead appreciated for monomer A binding to the first half of heparin (monomers 1-12) for which a greater fitting needs to be induced to generate the positive channel in the protein. More into the details, the first four monosaccharides of heparin, that interact with the N-ter of monomer A, are endowed with the higher fluctuation, with their $SO_3^-$ showing an average fluctuation up to 2Å, likely due to their solvation. RMSD demonstrates that the p17 C-α globular domains in the dimers are stable during the simulation, with an average fluctuation=1Å. This stability is shared also by the C-ter of monomer B, while the C-ter of monomer A and heparin reach stability only after 35ns (Fig. 07F). Electrostatic analysis of the docking model shows that some $SO_3^-$ of 24-mer heparin take contact with the N-ter basic motif of both the p17 monomers and with the C-ter of monomer B (Fig. 06B and Tab. 02) while some others remain instead exposed to the solvent. However, MD simulations show that these free $SO_3^-$, by inducing molecular drifts, succeed in taking contact with basic residues of the C-HBDs of both the p17 monomers. Thus, the heparin-induced conformational drift of p17 monomers creates sandwich-like structures that lead to the formation of the continuous positive channel, already described in literature, that prevents solvent exposition of bound heparin (Fig. 07B and Tab. 02). The results of the H-bonds analysis of the 24-mer heparin/p17 dimer complex are reported in Tab. 02. Considering the whole simulation, residues $A_{115}$ and $Q_{116}$ of the C-ter of monomer A, that exhibit low H-bond persistency with heparin (55 and 29%), increase their persistence after 35 ns to 99% and 85%, due to binding-induced stabilization of both heparin and p17 C-ter.

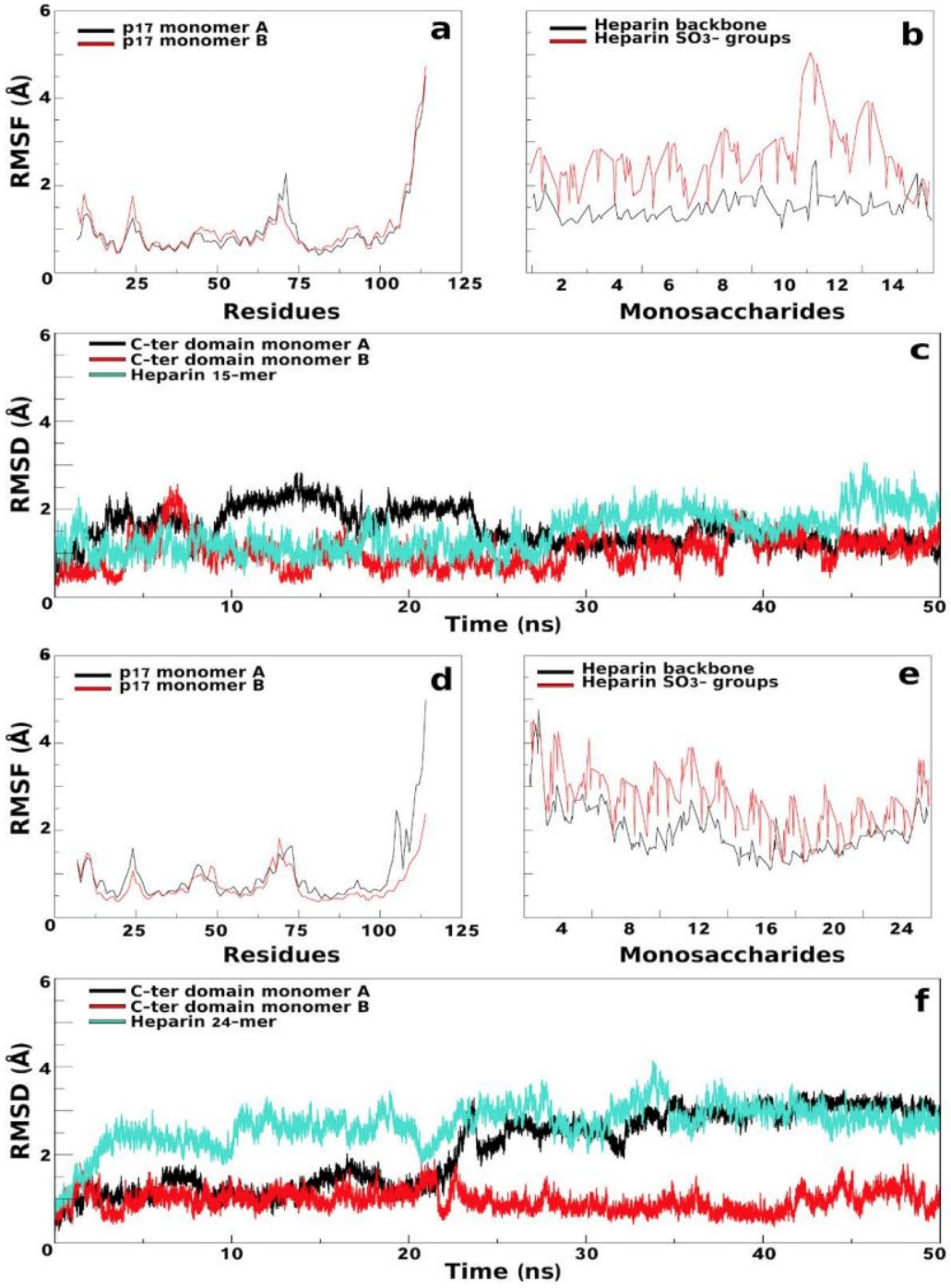



***Fig. 07****. RMSD and RMSF analyses were performed to evaluate the stability and conformational drift of the complexes. RMSD was used to evaluate the stability of unbound p17 (considering its globular domain and C-terl region separately), and heparin (considering its backbone and sulfate groups separately) while RMSF were calculated for single p17 amino acids or for heparin atoms. Starting from the trajectories and using CPPTRAJ, the most representative cluster and Hbond were assessed for p17/heparin interactions and stability. In figure are reported RMSF profile of p17 dimer in complex with 15-* ***(a)*** *or 24-* ***(d)*** *and of 15-mer* ***(b)*** *or 24-mer* ***(e)*** *heparin calculated per atoms and displayed per monosaccharides. RMSD along the simulation are also reported for 15-mer (c) or 24-mer* ***(f)*** *heparin in complex with the p17 dimer.*

## 2.1.4. Conclusions

The new computational protocol here adopted discloses the possibility that heparin can connect two p17 monomers with different binding modes depending on their lengths and that, during interaction, heparin induces dynamic conformational drift of the p17 dimer that facilitate the interaction between the GAG and the protein monomers, stabilizing the dimer and likely promoting further oligomerization. Protein oligomerization is involved in a variety of biological processes. It can occur in the form of "simple" dimers or higher order oligomers of cytokines, chemokines and growth factors that favour receptor clustering at the cell surface and transduction of the signal that, in turn, trigger cytoskeletal rearrangements, cell movement and other cellular responses including proliferation and survival involved in physiological and pathological processes such as immune responses, angiogenesis and cancerogenesis (65). Differently, some "amyloidogenic" proteins are prone to change conformation from an α-helix to β-sheet, tending to misfold and to form specific large aggregates responsible of the pathogenesis of many human diseases, including neurodegenerative pathologies such as Alzheimer and Parkinson' disease (66).

As already mentioned, the C-ter portion and amino acids $E_{42}$-$N_{47}$, $Q_{59}$, $Q_{63}$, and $Q_{69}$-$E_{74}$ of p17 mediate its oligomerization. Our MD simulations predict a heparin-induced drift of p17 monomers that leads to the resetting of globular domains and loss of interaction only between $A_{45}$ of monomer A and $S_{72}$ of monomer B that, in turn, induces the alignment of both the α-helix 2. Thus, heparin could bring together multiple copies of p17 through the relatively high affinity p17/heparin interaction (Kd ≅100 nM) (49), leaving exposed the p17 self-assembly regions on adjacent proteins as to facilitate the low affinity p17/p17 interaction (Kd ≅ 600nM). Also, it is possible that the formation of high order oligomers (trimer and tetramers) can be triggered by the initial assembly of the p17 dimer induced by heparin. In effect, oligomerization is often the outcome of a cooperative interaction process, during which a first binding makes the following ones (52). Thus, the action of heparin may consist in providing the appropriate scaffold to facilitate and stabilize the relatively weak p17/p17 interaction, as already observed for different other factors and cytokines (67).

Also, computational models for heparin/protein interaction have been so far restricted to short chains aimed at the identification of HBDs in proteins. Here, we have set up a new "step by step" protocol (Fig. 04) by which it has been possible to identify two heparin paths and then create two computational models of a p17 dimer in complex with 15- and 24-mer heparins whose validity is confirmed by their good agreement with the experimental data. Indeed, docking and MD simulations predict 15-mer and 24-mer heparin to bind to p17 dimer through an articulate network of interactions occurring mainly between the negative $SO_3^-$ of heparin and the positive amino acids of the proteins. Accordingly, these interactions have been experimentally demonstrated to mediate heparin/p17

interaction and p17 oligomerization by experiments with NaCl, desulfated heparin and the p17 K→A mutant.

MD simulations studies with the heparin models disclosed the capacity of heparin to interact with a p17 dimer with different binding modes depending on its length. Also, heparin can induce a "fitting-upon-binding" on the p17 dimer that causes the exposition of positively charged residues on the globular domains and the warping of both the C-ter to create a sandwich-like structure around heparin. Finally, RMSF showed that when complexed with the 24-mer heparin, the p17 dimer complex displays a lower fluctuation in respect to when complexed with 15-mer heparin, suggesting that the longer is heparin, the stronger is its stabilizing effect on the p17 dimer. MD simulations also show that, during the binding process, heparin fits together with the protein, simulating an annular-like structure already described in literature (44).

This computational protocol could be useful to understand at an atomic level the role of heparin in the processes of oligomerization of a variety of cytokines, chemokines and growth factors, and even in the formation of ternary complexes among heparin, heparin-binding proteins and their receptors.

## 2.2. New insight into the mechanistic effect of heparin and HSPGs on SARS-CoV-2 S glycoprotein

*The results from this work have been collected in a paper submitted to PNAS:* ***Paiardi G,*** *Rusnati M, Wade R.C "Mechanism of inhibition of Sars-CoV2 infection by the interaction of the spike glycoprotein with heparin".*
*This part of the work has been funded by Prace (Partnership for advanced computing in Europe) and carried out during the endorsement at the Heidelberg Institute for Theoretical Studies (HITS). Heidelberg - Germany.*

### 2.2.1. Introduction

In the last year, a sudden effort has been made by researchers to tackle the COVID-19 pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which has affected the lives of all people around the world.

SARS-CoV-2 virus is a lipid-enveloped positive-sense RNA virus belonging to the *Coronaviridae* family (68). Among the SARS-CoV-2 proteins, the spike S glycoprotein (spike) is highly conserved in the *Coronaviridae* family (76% and 96% amino acid similarity with SARS-CoV and BatCoV-RaTG13, respectively (69)) and mediates the virus entry into the human cells by binding to the angiotensin-converting enzyme 2 (ACE2) receptor (70). The prefusion spike is exposed on the virion surface as a homotrimer. Each spike subunit is composed of two domains, S1 and S2, connected by the S1/S2 junction, and involved in virus attachment and fusion, respectively (70). The S1/S2 junction, novelty feature of the spike SARS-CoV-2, consists in a multibasic sequence that is cleaved by the host furin protease that, in turn, activates the post-fusion conformation of spike necessary to SARS-CoV-2 entry (71). Finally, spike possess 22 N-linked glycans likely involved in protein stability and immune evasion of the virus (72).

Increasing experimental evidences suggest an essential role for heparan sulphate proteoglycans (HSPGs) as co-receptors that, by binding spike, favour SARS-CoV-2 attachment to human cells (73, 74). The binding of spike to HSPGs is mediated by cluster of positively charged amino acids present within the protein (the so called Cardin-Weintraub motifs or heparin binding motifs, HBDs), and the negatively charged sulfated groups present on the saccharidic chains attached to the HSPG core protein (9). To date, three HBDs have been identified in the spike sequence: in the RBD (RBD-HBD) and at the S2' site (both present in SARS-CoV-1) and the third HBD at the novel S1/S2 furin cleavage site (S1/S2-HBD) (74) (Fig.08A).

Heparin is a polysulfated glycosaminoglycan structurally similar to the saccharidic chain of HSPGs. It has been firstly used to treat COVID-19 patients for its strongly anticoagulant activity (75). However, due to its HSPGs-antagonist capacity (73, 74), it has also been taken in consideration for its direct antiviral effect (75). In respect to many other HSPGs-dependent viruses, for which the

mechanism of antiviral activity of heparin has been already deeply comprehended (76), for Sars-CoV2 this kind of study has been only recently started. Moreover, spike displays the peculiar feature to possess three different HBD overlapped to motifs endowed with distinct functions (RBD and furin cleavage site), making this area of research multifaceted and interesting but challenging from an experimental perspective.

Molecular models have become practically mandatory for a comprehensive knowledge of the physiological and pathological processes mediated by macromolecular interactions including virus infection (77). With these premises, here we report a thorough computational study whose results provide insight in the heparin/spike binding mode(s) useful to guide future experimental research, design novel heparin derivatives and better understand the mechanism by which HSPGs act as host-cell co-receptors.

## 2.2.2. Material and methods

*Model systems.* The models of the head SARS-CoV-2 S protein in closed and open conformations were released from the SwissModel website (https://swissmodel.expasy.org) and are based on PDBid 6ACC (seq. identity 76.47%) and PDBid: 6ACD (seq. identity 76.47%), respectively. The models were refined adding 18 N-glycans covalently attached in accordance with the glycomic profile (72) using the GLYCAM web server (http://glycam.org/) to obtain the S glycoprotein model. Protonation states were assessed using PROPKA3 (78). Heparin were designed using incremental docking and sliding window method (37) spanning from the S1/S2 multibasic site to the receptor-binding domain (RBD).

*All-atom MD simulations*. The Amber20 package (79) was used to run 1 μsecond of all-atom MD simulations in explicit solvent and four replicas for each model were carried out. S glycoprotein was parameterized with ff14SB force field (80) and GLYCAM-06j force field (26). Heparin was parameterized following the method published (37). All the systems were placed in a periodic-cubic water box using the TIP3P water model with 10Å between the systems and the edge of the box. The systems were salted and neutralized with NaCl to reach a concentration of 150mM in a neutral ph7 with standard protonation for the residues. Each system was minimized with 14 minimisation steps with decreasing positional restraints (from 1000 to 0 kcal/mol A2) with a cut-off for non-bonded interactions of 8 Å to remove bad contacts due to the modelling procedure. Two steps of heating (from 10 to 100 and from 100 to 310K°) in an NVT ensemble with Langevine thermostat were used to warm the systems. The systems were equilibrated in 4 steps of 2.5ns each one (2 fs time step) in NPT ensemble with Langevine thermostat in a periodic boundary condition. For each model, 4 independent replicas were carried out starting from randomly restart files by the last 5ns of equilibration. During the simulation, a cutoff of 8 Å for the evaluation of short-range non-bonded interactions and the

Particle Mesh Ewald method for the long-range electrostatic interactions have been used. The temperature was kept constant at 300 K° with Langevin thermostat. Bonds involving hydrogen atoms were constrained with the SHAKE algorithm.

*Hardware.* Thanks to the funding of computing time (3.52 million 'Prace' CPU hours) by PRACE (Partnership for Advanced Computing in Europe) (https://prace-ri.eu/), MD simulations were carried out on the accelerated cluster Marconi100 based on IBM Power9 architecture and Volta NVIDIA GPU. Marconi100 is equipped with a total of 8 nodes of 4xNVIDIA Volta V100 GPUs, Nvlink 2.0, 16 GB. The systems and the analysis were set up and analyzed on the in-house clusters equipped with high-performance clusters with both GPUs and CPU cores.

*Analysis.* MD trajectories were collected using AmberTools (79) and checked by visual inspection in Visual Molecular Dynamics (VMD) (81).

*Cluster analysis* was carried out along the last 100ns of all the MD simulations considering C-α, N and C for the protein residues and C, N and O for the heparins. N-glycans were excluded by the analysis. The hierarchical agglomerative (bottom-up) approach was used with a minimum distance between the clusters greater than 3.0 and using the average distance between members of two clusters.

*Hydrogen bond (H-bonds) analysis* was calculated along all the trajectories and all the replicas with 10 frames of time step and setting 3.5 Å as criteria for distance between heavy atoms. All the atoms of the systems were considered. Interaction fingerprint analysis (IFP) was calculated using the python script released by Kokh et al (82) to confirm the H-bond analysis. The interactions were collected along all the trajectory with a time step of 100 frames between the S protein and heparin.

*Root Mean square deviation (RMSD)* was calculated on C- α of the single monomers $S_A$ $S_B$ $S_C$ considering all the amino acids and on all the C, O and N of heparin. RMSD of the hinge regions was calculated on the C-α for residues 501-503 for $S_A$ $S_B$ $S_C$ separately.

*Solvent Accessible surface area (SASA).* Two separate SASA analyses were conducted on $S_A$ $S_B$ $S_C$ monomers separately: along the trajectory using AmberTools (79) and on the most representative clusters using Naccess (83). In both the analysis the Van der Waals radii considered was of 1.4 Å. Residues 682-685 were considered both along the trajectory and in the cluster in case of the S1/S2 multibasic site. In case of the receptor binding motif, all the residues of the domain (residues 437-510) were considered along the trajectory and only the residues suggested by Lan and co-workers (84) on the representative clusters.

*Dihedral Principal Component Analysis (dPCA).* The dihedral covariance matrix and the projection was calculated for the backbone of the phi/psi angles of residues 501-503 of $S_C$ monomer. The first four eigenvectors and eigenvalues were extracted and PC1 and PC2 were plotted for all the systems. All the systems were transformed into the same PC space to evaluate the simulation variance across the replicas.

*Essential Dynamics (ED)* were calculated with the PCA on the unbiased MD simulations for all the systems considering S protein and heparin along the trajectory. The main motion models were collected and visualized using VMD.

### 2.2.3. Results

- **Long mostly basic patches on spike glycoprotein accommodates heparin**

To characterize the potential effects exerted by an heparin treatment or by spike binding to cell-associated HSPGs, we take in consideration the active and inactive conformations of the homotrimeric spike head. The former, also defined closed conformation presents the three RBD of subunits $S_A$, $S_B$ and $S_C$ not available for the binding with ACE2 receptor, hereafter defined down-subunits. The latter, also called open conformation, presents the RBD of only one subunit ($S_C$-RBD) available for the binding to ACE2, hereafter defined as up-subunit. Based on these considerations, we set up a set of models of spike: (i) in closed conformation; (ii) in closed conformation with a single heparin chain bound; (iii) in closed conformation with three heparin chains bound; (iv) in open conformation; (v) in open conformation with three heparin chains bound.

The physiological likelihood of our models have been maintained adding the 18 N-glycans per subunit (72) covalently attached to the head portion of the SARS-CoV-2 spike (Fig.08) available at the SwissModel website (https://swissmodel.expasy.org) . Then, for modelling the heparin bound to glycosylated spike, we took into account the putative HBDs that were accessible to the solvent in the prefusion conformation of the spike, thus excluding the solvent-inaccessible S2'-HBD (Fig.08A) (7). To identify a continuous positively charged path on the protein surface at which the long heparin chains could bind, the electrostatic potential of spike was computed in both closed and open conformations. This analysis suggests that a heparin chain can follow a similar path in the two conformations of the spike head differing only in the interactions in the RBD-HBD (Fig.08B). The modelled polyanionic heparin chains run along mostly basic patches, partially through surface grooves, from the S1/S2 basic motif (R682, R683, R685), via the channel between the N-terminal domain of the same spike subunit and the RBD-HBD of an adjacent spike subunit (R346, R355, K356, R357).

Due to the structural similarity between heparin and HSPGs, it is expected that also the latter accommodate in the basic paths identified along the trimeric spike, inferring the possibility that, by burying these basic regions, heparin hinders the binding of HSPGs, reducing the amount of SARS-CoV-2 tethered to the cell surface, decreasing the subsequent binding to the ACE2 receptor. This model provides a mechanistic explanation for the experimental data reported by Esko and co-workers that demonstrated that heparin can prevent SARS-CoV-2 infection (73).

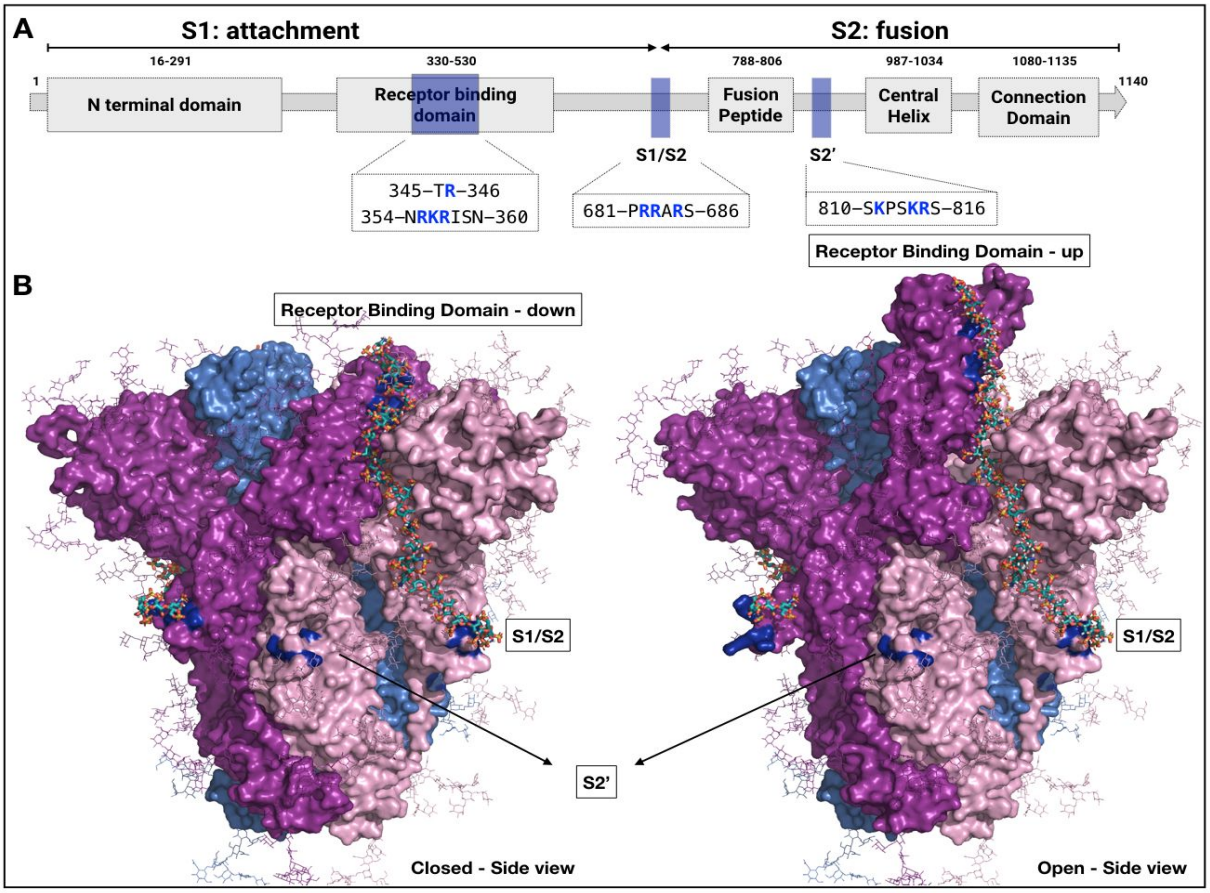



***Fig. 08**. Modeling of spike-heparin interactions. (A) Schematic diagram of the sequence of the SARS-CoV-2 spike (S) glycoprotein head (residues 1-1140), which is composed of S1 and S2 subunits. The boxes along the sequence show the positions of the main protein domains: N-terminal domain (16-291), receptor binding domain (RBD) (330-530), S1/S2 site (681-686), fusion peptide (788-806), S2' domain (810-816), central helix (987-1034) and connection domain (1080-1135). Putative HBDs are indicated by blue boxes with their corresponding sequences. **(B)** Side views of the spike head homotrimer models before the simulations in closed (left) and open (right) conformations with 3 heparin chains bound. Only one complete heparin chain can be seen from this view. The $S_A$ $S_B$ and $S_C$ subunits are shown as surfaces in blue, pink and magenta, respectively. The dark blue surface patches correspond to the putative HBDs. N-glycans covalently attached to the spike are shown in line representation, colored according to the subunit to which they are attached. The 31-mer heparin chains that span from the S1/S2 site HBD to the RBD-HBD are shown in stick representation, coloured by element with cyan carbons.*

- **Heparin stabilises the spike glycoprotein in closed conformation**

The models described in the previous section were simulated in multiple 1μs all-atom MD simulations. In Fig.09A, representative clusters of spike in closed and open conformations with 3 heparin chains bound are shown as electrostatic potential surfaces computed without N-glycans.

The structures of spike reached convergence within ~2-400ns in all simulations, as shown by the root mean square deviation (RMSD) relative to each starting structure (Fig.09). Moreover, when in complex with heparin, spike shows a decreased RMSD smaller by about 1Å, indicating that the

binding to the glycosaminoglycan induces a stabilization of the homotrimer (Fig.09B). These data were confirmed by root mean square analysis (RMSF) that shows an overall decrease of fluctuation of the RBD and of the S1/S2 site in spike with heparin bound. Accordingly, RMSD of heparin indicates an induced fit along the trajectory (Fig.09B). Comparing the RMSD and the electrostatic potential of closed and open spike bound to heparin, our models suggest an induced fitting upon binding that results in a well-defined partially grooved basic path where heparin is laid (Fig.09A)

Overall, the H-bond and MD-IFP analysis of all the MD simulations suggest that each heparin chain maintains stable interactions with 2 consecutive subunits in both the closed and open conformations of spike (Fig09A). Heparin binds through H-bonding interactions to the basic residues of the S1/S2-HBDs in the first subunit and to the RBD-HBDs of the second subunit throughout all the MD trajectories in both the active and inactive state (>90% occupancy) (Fig09A). However, along the heparin path, additional, less specific binding regions, which differ between the open and the closed conformations, were identified in over 75% of the MD simulations that can further stabilize the complexes (Fig09A).

In the closed spike model, the binding of one or three heparin chains invariably hinders the opening of the spike by stabilizing the closed conformation through the simultaneous binding of the RBD of one subunit (T345, R346, N354, R355, S359, N360, N450) and the N-ter domain of the adjacent subunit (N165, C166, T167, E169, V171, Q173, F220, N280, N282, T284, T286) (Fig09A). In this conformation, the RBDs should be harnessed by heparin and the spike activation should be prevented. Finally, heparin binding is also stabilised by interacting with residues near to the multibasic site (N606, S686, S689, S690).

On the other hand, the models of spike in the open state suggest that heparin does not induce allosteric effects that favour closure towards the inactive state. However, through the simultaneous binding to the up-RBD of subunit $S_C$ (T345, R346, N354, R355, R357, N360) and the N-terminal domain of the adjacent subunit (R34, T167, E169, Q173, L176, R-90, H207, T208, F220, S221), heparin induces in the open subunit a change of its orientation (Fig02A, Fig.S3, Tab.S2 and Tab.S3) further investigated in the next section. As for the closed models, some polar residues near to the multibasic site (N606, Y674, S686, Q689, S690) permanently interact with the polyanionic chain (Fig09A).

Taken together, these observations suggest that, due to the structural similarity with heparin, HSPGs may act in the same way, effectively increasing the concentration of SARS-CoV-2 on the host cell surface but hampering the opening and activation of closed spike and not favouring the closure of the open conformation.

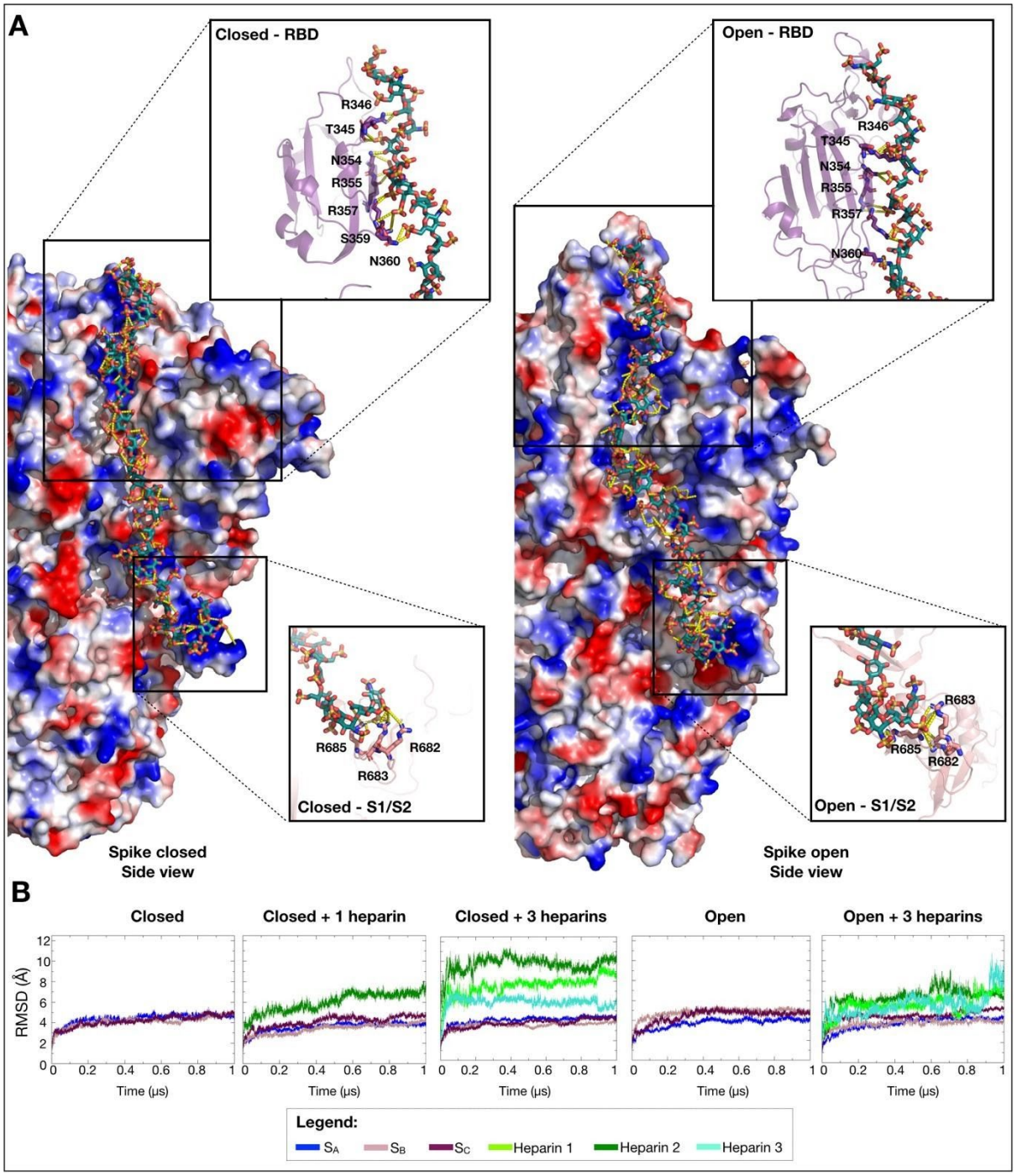



***Fig. 09.*** *Stability of the spike-heparin complexes. (A) Closed (left) and open (right) structures of spike bound to three heparin chains are displayed as molecular surfaces with electrostatic potential to show the partially grooved positively charged path occupied by heparin. Heparin is shown as sticks and colored by element with cyan carbons. The yellow dashes show the H-bonds interactions between spike and heparin. The insets highlight the H-bond interactions between heparin and the residues in the RBD* (T345, R346, N354, R355, N360) *and S1/S2 HBDs* (R682, R683, R685) *shown in stick representation with carbons colored according to the subunits to which they are bound both in the closed and open conformations.* ***(B)*** *Structural convergence of the five simulated systems. RMSD vs time is shown for the system components for one replica MD simulation for the indicated modelled systems.*

- **Heparin masks the S1/S2 furin cleavage site**

To assess the role of the novel S1/S2 furin cleavage site (71, 85) identified as a novel putative HBD in the spike of SARS-CoV-2, H-bond formation was monitored along the trajectories and surface exposure to the solvent was analysed for representative clusters (Fig.10).

In the initial complexes, the first monosaccharides of the heparin chain interact with the basic residues of the S1/S2-HBD. Salt-links with R682, R683 and R685 were maintained along all the trajectories (>90% occupancy), indicating strong interactions of the S1/S2-HBD with heparin (Fig.10A). The calculated solvent accessible surface area (SASA) of this multibasic site shows the persistence of the interactions along the simulation of the closed conformation (Fig.10B). Heparin halves the surface exposed in the closed models by binding directly to the S1/S2-HBD and the three heparin chains succeed in masking simultaneously all three multibasic sites on the closed spike trimer. In the open conformation, the reduced SASA in presence of heparin indicates a significant shielding effect primarily at the S1/S2-HBD of the up-subunit although the heparin chains maintain a lower shielding level on the same domains of the down-subunits. This difference in behaviour could be due to a more favourable conformation of the basic patches between the up-and-down subunits as compared to that between two-down subunits.

To further assess the shielding effect of heparin related to spike glycosylation, the SASA of the S1/S2-HBD calculated for the representative clusters was decomposed into the area exposed without consideration of the N-glycan and heparin sugars, the area exposed accounting for the N-glycans, and the area exposed accounting for both the N-glycans and heparin (Fig.10C). In agreement with the previous analysis, these calculations indicate that heparin directly binds to the S1/S2-HBDs halving the exposed surfaces in both the closed and open conformations. Again, on the open conformation, the shielding effect exerted by heparin on the S1/S2-HBDs of the two down-subunits is lower than on the same site of up-subunit. Moreover, the decomposition shows that the N-glycans of spike provide little contribution to the shielding of the multibasic sites. In summary, when comparing the binding of single or three heparin chains to the spike homotrimer, all the data indicate the ability of heparin to permanently occupy and shield the S1/S2 site without a significant contribution of N-glycans.

Based on these findings, we propose a dual role for the S1/S2 domain in spike activity. As already demonstrated, it acts as an essential furin cleavage site for the activation of the fusion conformation of spike (71). In addition, our MD simulations demonstrate a key role in the binding to heparin and possibly to HSPGs. The SASA analysis shows that the S1/S2-HBD is exposed to the solvent and available to interact with heparin or HSPGs, supporting the hypothesis that the this novel HBD may contribute to the increased affinity of SARS-CoV-2 spike for heparin/HSPGs compared to

previous coronavirus strains. Finally, we observed that, in the closed spike model, a single heparin chain succeeds in occupying the S1/S2-HBDs of only one subunit, without inducing allosteric or mechanistic effects on the other two subunits. This suggests that, to exert its full antiviral activity, heparin needs to be administered at doses as high as to saturate all the HBDs of the spike, otherwise leaving some of them still available for the tethering to cell-associated HSPG and consequent infection.

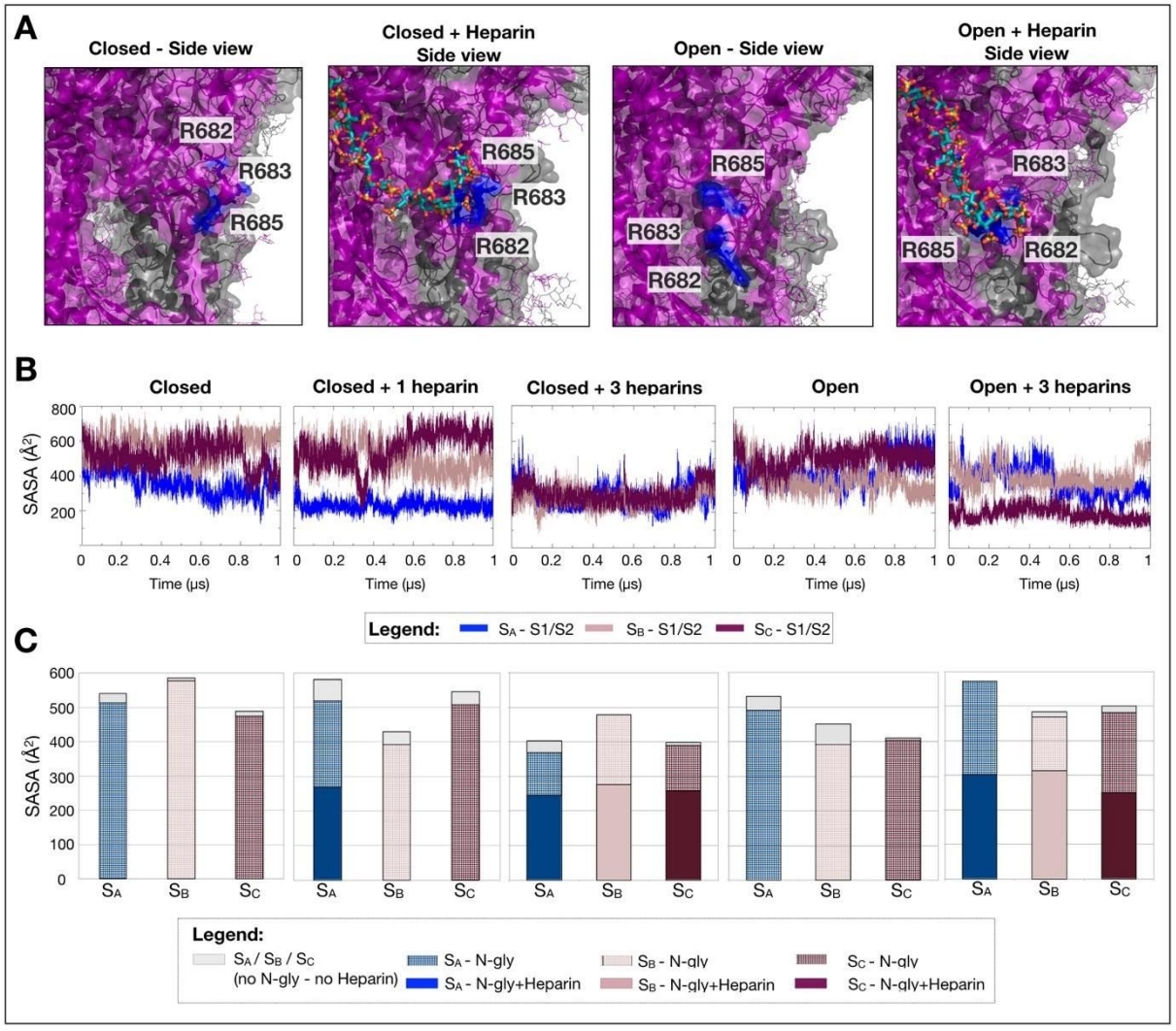



***Fig. 10.*** *Interaction of heparin with the spike S1/S2-HBD. (A) The spike S1/S2 site in closed (left) and open (right) conformations are shown with and without heparin. The $S_C$ and $S_B$ subunits are shown as cartoon and translucent surfaces in magenta and grey, respectively. Basic residues of S1/S2 are shown by blue surface. N-glycans are shown in line representation colored according to the corresponding subunit. Heparin is shown in stick representation and colored by element with cyan carbons. (B) SASA of the S1/S2 site plotted as a function of time for one trajectory for each simulated system: closed, closed with 1 heparin chain, closed with 3 heparin chains, open and open with 3 heparin chains (C) and computed for the most representative clusters in the MD simulations for different contributors to the burial of the S1/S2 site surface.*

- **Heparin allosterically affects the hinge region of the RBD and directly interact with the basic residues of the RBD-HBD**

To assess the mechanistic effect of heparin on the RBD, we analysed the stability of the hinge region associated with the activation of the RBD, the exposure of the up-RBD (RBD of the $S_C$ subunit) along the trajectory and, in particular, the potential shielding effect of heparin on the residues of the RBD involved in the interaction with ACE2, hereafter called the receptor binding motif (RBm) (84).

From a comparison between the crystal structures of the spike in closed and open conformations and the RMSF (data not shown), we identified residues 501-PKK-503 as the hinge region responsible for the conformational change which induces the opening of the spike protein. Importantly, no direct interactions occur between these residues and heparin, prompting us to investigate possible allosteric effects induced in this region by the binding of heparin to spike.

To this aim, the RMSD and dihedral principal component analysis (dPCA) of the hinge region along the trajectory were calculated for each subunit. As shown in Fig11A, the complex formed by the closed spike with three heparins bound shows structural stability of the three hinge regions along all the simulations. The closed conformation with one heparin bound shows stability for the two subunits that are directly involved in the interaction (Fig11A). However, the $S_C$ subunit that does not directly interact with heparin shows an increased RMSD of 0.5Å (Fig11A) and an RMSF lower than 0.5Å (data not shown), suggesting a compensatory effect due to the movement of the hinge region to maintain the closed conformation. In agreement, the dPCA of the hinge region of the closed-$S_C$ subunit in the spike complex with three heparins bound shows a limited sampling of the conformational space, in line with the stability shown by the RMSD. At variance, in the spike complex with one heparin bound, the dPCA shows an increased sampling for the hinge region, further supporting the hypothesis of a compensatory effect to maintain the closed conformation (Fig11B). On the other hand, the RMSD for the hinge region of the open spike with three heparins bound shows stability for the subunits $S_A$ and $S_B$ with the down-RBD and an increased deviation for the up-RBD of subunit $S_C$ compared to the starting structure (Fig11A). Accordingly, dPCA analysis shows a decrease in sampling the conformational space of the $S_C$ in presence of heparin associated with the stabilization of the hinge region (Fig.11B).

To evaluate if the induced fit promoted by heparin causes the masking of the ACE2 binding residues in the RBm, we calculated the SASA of these residues along the trajectory and their accessibility in the representative clusters. Both the analyses showed that the heparin chains do not significantly shield the residues of the RBm (Fig.11C-D). On one hand, these data suggest that heparins act indirectly on these domains with an induced fit mechanism as described in the previous paragraph. On the other hand, these data anticipate the capability of spike to simultaneously bind both

the HSPGs and ACE2, as suggested by the surface plasmon resonance (SPR) analysis of Clausen and co-workers (73).

Finally, to obtain further insights into the effect of heparin on the open RBD, we performed essential dynamics (ED) analysis (Fig.11E). The analysis on the closed conformation shows an overall stabilization of the spike without significant effects on the RBDs (data not shown). ED analysis on the open conformation highlights a different magnitude of motion with and without heparin and in particular, the motion described by the first eigenvector on RBDs in absence or presence of heparin is completely different (Fig.11E). Although the different starting conformations and the independent sampling, all the up-subunits in absence or with heparin bound shown an univocal orientation of motion, suggesting that the presence of heparin, thus also for the HSPGs, could affect the motion, induce a gating-like conformational change schematized in Fig11E.

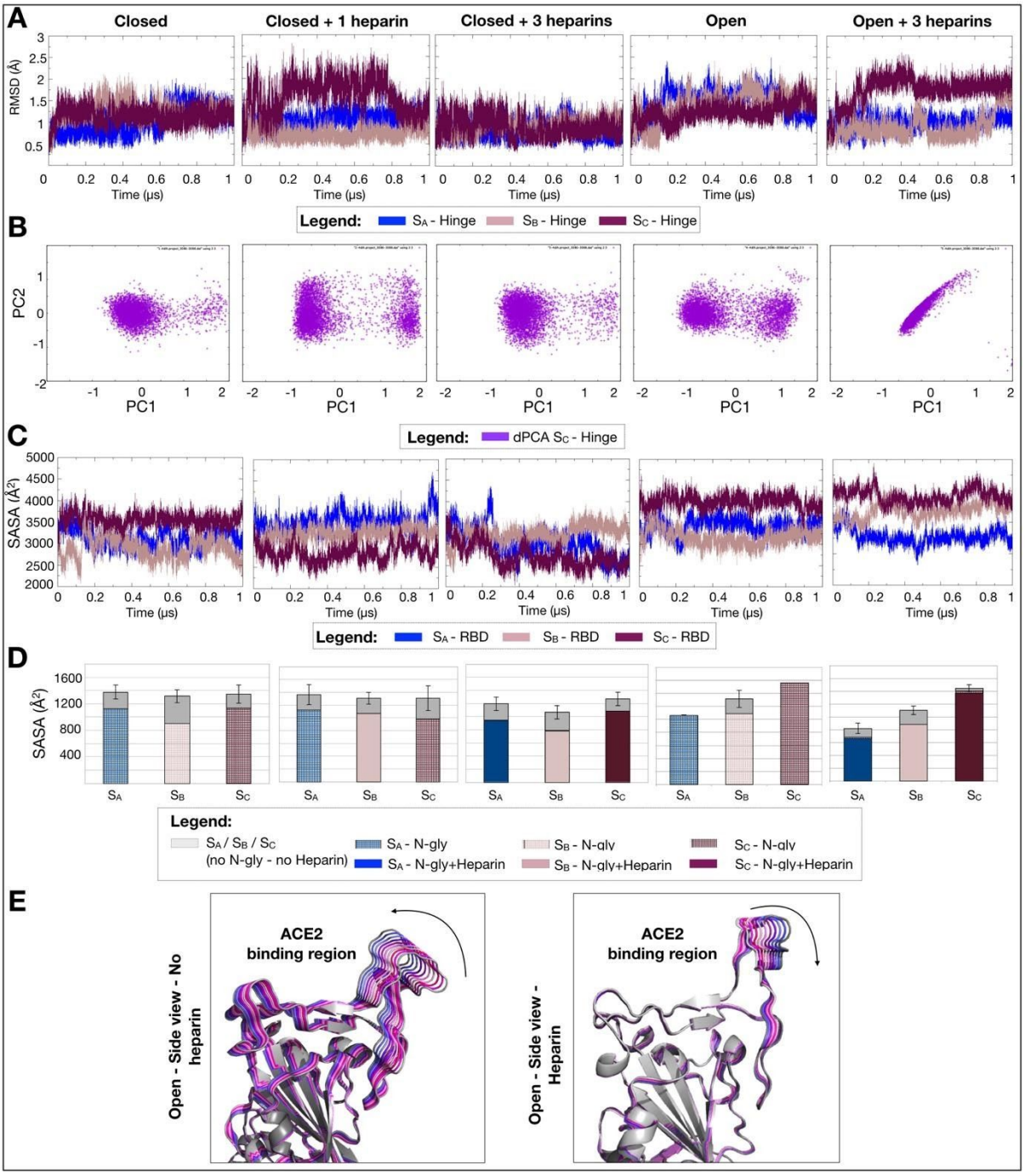



***Fig. 11***. *Conformational flexibility of the hinge region (residues 501-503) and exposition of the RBD. (A) RMSD and (B) dPCA of the hinge region vs. time along a trajectory for each simulated system. SASA of RBm along the corresponding trajectories and in the most representative cluster for each system are reported in (C) and (D), respectively. (E) Superimposition of 10 representative configurations obtained projecting the C-alpha motion onto the firsts essential eigenvector of fluctuation involved in the motion of the up-RBD without (left) and with (right) heparin showing the rigidification of up-RBD induced by heparin. The RBD is shown in cartoon representation and colored from magenta to blue with increasing time. Heparin is omitted for ease of visualization.*

- **N-glycans can modulate binding of the spike for heparin**

In this last section, we aimed to briefly describe the direct interaction between heparin and the N-glycans. Heparin chains in both closed and open models of spike were modelled to interact with K444 and N448 residues, that are in close proximity of the ACE2 binding residues. Visual inspection of the trajectories and the H-bond analysis shows that N-glycans can aspecifically and transiently cause the detachment only of a limited portion of the heparin chains from spike. In particular, we identified two N-glycans with key roles in this mechanism: the one attached to N-122 (in the N-terminal domain) acts only in the closed conformation while, that attached to N-343 (in the RBD) is mainly responsible for the detachment in both in the closed and the open conformations (Fig.12). In this context, as the SPR and circular dichroism spectroscopy were done using unfractionated heparins (13.5-15kDa - ~48 monosaccharides), we cannot exclude that longer heparin chains could maintain the interaction with these residues (86).

In conclusion, our simulations suggest that N-glycans have exerted a shielding effect resulting only in a non-specific, partial and transient detachment of heparin.

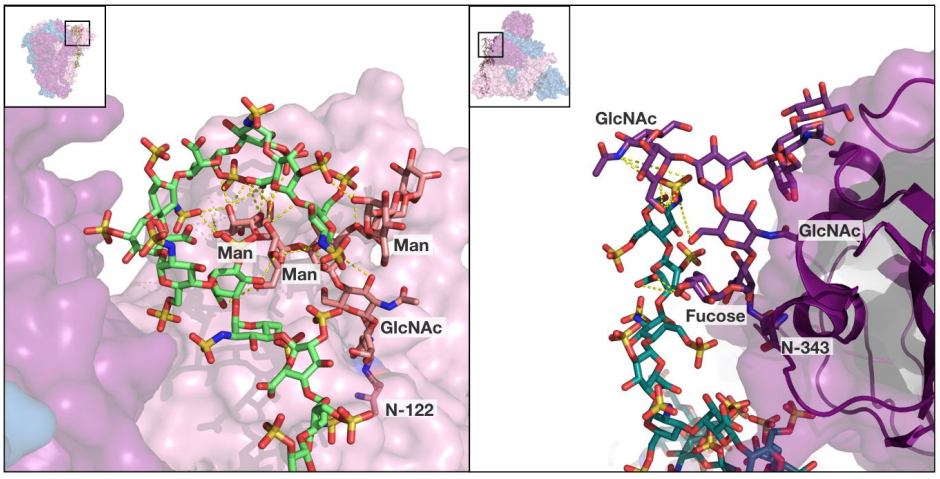



***Fig. 12.*** *Spike N-glycans can transiently displace heparin from the spike surface. Interactions between the N-122 (left) and N-343 (right) N-glycans with heparin in closed and open models, respectively. The region shown in the spike is indicated by the squares in the insets. The $S_B$ and $S_C$ subunits are shown as translucent surfaces and cartoons in pink and magenta, respectively. N-122, N-343 and the N-glycans are shown in stick representation and colored by element with pink/magenta carbons. Heparin chains are shown in stick representation with carbons in light green (left) and cyan (right). Dashed lines indicate the H-bond interactions between the glycans and heparin.*

## 2.2.4. Conclusions

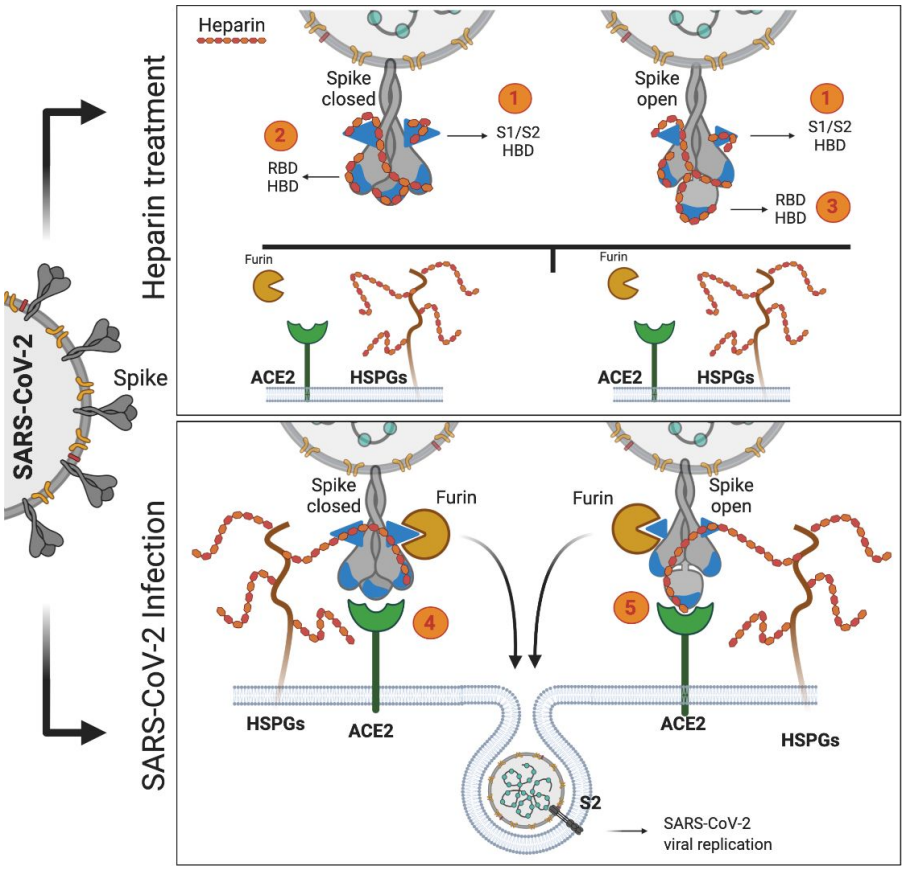
We simulated several systems composed of the head of the SARS-CoV-2 spike glycoprotein in either closed or open conformations in the presence of zero, one or three heparin chains. We analysed the dynamics and stability of these systems to gain mechanistic insights into the antiviral effect of heparin and, based on the similarity with HSPGs to anticipate their role as co-receptor in SARS-CoV-2 infection. Our models shows that the minimal length required to connect the S1/S2-HBD of one subunit to the RBD-HBD of the adjacent subunit is around 24 monosaccharides, although a 31 monosaccharides polyanionic chain is needed to fully cover both the basic domains and characterize the direct and allosteric effects exerted.

Overall, our multiple models of spike with heparin bound allow us to identify similar long, mostly basic patches between the closed and open conformations, where heparin exerts its direct antiviral effect burying these basic regions and hindering the binding to HSPGs. Multiple models of the spike in a closed conformation and in the presence of a single heparin chain, show that on a timescale of 1 μs, heparin/HSPGs are not able to directly induce the opening of the closed spike, suggesting that the role of HSPGs is likely only to increase the concentration of the virus on the host-cell surface. However, HSPGs may also represent a sort of "shortcut" by mediating the activation of the closed spike due to its exposition to the human ACE2. Similarly, models of the closed spike with three heparin chains confirm the ability of heparin to stabilize the spike in a closed conformation, hinder both the opening/activation and prevent the binding to the HSPGs due to the masking of all the basic regions exposed to the solvent. Finally, the open models in the presence of heparin demonstrate that heparin masks the RBD-HBD but is not able to induce the closure of the RBD and the masking of the RBm. Thus, HSPGs may favour the formation of a ternary complex with the up-subunit of spike and ACE2.

Interestingly, all the simulations in the presence of heparin confirm the ability of heparin to act both directly and allosterically. On one hand, heparin masks the S1/S2 multibasic site, and could thereby prevent the cleavage by furin and the activation of the prefusion conformation. On the other hand, heparin masks the basic residues of the RBD and allosterically acts on the hinge region that is suggested to be key for the opening and mobility of the RBD.

Finally, open and closed models with heparin bound suggest a modulatory effect of N-glycans on the binding between spike and heparin.

Taken together, our computational studies can provide useful insights for the rational optimization of tailored heparin derivatives for the development of new antivirals.




***Graphical representation of the conclusions.*** *Heparin exert its function as antiviral preventing the binding between spikes both in closed and open conformation to the HSPGs. In particular, heparin binds the S1/S2-HBD of spike in both the conformations (1), hinders the opening/activation of the closed spike (2) and acts allosterically on the hinge region associated with the movement of the RBD and directly masking the RBD-HBD (3). Moreover, based on the structural similarity between heparin and HSPGs we anticipate that HSPGs are able to bind both the closed (4) and open (5) conformation of spike favouring interaction with the ACE2 receptor, the subsequent furin cleavage and the SARS-CoV-2 infection.*

## 2.3 New insight into the role of heparin and HSPGs in the interaction of VEGF and its receptor VEGFR2

*This work is still ongoing and the related manuscript is in preparation.*

### 2.3.1. Introduction

Neovascularization, also termed angiogenesis, is the formation of new capillaries that takes place during cancer and represents the outcome of the interactions occurring at the endothelial surface among angiogenic growth factors and their receptors (87).

VEGFs family comprises VEGF-A, B, C, D, E and placental growth factor (PlGF), with VEGF-A representing the most important member involved in angiogenesis. VEGF-A interacts with three tyrosine kinase receptors (VEGFRs) expressed on ECs, among which VEGFR2 (KDR) represents the primary pro-angiogenic receptor, with the VEGF-A/VEGFR2 system emerging as the most studied target for the development of antiangiogenic drugs (88). VEGF exists as a dimer that, by engaging two VEGFR2 molecules, induces their dimerization (89). VEGF dimerization and interaction with VEGFR2 are regulated by free heparin and structurally related HSPGs expressed on the surface of endothelial cells.

In effect, the various VEGFs are endowed with different capacities to bind heparin and HSPGs. Through alternative mRNA splicing, the *VEGF-A* gene codifies for various isoforms that differ by the presence or absence of a C-ter HBD into its C-ter 55 residues: VEGF-$A_{121}$ lacks the HBD and does not bind HSPGs, being found mainly as a free protein in body fluids. VEGF-$A_{189}$ is instead found mainly tethered to the HSPGs of the extracellular matrix in an inactive form whose enzymatic activation generates VEGF-$A_{110}$ that lacks the HBD (90). VEGF-$A_{165}$ (from here on referred to as VEGF) interacts with heparin (91) and with HSPGs that act as coreceptors for its subsequent interaction with VEGFR2 (92). Although all the sulfate groups of heparin contribute to the interaction with VEGF, *6-O*-sulfate groups appear to be particularly important. An hexa/eptasaccharide is sufficient to bind a VEGF monomer (93) but heparin length represents an important parameter in determining the biological consequences of the interaction of VEGF with heparin/HS: low and high molecular weight heparins inhibit and potentiate VEGF binding to its receptors, respectively (60). Accordingly, *in vivo*, short heparins suppress VEGF-mediated angiogenesis, while longer heparins promote the process (94). It is thus tentative to hypothesize that, depending on its length, heparin differently binds VEGF alone (mainly in case of an inhibitory effect, classically due to a sequestration of the protein in the extracellular matrix) or VEGF complexed to VEGFR2 (promoting receptor activation and hence angiogenesis).

The accurate characterization of the binding mode(s) of macromolecular complexes by means of computational studies (molecular docking and dynamics) is by now almost mandatory for the

characterization of biological macromolecular complexes. In the field of neovascularization, this approach has been extensively used for the study the interaction of the angiogenic fibroblast growth factors (FGFs) with heparin and with their tyrosine kinase receptor (FGFRs) (95). At variance, this approach for the VEGF/heparin complex has been so far delayed since only the N-ter and C-ter domains of VEGF, containing its RBD and HBD are available separately (PBDid: 1KAT and 1KMX, respectively). The dodecapeptide that comprises residue 133-145, acting as a link between the RBD and HBD domains has not yet been resolved. As a consequence, many molecular docking and dynamics have been performed on the two separate VEGF domains, with some studies in which different strategies have been applied to model a 3D model of VEGF containing both RBD and HBD (60).

It must be pointed out that, whatever the binding protein, computational docking and molecular dynamics with heparin have been so far limited only to short heparins (from 2-mer to 14-mer). This limitation is mainly due to heparin conformational flexibility, large number of torsional angles between glycosidic bonds, high charge density and large electrostatic surface involved, weak surface complementarity of heparin/protein interactions, absence of well-defined binding pockets and difficulty to define the impact of solvation/desolvation on the interaction (96). This type of "limited" approach has been instrumental for the identification of VEGF HBD, to study the role of heparin in VEGF dimerization and to help the design of VEGF binders endowed with anti-angiogenic potential. However, it could not be applied to understand the composite interaction occurring between (longer) natural heparin/HS and the complex formed *in vivo* by VEGF and VEGFR2 dimers. Relevant to this point, it has been proposed that heparin may interact directly with VEGFR2 (97), as already demonstrated in other pro-angiogenic tyrosine kinase receptors such as FGFRs (98).

To overcome these limits, we have recently proposed a new protocol that allows the modelling of longer heparins (up to 24-mer), with which it has been possible to study the role of the GAG in the process of HIV-1 p17 protein oligomerization (37). Using this method, we are aimed to clarify the role of heparin in the oligomerization process of full-length VEGF and the role of the strong anionic sugar in the exposition of the growth factor to its receptor. Equally important, being an appropriate full-length model of VEGF unavailable, we applied different molecular dynamics methods to predict the full-length structure of the growth factor.

## 2.3.2. Material and methods

*REMD* was carried out using Amber20 software in implicit solvent (79). The linker decapeptide that connects the VEGF RBD to its HBD was at first divided into two partially overlapping peptides (RPKKDRARQ and KDRARQENP). The number of replicas was determined on the basis of the number of atoms in the system. For all the peptides 8 replicas were used and the temperature were started using the temperature generator for REMD simulation (http://folding.bmc.uu.se/remd/) with 300K as temperature of the lowest replica and 0.25 as expected probability rate for each replica. To prevent unwanted rotations around the peptide bond due to high temperature used during the simulation, chirality restraints were generated using the "makeCHIR_RST" script provided in the Amber20Tools (79) with default parameters. The simulations were carried out using Langevin temperature coupling with collision of frequency parameter $\gamma$ = 1ps (gamma_ln=1.0). The SHAKE algorithm was used to constrain all bonds containing hydrogen atoms and a 1 fs time integration step was used. For each temperature, 200ps of equilibration and 50ns of effective MD run were carried out and trajectory were recorded every 10ps. There were 10 attempts in the simulation, which corresponds to one attempt per 1ps. The trajectory corresponding to 300.0 K was extracted and analysed using Cpptraj, evaluating both most representative clusters based on C- $\alpha$ distance and dPCA obtained by the analysis of phi and psi angles of the peptide. The simulation was performed in a cluster of 8 Tesla K20i GPU.

*Molecular modelling of complete VEGF and D2D3-VEGF.* Modeller with default parameters and advanced modelling protocol was used to combine the structures of dimeric VEGF RBD (PDBid 1KMX_V), decapeptide RPKKDRARQENP (corresponding to the lowest energy from the REMD simulation) and VEGF HBD (aa 132-143) (PDBid 1KAT, aa 1-55). The obtained structures were compared by superimposition of backbone with the structure of VEGF/VEGFR2-D2D3 (PDBid 3V2A) and the ones with spatial overlap were discarded. The structure of complete dimeric VEGF/VEGFR2-D2D3 was obtained by overlapping starting from the 3v2a structure.

*aMD* was carried out using Amber 20 software (79) in explicit solvent. During the preparation of the input file, all the disulfide bonds in the system were imposed to prevent unwanted modification in the structure induced by the boosts during the simulation. Energy minimization was carried out with the non-bounded cut-off of 8 Å through the following steps: (i) whole VEGF and counter ions were restrained by decreasing harmonic potential from 50 to 10 kcal/mol × Å2, while water molecules were relaxed using 5000 cycles of steepest descendent method; (ii) counter ions and then hydrogen were relaxed using 5000 cycles of steepest descendent method whereas the protein is already restrained by harmonic potential of 10 kcal/mol × Å2; (iii) the full-lenght VEGF was restrained by decreasing harmonic potential from 8 to 2 kcal/mol × Å2 by 5000 cycles of steepest descendent

method; (iv) the system was relaxed using two steps composed by 5000 and 50000 steps of steepest descendent method without any restrain. The system was heated from 0.1 K to 100.0 K in NVT and from 100.0 to 300.0 K in NPT. Equilibration (3 ns) and simulation were validated using the physical observables parameters of the system confirming that the complexes obeyed the NPT ensemble. The complex was simulated in periodic boundary conditions using the Langevin algorithm at 300.0 K and the SHAKE algorithm was used to constrain all bonds containing hydrogen atoms. aMD simulations were performed over 50ns using pmemd CUDA and a server with Tesla K20i GPU.

*Heparin modelling***.** Heparin was designed as described by Bugatti and co-workers (37). ClusPro web-server (16) was adopted for the blind docking simulation , Autodock4.2 (17) and Pymol for the refinement of the models as described in section 2.1.

*MD simulations.* Amber20 package (79) was used for MD simulations of VEGF dimer in complex with heparins that were carried out using ff14SB force field (80) for protein and GLYCAM_06 for heparins. Each complex was neutralized by adding $Ca^{2+}$ restrained away from the protein and solvated with TIP3P water model. Energy minimization was carried out with the non-bonded cut-off of 8 Å through the following steps: i) VEGF/heparin complexes and counter ions were restrained by a harmonic potential of 5 kcal/mol×$Å^2$, while water molecules were relaxed using 2,500 cycles of steepest descent and conjugate gradient methods; ii) counter ions and hydrogens were relaxed using 5,000 cycles of steepest descent and conjugate gradient methods and restrained by a harmonic potentials of 3 kcal/mol´$Å^2$ and then of 1 kcal/mol´$Å^2$; iii) the system was relaxed using 5,000 cycles of steepest descent and conjugate gradient methods without any restraint; iv) the system was heated from 0.1 K to 100.0 K in NVT (constant volume) and from 100.0 to 300.0 K in NPT (1.0 atm constant pressure). The two generated complexes were simulated in periodic boundary conditions using the Langevin algorithm at 300.0 K. During heating and simulations, the $Ca^{2+}$ ions were restrained at 500 kcal/mol×$Å^2$. Equilibration (3 ns) and simulation were validated using the physical observables parameters of the system confirming that the complexes obeyed the NPT ensemble. Electrostatic interactions were calculated using the Particle Mesh Ewald method. A cut-off of 8.0 Å was applied to van der Waals forces. Integration time step was set to 1.0 fs during equilibration and 2.0 fs during simulation. MD simulations were performed over 50 ns using the pmemd CUDA program of the Amber14 package and a server Tesla K20 GPU.

## 2.3.3. Results

- **Characterization of the full-length VEGF.**

Due to the absence of a full-length model of VEGF, we applied MD simulations to predict the whole growth factor. Firsty, replica exchange MD simulations were performed to characterize the behaviour of the dodecapeptide RPKKDRARQENP itself. The simulations were performed by splitting the dodecapeptide in 2 smaller nona-peptides (RPKKDRARQ and KDRARQENP) to set up the protocol and analysed by dPCA integrated with the Boltzman constant and to obtain the free energy landscape (data not shown). The same protocol was then used to characterize the ternary structure of the dodecapeptide. The result of the free energy landscape on REMD simulation of the dodecapeptide, associated with the 3D structure collected within 2kcal/mol from the global minima suggests that the peptide tends to fold as an α-helix in its first half thanks to H-bond interactions between the backbone of residues $P_{133}$, $K_{135}$, $D_{136}$, $R_{137}$ and $A_{138}$. The last portion of the peptide tends instead to remain unfolded possibly due to the presence of $P_{143}$ which acts as a disorder promoting agent, differently from the $P_{133}$, which disorder activity is neutralized by the H-bond network to fold as an α -helix (Fig. 13). The representative structure corresponding to energy minima was used to model the full length VEGF. Dimeric VEGF was modelled using Modeller and based on the crystal structure of the dimeric RBD (PDBid 1KMX_V) and the HBD (PDBid 1KAT) (Fig. 14).

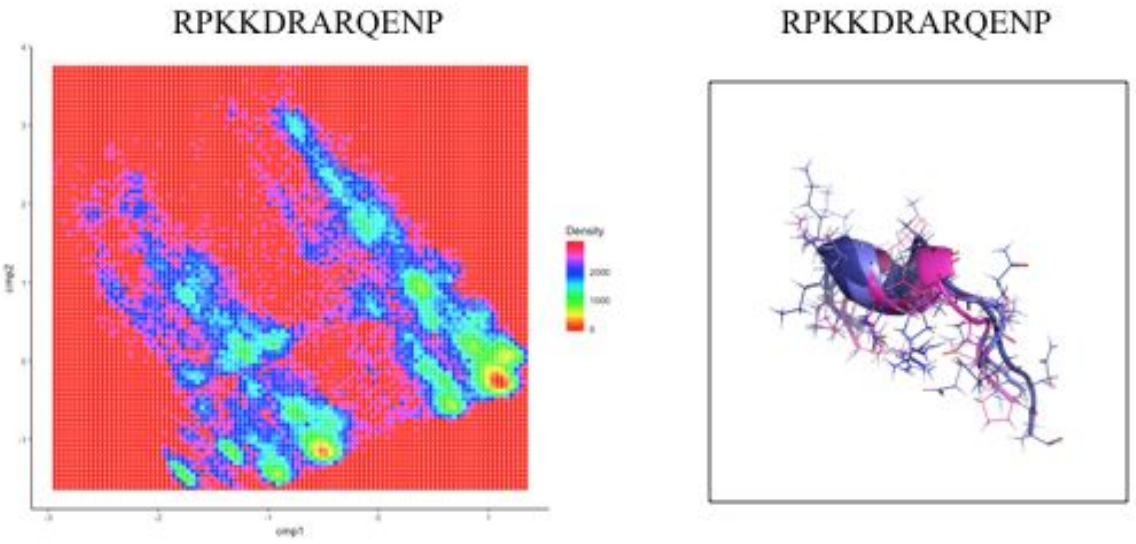



***Fig. 13.*** ***(Left)*** *Free energy landscape analysis on PCA1 (cmp1) and PCA2 (cmp2) from the REMD simulation of the linker domain (RPKKDRARQENP) between RBD and HBD of VEGF.* ***(right)*** *Representative structure of the peptide corresponding to the energy minima (within 2kcal/mol from the global minima) superimposed on the backbone. The linker is shown as cartoons and lines.*

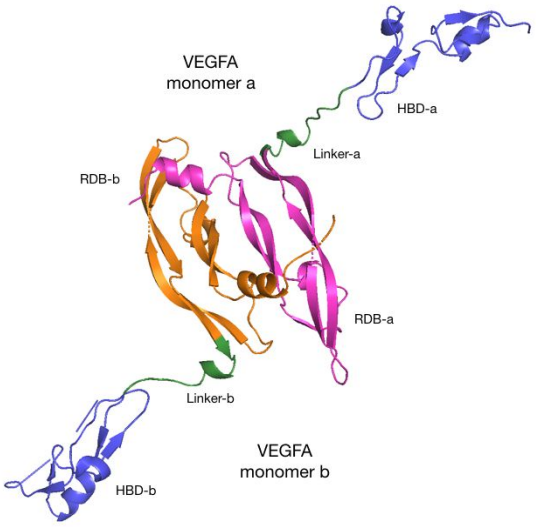



***Fig.14***. *Model of dimeric full-length VEGF obtained by Modeller. VEGF is shown in cartoons and colored based on the domains. RBD of monomer a and b are colored in magenta and orange, respectively. Both the linker and the HBD are shown in green and slate, respectively.*

The resulting models were filtered by superimposition of the C-α with the crystal structure of VEGF/VEGFR2 (PDBid 3V2A) to exclude the models that classes with the Ig-like domains of VEGFR2. Finally, to cross check the results obtained by REMD on the linker and by Modeller, we decided to characterize the behaviour of the VEGF using aMD, which allows us to overcome the local minima of the 3D structures and sample the conformational space by applying boost during the simulation. aMD simulation was carried out without any positional restraint to evaluate the behaviour of the linker connected to the RBD and HBD. The trajectory was analysed through dPCA integrated with the Boltzman constant (Fig. 15). The free energy landscapes on the PC1 and PC2 showed a good quality of the sampling space and suggest the presence of a principal local minima that was extracted from the simulation and used as full-length VEGF.

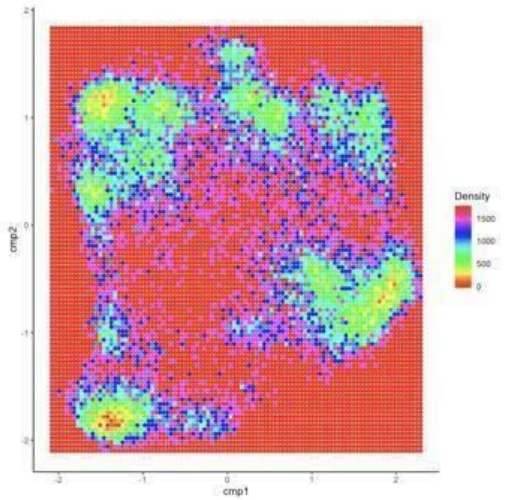



***Fig.15*** *Free energy landscape analysis on PC1 (cmp1) and PC2 (cmp2) of the aMD simulation on the full-length model of VEGF obtained by the modelling with Modeller.*

- **Heparin stabilizes VEGF inducing a conformational drift in the growth factor.**

HSPGs act as reservoirs, increasing the concentration of a ligand at the cell surface presenting it to the signalling receptor. In cell assays, heparins longer than 22 mer effectively promote VEGF/VEGFR2 interaction (90). To elucidate if heparins longer than 22 mer are able to induce mechanistic effects on dimeric full length VEGF, we decided to model the complex between VEGF and a heparin long enough to span from the HBD of one monomer to the HBD of the second monomer. The model of this heparin was set up following the protocol already reported (37). As a result, a 36-mer heparin was modelled, complexed with the VEGF dimer and subjected to 50ns MD simulations to evaluate the stability of the complex and possible mechanistic effects induced on the proteins. Fig. 16 shows the most representative cluster extracted from the first MD simulation that suggests that in effect heparin could induce a change in the conformation of both the linker dodecapeptide and HBD of VEGF. Further studies are ongoing to confirm and better characterize these preliminary results.

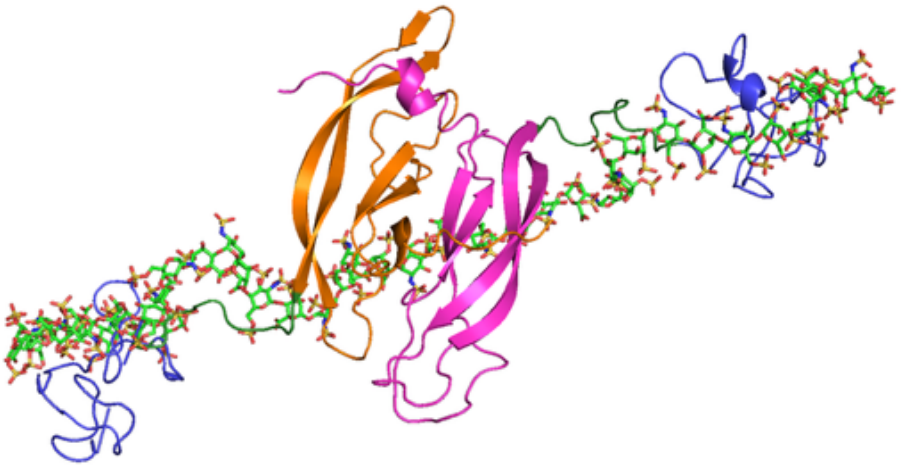


***Fig.16**. Representative cluster of full-length VEGF in presence of 36-mer heparin chain extracted from the 50ns of MD simulation. Full length VEGF is shown in cartoons and colored based on the domains. RBD of monomer 1 and 2 are colored in magenta and orange, respectively. Both the linkers and the HBD of monomer 1 and 2 are colored in green and slate, respectively. Heparin is shown in sticks and colored by elements in green (oxygens are in red, sulfates are in yellow and nitrogens are in blue).*

- **Heparin promotes the interaction of VEGF with VEGFR2.**

As mentioned above, heparins longer than 22 mer promote the VEGF/VEGFR2 interaction in cell assays (90). We thus decided to clarify the role of heparin in the promotion of VEGF/VEGFR2 interaction. Then, we coupled the full-length VEGF model described above to the model of the domains D2 and D3 of VEGFR2 by superimposition of the domains starting from the PDBid 3V2A. The model of the VEGF/VEGFR2 complex was then coupled to the heparin model described above.

The electrostatic surface potential analysis suggests that both the linker and the closest part of D2 act as basic domains which favour the formation of the trimeric complex. We then decided to design another heparin structure that spas from the linker of one monomer to the linker of the second monomer, thus interacting with both the basic residues of the linker and D2 of VEGFR2, allowing to evaluate both the importance of the basic domains in the formation of trimeric complex and the capacity of heparin to wrap around the VEGF HBD. In Fig. 17 is shown the complex set up that will be subjected to MD simulations for further interpretations.

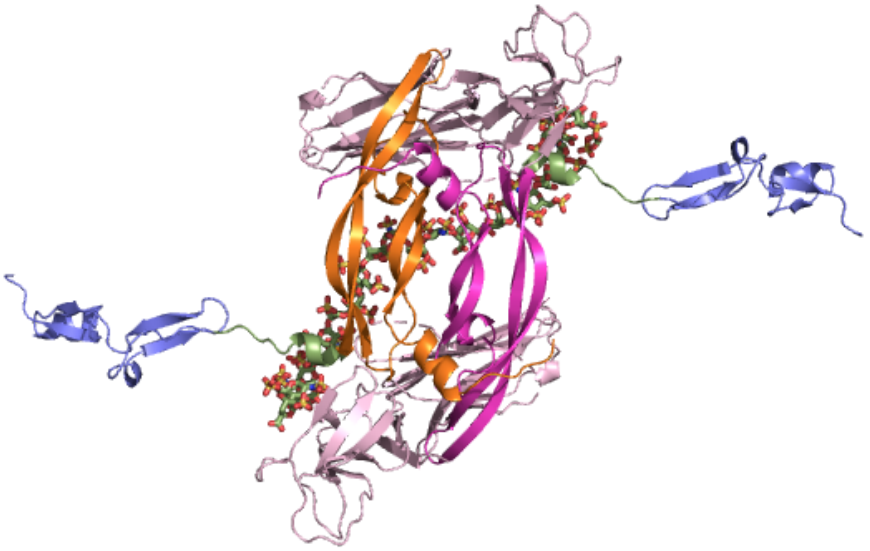


***Fig.17.*** *Trimeric complex between full-length VEGF, VEGFR2 D2D3 and 22-mer heparin. Full length VEGF is shown in cartoons and colored based on the domains. RBD of monomer 1 and 2 are colored in magenta and orange, respectively. Both the linkers and the HBD of monomer 1 and 2 are colored in green and slate, respectively. D2D3 VEGFR2 are shown in cartoons and colored in light pink. 22-mer heparin is shown in sticks and colored by elements in green (oxygens are in red, sulfates are in yellow and nitrogens are in blue).*

## 2.3.4. Conclusions

Despite its preliminary nature, this work already provides us with novel models of dimeric VEGF in its apo form or complexed to heparin or VEGFR2 or to both simultaneously. In particular, the trimeric complex emerges as promising for further computational analysis aimed at the comprehension at an atomic level of the formation of such a complex that, in turn, will provide invaluable insights useful to comprehend the wide range of biological phenomena leading to VEGF-induced neovascularization.

# Chapter 3.

# Sialic Acid as a New Target for Drug Discovery

*The introduction of this chapter is extracted from the paper: Chiodelli P, Urbinati C, **Paiardi G**, Monti E, Rusnati M. Sialic acid as target for development of novel antiangiogenic strategies. Future Medicinal Chemistry. (2018) 10:2835-54.*

Neuraminic acids, also termed sialic acids (Sia) belong to a large family of related acidic monosaccharides that share a core structure of a nine-carbon polyhydroxylated α-keto acid derived from neuraminic acid. The most abundant member of the family carries an acetyl moiety linked to the amino group of carbon 5 giving the N-acetylneuraminic acid (Neu5Ac). O-acetylation of the glycerol portion of Neu5Ac, as well as variation to the carbon 5 substituents (Fig. 18) gives origin to several other Sia derivatives. Sia can be found in a free form (polysialic acid), associated with glycosphingolipids (gangliosides), or as terminal components of the oligosaccharide chain of glycoproteins. In these structures, Sia is bound through α(2-3) and α(2-6) glycosidic linkages to galactose residues or through α(2-8) linkage to other Sia residues (99).

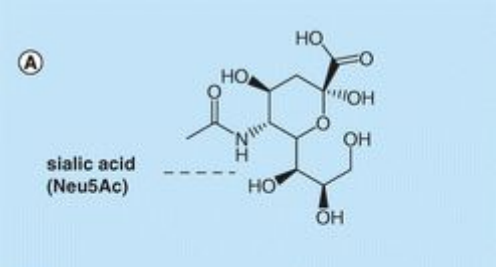



***Fig18**. Structure of Sia. Chemical structure of Sia. C2 is involved in glycosidic linkages (see text for details) and C5 bears the N-acetyl moiety (dotted line).*

Several enzymes are involved in Sia metabolism. Sialyltransferases (STs) are membrane proteins localized within the trans-Golgi networks along the secretory pathway that utilize cytidine monophosphate-Sia as activated sugar donor and galactose or Sia as acceptors (100). STs are divided in four subfamilies (ST3Gal, ST6Gal, ST6GalNAc and ST8Sia) depending on the glycosidic linkage formed and the monosaccharide acceptor recognized (101). STs expression is finely regulated in different organs during physiological or pathological conditions (102). ST-mediated addition of Sia on glycans usually stops their further growth and modifies charge, steric hindrance, conformation and flexibility, underlying the importance of STs in shaping the structures and functions of sialoglycans (103). On the other hand, the catabolism of sialo-glycans involves sialidases (NEUs), members of the CAZy family GH33 (104) that specifically release Sia through hydrolysis of the glycosidic bond between the acidic sugar(s) and the internal acceptor.

In conclusion, the whole set of sialoglycans of eukaryotic cells is the outcome of the fine tuning of the biosynthetic and trimming actions exerted by STs and NEUs, whose balanced activities are altered during cancer (104). Indeed, STs over-expression and consequent aberrant sialylation of

glycoconjugates modulate tumor cell invasiveness and dissemination, and correlate with poor prognosis (105). Thus, STs, NEUs and Sia itself represent important therapeutic targets for the development of novel anticancer therapies (106).

As already mentioned, computational modelling can greatly contribute to drug discovery. In the field of sialoglycans, computational studies have been mainly devoted to the interaction of STs with Sia-donors and acceptors (100). These studies predicted the donor-binding site to be highly conserved among STs, with the major variations occurring in the acceptor site. This makes it possible to exploit computational tools to assist structure-based design of inhibitors either common or specific for the various STs involved (105). Also, the complete tridimensional models (including the glycan portion and Sia) of some sialylated proteins have been already resolved and used in computational studies (Tab. 03).

Thus, pinpointing the activity of STs and their interaction with bi-antennary N-glycans at a molecular level is almost mandatory to develop new specific inhibitors on this reaction. For this reason, the aim of this part of my work has been devoted to the modelling and characterization of VEGFR2-ST6Gal complex exploitable for the identification of druggable pockets and related inhibitors.

| | **compound** | **complex with** | **PDB code** |
|---|---|---|---|
| **sialylated proteins** | Plasminogen | α2-antiplasmin | 4A5T |
| | Fibrinogen | n.c. | 3GHG |
| | PSA | anti-PSA antibody | 3QUM |
| | IgG1-Fc | FcγRI | 5GSQ |
| | IgA1-Fc | FcγRI | 1OW0 |
| | CSF receptor-1 | colony stimulating factor-1 | 3WSR |
| | Podoplanin | C-type lectin-like receptor 2 | 3WSR |
| | SLx-containing glycoproteins | E- and P-selectin | 4CSY |
| | serum amyloid P component | hexanoyl bis(D-proline) | 4AVV |
| | Transferrin | Neisseria TbpA-TbpB proteins | 3V8X |
| | Lithostathine | n.c. | 1QDD |
| | Zn- α 2-glycoprotein | n.c. | 1ZAG |
| | HIV-1 gp120 | anti-gp120 antibody | 5T3X |
| | HSV-1 glycoprotein B | Siglecs | 5XOF/5XO2 |
| **Sialylated oligo-saccharides** | GT1b sialylated glycan | siglec-7 | 2HRL |
| | GM1 sialylated glycan | complement factor H<br>galectin-3 | 4ZH1<br>3AYC |
| | ganglioside analogue DSLc4 | siglec-7 | 2DF3 |

| | | | |
|---|---|---|---|
| | Sialyllactoses | siglec-5<br>siglec-2 (CD22)<br>galectin-3<br>galectin-8<br>adenovirus 37 fiber head | 2ZG1<br>5VKM<br>4LBO/4LBL<br>3VKM/3VKO<br>2W9L |
| **sialic acid** | 6'sulfo sialyl Lewis(x). | siglec-8 | 2N7B |
| | sialic acid | galectin-8<br>adenovirus 37 fiber head<br>carboxylesterase 1 | 3AP7<br>2WBW<br>2HRQ |

***Tab.03.*** *Sialylated structures for which a tridimensional model has been resolved. n.c.: not complexed.*

# 3.1. Modelling of the VEGFR2-ST6Gal complex

*This work is still ongoing and the related manuscript is in preparation.*

## 3.1.1. Introduction

As already described in Section 2.3, angiogenesis is the formation of new blood vessels from pre-existing ones. Physiologically, it is tightly controlled during embryonic development and wound repair. Its pathological unbalance, characterized by the uncontrolled release of angiogenesis growth factors or impediment of natural angiogenic inhibitors, contributes to the development of many diseases, including cancer. Sia is the major surface anion on the endothelial cell membrane where it regulates various molecular and cellular interactions during angiogenesis and tumorigenesis (104,107).

The VEGF/VEGFRs system plays a key role in pathological angiogenesis (see Section 2.3 for further details). Interestingly, the extracellular portion of VEGFR2 bears N-glycans with terminal α(2-6) Sia. Although mainly localized on vascular endothelial cells, sialylated VEGFRs are also found on a variety of tumor cells, indicating that, beside angiogenesis, they can directly influence tumor cell growth and invasion *via* an autocrine mechanism (108). More recently, it has been demonstrated that VEGFR2-associated α(2,6) Sia is mandatory for VEGF binding to VEGFR2 and consequent VEGF-dependent angiogenesis (109). Taken together, these observations point to VEGFR2 sialylation as a promising target for the development of new antiangiogenic, anticancer therapies.

Based on these considerations, the aim of this part of my work has been the optimization of the available 3D structures to set up and validate a complete model of glycosylated-VEGFR2 in complex with ST6Gal to be used for the identification of druggable pockets and the subsequent screening of library-compounds with the aim of identifying inhibitors of VEGFR2 sialylation.

## 3.1.2. Material and methods

*Molecular modelling of ST6Gal1-glycan complex and D23-VEGFR2.* The structure of ST6Gal1 was refined starting from the PDBid 4JS2 through homology modeling using Swiss Model web server (https://swissmodel.expasy.org/). The CMP-Sia was positioned in the catalytic site of the enzyme through superimposition using Pymol starting from the crystal structure of the *Campylobacter jejuni CstII* ST3Gal2 (PDBid 2P2V). The bi-antennary complex-type N-glycan was at first modelled using the GLYCAM web server (http://glycam.org/) and then positioned *via* semi rigid docking using Autodock 4.2 (17). The docking poses were collected keeping only the glycosidic linkages of the branch 1-3 free to move. Molecular docking was carried out with the default parameters and the pose was selected based on the H-bond analysis and the orientation of the glycan against the CMP-Sia. The structure of the D2D3-VEGFR2 was refined starting from the PDBid 2X1X using the SwissModel

web server. The bi-antennary complex type N-glycan was covalently attached to the $N_{145}$ using GLYCAM web-server. 7 different systems were set up based on the orientation of $N_{145}$ and the steric bulk of the residue and the N-glycan covalently attached.

*MD simulation.* All-atom MD simulations were independently carried out using the Amber20 package (79). Both the systems of the ST6Gal1 and D2D3-VEGFR2 were set up in explicit solvent using *leap*. ff14SB force field (80) was used to parametrize both ST6Gal1 and D2D3-VEGFR2. N-glycan was parameterized using the GLYCAM_06 force field (26) and the GAFF in case of the CMP-Sia. The simulated systems were first energy relaxed with 2000 steps of steepest descent energy minimization followed by further 2000 steps of conjugate gradient energy minimization to remove possible bad contacts from the initial structures. For each simulation, the system was heated until 300K in NVT ensemble and equilibrated by 5ns of MD runs without position restraints in NPT ensemble by weak coupling to a bath of constant pressure. 100 and 50 ns of MD simulations were carried out for the ST6Gal1 and D2D3-VEGFR2, respectively. The electrostatic term was described by using the particle mesh Ewald algorithm. The SHAKE algorithm was used to constrain all bond lengths. The analysis of the trajectories was carried out using Amber20Tools (79). The structural cluster analysis was carried out using the hierarchical algorithm.

### 3.1.3. Results

- **Characterization of the Michaelis complex ST6Gal1-N-glycan.**

The comparison between the crystal structure of ST6Gal1 with the bi-functional ST3/8 from Campylobacter jejuni (110) and porcine ST3Gal-I (111) reveals that the protein conformations are quite similar, in turn suggesting that the nucleotide-binding pocket is pre-formed. Thus, we positioned the natural substrate of reaction CMP-Sia through superimposition of the *C. jejuni CstII* ST3Gal2 (PDBid 2P2V). The system was validated carrying out 50ns of all-atom MD simulation which allow to identify the important residues involved in the stabilization of CMP-Sia. The H-bond analysis suggests that the exocyclic amino group of CMP is stabilized by an H-bond network with $C_{353}$, $T_{365}$ and $K_{376}$. Whereas, the phosphate group is stabilized by an extended hydrogen-bonding network involving $N_{233}$, $S_{323}$ and $H_{370}$, identified as the catalytic triad which coordinates the reaction (data not shown). Subsequently, to characterize the Michaelis complex, which consists in the initial step of an enzymatic reaction, an N-glycan was added to the system. The glycan was designed using the GLYCAM web server as a bi-antennary, complex-type glycan following the mass spectroscopy results released by Chandler and co-workers (112). Next, a semi rigid docking using Autodock 4.2 was performed to fit the glycan in the active site of ST6Gal1 taking in consideration the branch 1-3, directly involved in the reaction (113). The docking pose was selected evaluating the H-bonds and the atomic distances between the glycan, CMP-Sia and the catalytic triad identified. The features and the

stability of ST6Gal1-CMP-Sia-glycan was evaluated carrying out 100ns of all-atom MDs. The analysis show that the protein and CMP-Sia were stable and the H-bonds between ST6Gal1 and the branch 1-3 of the glycan near to the active site were maintained as well as the distances between the keys atoms involved in the VEGFR2 sialylation (Fig. 19). Interestingly, during all the MD simulation the bi-antennary glycan adopts a "bird" conformation with both the branches of the bi-antennary structure exposed to the solvent (103).

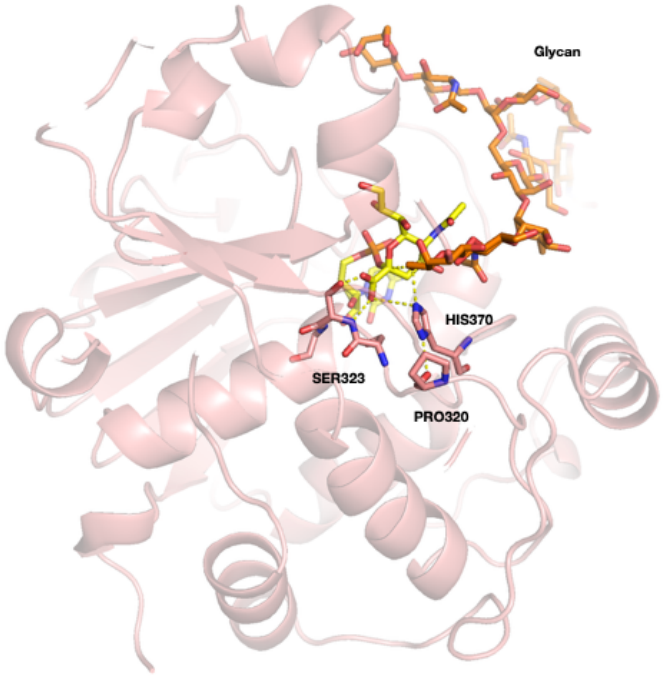



***Fig.19.*** *Michaelis complex between St6Gal1 and the biantennary glycan. St6Gal1 is shown as pink cartoon. CMP-NeuAc, the biantennary complex type glycan and the catalytic triad are shown in sticks and colored by elements (nitrogen in blue and oxygen in red) in yellow, orange and pink, respectively. The H-bond networking that characterize the nucleophilic attack are shown in yellow dots. In this model, the His-370 is pre-oriented by Pro-320 that acts as the catalytic base for deprotonation. The subsequent histidinium is stabilized by the NeuAc. The orientation between the galactose and the CMP allows the nucleophilic attack.*

- **Characterization of dynamics of glycosylated D2D3-VEGFR2.**

It is known that N-glycans adopt different conformations if simulated in solution or covalently attached to proteins, as in the case of VEGFR2 (103). Thus, we decided to model the VEGFR2 D2D3 region to characterize the trend of the N145 which could have a relevant role in favouring the interaction between VEGFR2 and VEGF. This supposition is based on the position of $N_{145}$ within the domain D2 of VEGFR2 that is directly involved in the interaction with VEGF and further sustained by mass spectroscopy data that suggest that this glycan is the mainly sialylated one among all the sugars associated to the receptor (112).

To refine the VEGFR2 D2D3 structure, we used the comparative modelling method *via* SwissModel web server to fill the gaps in the structure and investigate the flexibility of the hinge region. Then, GLYCAN web server was adopted to design the bi-antennary N-glycan covalently attached to $N_{145}$. To characterize the dynamic behaviour of the N-glycan, we set up 7 different rotamers for $N_{145}$ and 7 different starting orientations of the sugar (Fig. 20A). 50ns of independent all-atom MD simulations were carried out for each system. Interestingly, the visual inspection of the trajectories clearly confirms the capability of the N-glycan to create the classical cloud and to partially shield the D2-VEGFR2 but not the hinge region between D2 and D3. Next, the trajectories of the MD simulations were clustered based on the conformation of the N-glycans using the hierarchical algorithm to extract the most representative structure and analyzed the orientation of the N-glycan (Fig. 20B).

Interestingly, the bi-antennary sugar covalently attached to the D2D3 region primarily adopts a "broken-wing" conformation in which the branch 1-6 wrap ad is stabilized by H-bond interactions with the fucose attached to the root of the N-glycan. This conformation allows a better exposition to the solvent of the branch 1-3 that is supposed to be primarily sialylated.

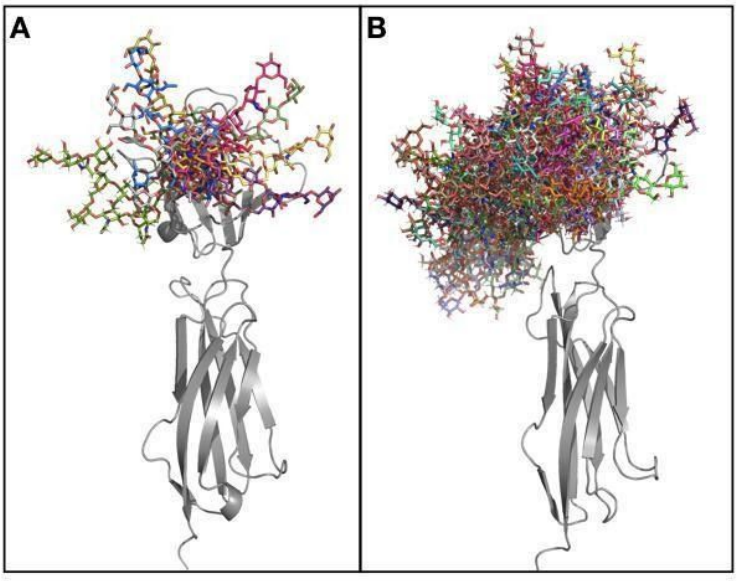



***Fig.20**. N-glycan cloud. D2D3-VEGFR2 is shown in grey cartoons. **(A)** Superimposition of the starting structure of D2D3-VEGFR2 with 7 different orientations of the N-glycans associated with 7 different rotamers of Asn-145. The N-glycans are shown in sticks and colored by elements in different colours for each glycan. **(B)** Superimposition of structures extracted along the MD simulations of the systems previously described. Independent from the starting orientation of the N-145, N-glycans are able to shield the D2 but not the hinge region between D2 and D3.*

### 3.1.4. Conclusions

Our preliminary results about the simulation of the Michaelis complex between ST6Gal1-CMP-Sia and bi-antennary glycan complex are in line with the data released by Kuhn B et al. (113) prompting us to consider the H-bond networks identified as a starting structural framework for the reposition or development of new inhibitors that could prevent this reaction. Moreover, the simulations on the D2D3 region of VEGFR2 provide novel information on the shielding effect of the N-glycan N145 on VEGFR2 and could help in the identification of druggable pockets once merged with the enzyme ST6Gal1 system. Taken together, our data and those available in literature prompt us to proceed to: 1) further validate the Michaelis complex among ST6Gal1, CMP-Sia and glycosylated D2D3-VEGFR2; 2) exploit it for the identification of druggable pockets and compounds able to interfere with the formation of the ternary complex and hence with the enzymatic reaction; 3) evaluate the possible hits for their effective capacity to inhibit VEGFR2 sialylation and angiogenic activity.

# Concluding Remarks

Many biological phenomena are regulated by sugars, especially HSPGs and Sia. Their involvement in many physiological processes and in the pathogenesis of different diseases have been widely demonstrated by experimental cellular and animal models. However, the deep understanding of their mechanism(s) of action (as well as those of heparin, a sugar widely used in therapy) has been so far limited lacking a full exploitation of computational approaches that has instead demonstrated their full potential when applied to proteins (i.e. receptors or ligands), allowing the comprehension at an atomic level of their structure/function relationship and their interaction.

The delay of computational studies of sugars is due to their challenging structure, including a high conformational flexibility, length and branching of their chains. Also, the structure of sugar can greatly differ among diverse cell types since their synthesis is not under direct control of genes but subjected to the action of different modifying enzymes. On the other side, there is a shortage in algorithms specifically dedicated to study sugar and their interaction with proteins.

It derives that the development of new and reliable algorithms and workflows dedicated to sugars and aimed at overcoming the above mentioned limitations is essential for the throughout understanding of molecular processes underlying the arise and progression of those diseases in which sugars are involved. Equally important, appropriated computational models will speed up drug discovery and the development of sugar-based therapeutic strategies. This scenario is tantalizing as it lays the foundation for future perspectives of this journey.

In this path, my research work has been devoted to: i) develop novel computational methods to pinpoint the role of heparin and HSPGs in different pathological settings including viral infection and tumor neovascularization; ii) set up a new pipeline to allow the molecular characterization of the enzymatic reaction leading to sialylation of receptors involved in pathological neovascularization. This double-aimed purpose allowed the unmasking of shared molecular mechanisms of action and overlapped interactions of heparin/HSPGs and Sia in the pathogenesis of diseases that are very different when considered at an etiological and clinical level (Fig. 21).

The results obtained point to the necessity to adopt a mixed approach that merges experimental techniques and computational molecular models to better understand the role and mechanism of action of sugars in the pathological settings taken in consideration. Also, the models here developed for the different sugars and proteins (Fig. 21) will help to identify druggable pockets in their structure to be used as targets for the developments of compounds able to prevent pathological sugar-protein interactions to be transformed in novel therapies for sugar-based pathologies.

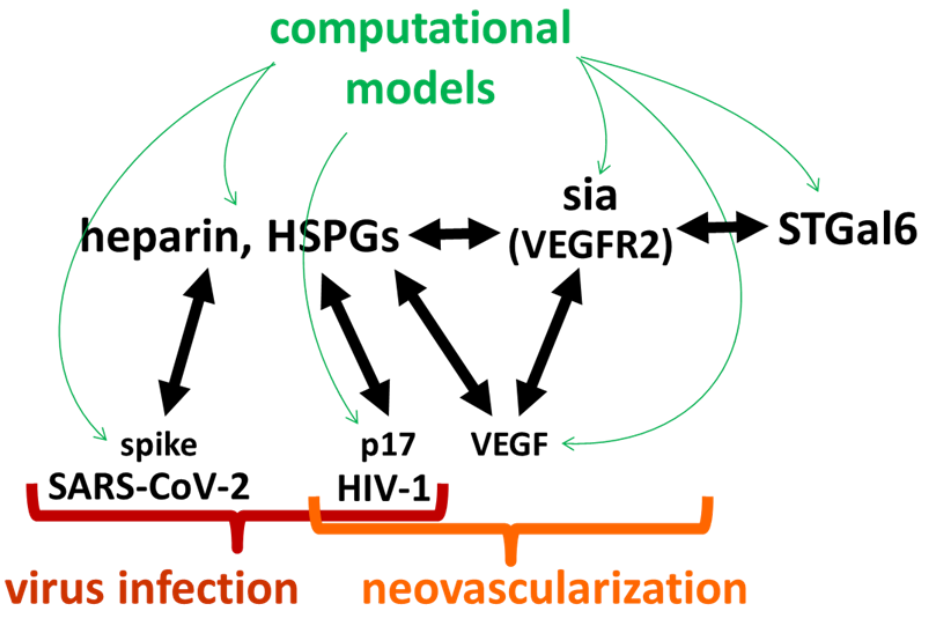



***Fig. 20.*** *Schematic representation of the computational studies performed in my research work with highlighted the models developed for proteins and sugars, their mutual interactions and the pathological processes in which they are involve*

# Bibliography


1. Shah ET, Upadhyaya A, Philp LK, Tang T, Skalamera D, Gunter J, et al. Repositioning "old" drugs for new causes: identifying new inhibitors of prostate cancer cell migration and invasion. Clin Exp Metastasis. 2016;33(4):385-99.
2. Tat J, Liu M, Wen XY. Zebrafish cancer and metastasis models for in vivo drug discovery. Drug Discov Today Technol. 2013;10(1):e83-9.
3. Zhu Q, Tao C, Shen F, Chute CG. Exploring the pharmacogenomics knowledge base (PharmGKB) for repositioning breast cancer drugs by leveraging Web ontology language (OWL) and cheminformatics approaches. Pac Symp Biocomput. 2014:172-82.
4. Lee H, Kang S, Kim W. Drug Repositioning for Cancer Therapy Based on Large-Scale Drug-Induced Transcriptional Signatures. PLoS One. 2016;11(3):e0150460.
5. Richon A. An introduction to molecular modeling,. Mathematech. 1994;1:83-90.
6. Hollingsworth SA, Dror RO. Molecular Dynamics Simulation for All. Neuron. 2018;99(6):1129-43.
7. Pinzi L, Rastelli G. Molecular Docking: Shifting Paradigms in Drug Discovery. Int J Mol Sci. 2019;20(18).
8. Pérez S, Tvaroška I. Carbohydrate-protein interactions: molecular modeling insights. Adv Carbohydr Chem Biochem. 2014;71:9-136.
9. Xu D, Esko JD. Demystifying heparan sulfate-protein interactions. Annu Rev Biochem. 2014;83:129-57.
10. Dwek RA. Glycobiology: Toward Understanding the Function of Sugars. Chem Rev. 1996;96(2):683-720.
11. Kornfeld R, Kornfeld S. Assembly of asparagine-linked oligosaccharides. Annu Rev Biochem. 1985;54:631-64.
12. Vliegenthart JFG. Introduction to NMR Spectroscopy of Carbohydrates. American Chemical Society. 2006.
13. Lengauer T, Rarey M. Computational methods for biomolecular docking. Curr Opin Struct Biol. 1996;6(3):402-6.
14. Nivedha AK, Thieker DF, Makeneni S, Hu H, Woods RJ. Vina-Carb: Improving Glycosidic Angles during Carbohydrate Docking. J Chem Theory Comput. 2016;12(2):892-901.
15. Zundert GCPv, Rodrigues JPGLM, Trellet M, Schmitz C, Kastritis PL, Karaca E, et al. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. J Mol Biol. 2016;428(4):720-5.

16. Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C, et al. The ClusPro web server for protein-protein docking. Nat Protoc. 2017;12(2):255-78.

17. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem. 2009;30(16):2785-91.

18. McCammon JA, Gelin BR, Karplus M. Dynamics of folded proteins. Nature. 1977;267, pages585–590:585–90.

19. Levitt M, Lifson S. Refinement of protein conformations using a macromolecular energy minimization procedure. J Mol Biol. 1969;14(2):269-79.

20. Frank M, Schloissnig S. Bioinformatics and molecular modeling in glycobiology. Cell Mol Life Sci. 2010;67(16):2749-72.

21. Casalino L, Gaieb Z, Goldsmith JA, Hjorth CK, Dommer AC, Harbison AM, et al. Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. ACS Cent Sci. 2020;6(10):1722-34.

22. Frenkel D, Smit B. Understanding Molecular Simulation. From Algorithms to Applications. edition n, editor2001.

23. Kulakova L, Arampatzis G, Angelikopoulos P, Hadjidoukas P, Papadimitriou C, Koumoutsakos P. Data driven inference for the repulsive exponent of the Lennard-Jones potential in molecular dynamics simulations. Sci Rep. 2017;7(1):16576.

24. Rahman AR, Bishov S, Westcott DE. Reversible Compression of Dehydrated Peas. J Texture Stud. 1971;2(2):240-5.

25. Jorgensen WL, Chandrasekhar J, Madura JD. J Chem Phys. 1983;79:926-35.

26. Kirschner KN, Yongye AB, Tschampel SM, Gonzalez-Outeirino J, Daniels CR, Foley BL, et al. GLYCAM06: a generalizable biomolecular force field. Carbohydrates. J Comput Chem. 2008;29(4):622-55.

27. Kirschner KN, Woods RJ. Solvent interactions determine carbohydrate conformation. Proc Natl Acad Sci U S A. 2001;98(19):10541-5.

28. Allen MP. Introduction to Molecular Dynamics Simulation. In: Norbert Attig KB, Helmut Grubmuller ¨ , Kurt Kremer, editor. Computational Soft Matter: From Synthetic Polymers to Proteins. 232004. p. 1-28.

29. Darden T, York D, Pedersen L. Particle mesh Ewald: An N · log(N) method for Ewald sums in large systems. The Journal of Chemical Physics. 1993;98(12):10089- 92.

30. Tuckerman ME, Martyna GJ. Understanding Modern Molecular Dynamics: Techniques and Applications. J Phys Chem B. 2001;105(7598).

31. Berendsen HJC, Postma JPM, Gunsteren WFv, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys. 1984;81:3684-90.

32. Sugita Y, Okamoto Y. Replica-exchange molecular dynamics method for protein folding. Chemical Physics Letters. 1999; 314:141–51.

33. R Q, G W, B M, R N. Replica Exchange Molecular Dynamics: A Practical Application Protocol with Solutions to Common Problems and a Peptide Aggregation and Self-Assembly Example. Methods in Molecular Biology. 2018;1777:101-19.

34. Hamelberg D, Mongan J, McCammon JA. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. J Chem Phys. 2004;120(24):11919-29.

35. Jin L, Abrahams JP, Skinner R, Petitou M, Pike RN, Carrell RW. The anticoagulant activation of antithrombin by heparin. Proc Natl Acad Sci U S A. 1997;94(26):14683-8.

36. Cardin AD, Weintraub HJ. Molecular modeling of protein-glycosaminoglycan interactions. Arteriosclerosis. 1989;9(1):21-32.

37. Bugatti A, Paiardi G, Urbinati C, Chiodelli P, Orro A, Uggeri M, et al. Heparin and heparan sulfate proteoglycans promote HIV-1 p17 matrix protein oligomerization: computational, biochemical and biological implications. Sci Rep. 2019;9(1):15768.

38. PW P, O R, M B. Cell surface heparan sulfate proteoglycans: selective regulators of ligand-receptor encounters. J Biol Chem. 2000;275:29923-6.

39. Christianson HC, Belting M. Heparan sulfate proteoglycan as a cell-surface endocytosis receptor. Matrix Biol. 2014;35:51-5.

40. Nagarajan A, Malvi P, Wajapeyee N. Heparan Sulfate and Heparan Sulfate Proteoglycans in Cancer Initiation and Progression. Front Endocrinol (Lausanne). 2018;24:483.

41. Chiodelli P, Mitola S, Ravelli C, Oreste P, Rusnati M, Presta M. Heparan sulfate proteoglycans mediate the angiogenic activity of the vascular endothelial growth factor receptor-2 agonist gremlin. Arterioscler Thromb Vasc Biol. 2011;31(12):e116-27.

42. Cagno V, Tseligka ED, Jones ST, Tapparel C. Heparan Sulfate Proteoglycans and Viral Attachment: True Receptors or Adaptation Bias? Viruses. 2019;11(7).

43. Cannon PM, Matthews S, Clark N, Byles ED, Iourin O, Hockley DJ, et al. Structure-function studies of the human immunodeficiency virus type 1 matrix protein, p17. J Virol. 1997;71(5):3474-83.

44. Zeinolabediny Y, Caccuri F, Colombo L, Morelli F, Romeo M, Rossi A, et al. HIV-1 matrix protein p17 misfolding forms toxic amyloidogenic assemblies that induce neurocognitive disorders. Sci Rep. 2017;7(1):10313.

45. Martorelli D, Muraro E, Mastorci K, Dal Col J, Fae DA, Furlan C, et al. A natural HIV p17 protein variant up-regulates the LMP-1 EBV oncoprotein and promotes the growth of EBV-infected

B-lymphocytes: implications for EBV-driven lymphomagenesis in the HIV setting. Int J Cancer. 2015;137(6):1374-85.

46. Caccuri F, Giagulli C, Bugatti A, Benetti A, Alessandri G, Ribatti D, et al. HIV-1 matrix protein p17 promotes angiogenesis via chemokine receptors CXCR1 and CXCR2. Proc Natl Acad Sci U S A. 2012;109(36):14580-5.

47. de Francesco Daher E, de Sousa Barros FA, da Silva Junior GB, Takeda CF, Mota RM, Ferreira MT, et al. Risk factors for death in acquired immunodeficiency syndrome-associated disseminated histoplasmosis. Am J Trop Med Hyg. 2006;74(4):600-3.

48. Rusnati M, Oreste P, Zoppetti G, Presta M. Biotechnological engineering of heparin/heparan sulphate: a novel area of multi-target drug discovery. Curr Pharm Des. 2005;11(19):2489-99.

49. Bugatti A, Giagulli C, Urbinati C, Caccuri F, Chiodelli P, Oreste P, et al. Molecular interaction studies of HIV-1 matrix protein p17 and heparin: identification of the heparin-binding motif of p17 as a target for the development of multitarget antagonists. J Biol Chem. 2013;288(2):1150-61.

50. Roderiquez G, Oravecz T, Yanagishita M, Bou-Habib DC, Mostowski H, Norcross MA. Mediation of human immunodeficiency virus type 1 binding by interaction of cell surface heparan sulfate proteoglycans with the V3 region of envelope gp120-gp41. J Virol. 1995;69(4):2233-9.

51. Urbinati C, Nicoli S, Giacca M, David G, Fiorentini S, Caruso A, et al. HIV-1 Tat and heparan sulfate proteoglycan interaction: a novel mechanism of lymphocyte adhesion and migration across the endothelium. Blood. 2009;114(15):3335-42.

52. Rusnati M, Tulipano G, Spillmann D, Tanghetti E, Oreste P, Zoppetti G, et al. Multiple interactions of HIV-I Tat protein with size-defined heparin oligosaccharides. J Biol Chem. 1999;274(40):28198-205.

53. Hoogewerf AJ, Kuschert GS, Proudfoot AE, Borlat F, Clark-Lewis I, Power CA, et al. Glycosaminoglycans mediate cell surface oligomerization of chemokines. Biochemistry. 1997;36(44):13570-8.

54. Matteis MAD, Godi A. PI-loting membrane traffic. Nat Cell Biol. 2004;6(6):487-92.

55. Verli H, Calazans A, Brindeiro R, Tanuri A, Guimaraes JA. Molecular dynamics analysis of HIV-1 matrix protein: clarifying differences between crystallographic and solution structures. J Mol Graph Model. 2007;26(1):62-8.

56. Hill CP, Worthylake D, Bancroft DP, Christensen AM, Sundquist WI. Crystal structures of the trimeric human immunodeficiency virus type 1 matrix protein: implications for membrane association and assembly. Proc Natl Acad Sci U S A. 1996;93(7):3099-104.

57. Mottarella SE, Beglov D, Beglova N, Nugent MA, Kozakov D, Vajda S. Docking server for the identification of heparin binding sites on proteins. J Chem Inf Model. 2014;54(7):2068-78.

58. Babik S, Samsonov SA, Pisabarro MT. Computational drill down on FGF1-heparin interactions through methodological evaluation. Glycoconj J. 2017;34(3):427-40.

59. Brown AJ, Sepuru KM, Rajarathnam K. Structural Basis of Native CXCL7 Monomer Binding to CXCR2 Receptor N-Domain and Glycosaminoglycan Heparin. Int J Mol Sci. 2017;18(3).

60. Uciechowska-Kaczmarzyk U, Babik S, Zsila F, Bojarski KK, Beke-Somfai T, Samsonov SA. Molecular dynamics-based model of VEGF-A and its heparin interactions. J Mol Graph Model. 2018;82:157-66.

61. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem. 2004;25(13):1605-12.

62. Case DA, Babin V, Berryman JT, Betz RM, al. e. AMBER2014. University of California, San Francisco2014.

63. Drabik P, Liwo A, Czaplewski C, Ciarkowski J. The investigation of the effects of counterions in protein dynamics simulations. Protein Eng. 2001;14(10):747-52.

64. Ohori Y, Okazaki H, Watanabe S, Tochio N, Arai M, Kigawa T, et al. Flexible and rigid structures in HIV-1 p17 matrix protein monitored by relaxation and amide proton exchange with NMR. Biochim Biophys Acta. 2014;1844(3):520-6.

65. Chiodelli P, Bugatti A, Urbinati C, Rusnati M. Heparin/Heparan sulfate proteoglycans glycomic interactome in angiogenesis: biological implications and therapeutical use. Molecules. 2015;20(4):6342-88.

66. Chiti F, Dobson CM. Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade. Annu Rev Biochem. 2017;86:27-68.

67. Ornitz DM, Yayon A, Flanagan JG, Svahn CM, Levi E, Leder P. Heparin is required for cell-free binding of basic fibroblast growth factor to a soluble receptor and for mitogenesis in whole cells. Mol Cell Biol. 1992;12(1):240-7.

68. Jaimes JA, Andre NM, Chappie JS, Millet JK, Whittaker GR. Phylogenetic Analysis and Structural Modeling of SARS-CoV-2 Spike Protein Reveals an Evolutionary Distinct and Proteolytically Sensitive Activation Loop. J Mol Biol. 2020;432(10):3309-25.

69. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. A new coronavirus associated with human respiratory disease in China. Nature. 2020;579(7798):265-9.

70. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell. 2020;183(6):1735.

71. Hoffmann M, Kleine-Weber H, Pohlmann S. A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells. Mol Cell. 2020;78(4):779-84 e5.

72. Watanabe Y, Allen JD, Wrapp D, McLellan JS, Crispin M. Site-specific glycan analysis of the SARS-CoV-2 spike. Science. 2020;369(6501):330-3.

73. Clausen TM, Sandoval DR, Spliid CB, Pihl J, Perrett HR, Painter CD, et al. SARS-CoV-2 Infection Depends on Cellular Heparan Sulfate and ACE2. Cell. 2020;183(4):1043-57 e15.

74. Kim SY, Jin W, Sood A, Montgomery DW, Grant OC, Fuster MM, et al. Characterization of heparin and severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) spike glycoprotein binding interactions. Antiviral Res. 2020;181:104873.

75. Hippensteel JA, LaRiviere WB, Colbert JF, Langouet-Astrie CJ, Schmidt EP. Heparin as a therapy for COVID-19: current evidence and future possibilities. Am J Physiol Lung Cell Mol Physiol. 2020;319(2):L211-L7.

76. Rusnati M, Vicenzi E, Donalisio M, Oreste P, Landolfo S, Lembo D. Sulfated K5 Escherichia coli polysaccharide derivatives: A novel class of candidate antiviral microbicides. Pharmacol Ther. 2009;123(3):310-22.

77. Sbalzarini IF, Greber UF. How Computational Models Enable Mechanistic Insights into Virus Infection. Methods Mol Biol. 2018;1836:609-31.

78. Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, Klebe G, et al. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. Nucleic Acids Res. 2007;35(Web Server issue):W522-5.

79. Case DA BK, Ben-Shalom IY, Brozell SR, Cerutti DS, et al. AMBER 2020. University of California, San Francisco. 2020.

80. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J Chem Theory Comput. 2015;11(8):3696-713.

81. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph. 1996;14(1):33-8, 27-8.

82. Kokh DB, Doser B, Richter S, Ormersbach F, Cheng X, RC. W. A Workflow for Exploring Ligand Dissociation from a Macromolecule: Efficient Random Acceleration Molecular Dynamics Simulation and Interaction Fingerprints Analysis of Ligand Trajectories. J Chem Phys. 2020;158

83. Hubbard S, Thornton J. NACCESS. 1992-6.

84. Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature. 2020;581(7807):215-20.

85. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. Nat Med. 2020;26(4):450-2.

86. Mycroft-West CJ, Su D, Pagani I, Rudd TR, Elli S, Gandhi NS, 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. 2020;120(12):1700-15.

87. Plouet J, Moro F, Bertagnolli S, Coldeboeuf N, Mazarguil H, Clamens S, et al. Extracellular cleavage of the vascular endothelial growth factor 189-amino acid form by urokinase is required for its mitogenic effect. J Biol Chem. 1997;272(20):13390-6.

88. Wijelath E, Namekata M, Murray J, Furuyashiki M, Zhang S, Coan D, et al. Multiple mechanisms for exogenous heparin modulation of vascular endothelial growth factor activity. J Cell Biochem. 2010;111(2):461-8.

89. Cochran S, Li CP, Ferro V. A surface plasmon resonance-based solution affinity assay for heparan sulfate-binding proteins. Glycoconj J. 2009;26(5):577-87.

90. Gitay-Goren H, Soker S, Vlodavsky I, Neufeld G. The binding of vascular endothelial growth factor to its receptors is dependent on cell surface-associated heparin-like molecules. J Biol Chem. 1992;267(9):6093-8.

91. Robinson CJ, Mulloy B, Gallagher JT, Stringer SE. VEGF165-binding sites within heparan sulfate encompass two highly sulfated domains and can be liberated by K5 lyase. J Biol Chem. 2006;281(3):1731-40.

92. Beckouche N, Bignon M, Lelarge V, Mathivet T, Pichol-Thievend C, Berndt S, et al. The interaction of heparan sulfate proteoglycans with endothelial transglutaminase-2 limits VEGF165-induced angiogenesis. Sci Signal. 2015;8(385):ra70.

93. Koehler L, Ruiz-Gomez G, Balamurugan K, Rother S, Freyse J, Moller S, et al. Dual Action of Sulfated Hyaluronan on Angiogenic Processes in Relation to Vascular Endothelial Growth Factor-A. Sci Rep. 2019;9(1):18143.

94. Samsonov SA, Freza S, Zsila F. In silico analysis of heparin and chondroitin sulfate binding mechanisms of the antiprotozoal drug berenil and pentamidine. Carbohydr Res. 2019;482:107742.

95. Balogh G, Gyongyosi T, Timari I, Herczeg M, Borbas A, Feher K, et al. Comparison of Carbohydrate Force Fields Using Gaussian Accelerated Molecular Dynamics Simulations and Development of Force Field Parameters for Heparin-Analogue Pentasaccharides. J Chem Inf Model. 2019;59(11):4855-67.

96. Grant OC, Woods RJ. Recent advances in employing molecular modelling to determine the specificity of glycan-binding proteins. Curr Opin Struct Biol. 2014;28:47-55.

97. Teran M, Nugent MA. Synergistic Binding of Vascular Endothelial Growth Factor-A and Its Receptors to Heparin Selectively Modulates Complex Affinity. J Biol Chem. 2015;290(26):16451-62.
98. Ornitz DM. FGFs, heparan sulfate and FGFRs: complex interactions essential for development. Bioessays. 2000;22(2):108-12.
99. Schauer R. Sialic acids as regulators of molecular and cellular interactions. Curr Opin Struct Biol. 2009;19(5):507-14.
100. Bhide GP, Colley KJ. Sialylation of N-glycans: mechanism, cellular compartmentalization and function. Histochem Cell Biol. 2017;147(2):149-74.
101. Buschiazzo A, Alzari PM. Structural insights into sialic acid enzymology. Curr Opin Chem Biol. 2008;12(5):565-72.
102. Harduin-Lepers A, Krzewinski-Recchi MA, Colomb F, Foulquier F, Groux-Degroote S, Delannoy P. Sialyltransferases functions in cancers. Front Biosci (Elite Ed). 2012;4:499-515.
103. Guillot A, Dauchez M, Belloy N, Jonquet J, Duca L, Romier B, et al. Impact of sialic acids on the molecular dynamic of bi-antennary and tri-antennary glycans. Sci Rep. 2016;6:35666.
104. Groux-Degroote S, Guerardel Y, Delannoy P. Gangliosides: Structures, Biosynthesis, Analysis, and Roles in Cancer. Chembiochem. 2017;18(13):1146-54.
105. Szabo R, Skropeta D. Advancement of Sialyltransferase Inhibitors: Therapeutic Challenges and Opportunities. Med Res Rev. 2017;37(2):219-70.
106. Bauer J, Osborn HM. Sialic acids in biological and therapeutic processes: opportunities and challenges. Future Med Chem. 2015;7(16):2285-99.
107. Schnaar RL. Glycobiology simplified: diverse roles of glycan recognition in inflammation. J Leukoc Biol. 2016;99(6):825-38.
108. West DC, Rees CG, Duchesne L, Patey SJ, Terry CJ, Turnbull JE, et al. Interactions of multiple heparin binding growth factors with neuropilin-1 and potentiation of the activity of fibroblast growth factor-2. J Biol Chem. 2005;280(14):13457-64.
109. Chiodelli P, Rezzola S, Urbinati C, Federici Signori F, Monti E, Ronca R, et al. Contribution of vascular endothelial growth factor receptor-2 sialylation to the process of angiogenesis. Oncogene. 2017;36(47):6531-41.
110. Chiu CP, Watts AG, Lairson LL, Gilbert M, Lim D, Wakarchuk WW, et al. Structural analysis of the sialyltransferase CstII from Campylobacter jejuni in complex with a substrate analog. Nat Struct Mol Biol. 2004;11(2):163-70.
111. Rao FV, Rich JR, Rakic B, Buddai S, Schwartz MF, Johnson K, et al. Structural insight into mammalian sialyltransferases. Nat Struct Mol Biol. 2009;16(11):1186-8.

112. Chandler KB, Leon DR, Meyer RD, Rahimi N, Costello CE. Site-Specific N-Glycosylation of Endothelial Cell Receptor Tyrosine Kinase VEGFR-2. J Proteome Res. 2017;16(2):677-88.

113. Kuhn B, Benz J, Greif M, Engel AM, Sobek H, Rudolph MG. The structure of human alpha-2,6-sialyltransferase reveals the binding mode of complex glycans. Acta Crystallogr D Biol Crystallogr. 2013;69(Pt 9):1826-38.

# Acknowledgements

Ringrazio il Professor Marco Rusnati
per avermi insegnato che solo dal confronto si cresce e solo due “figli unici” possono scontrarsi ad armi pari.

Ringrazio il Professor Paolo Bergese
per avermi insegnato che l’unico ostacolo alla ricerca sono i limiti della nostra mente.

Ringrazio il Professor Luciano Milanesi, la Dottoressa Pasqualina D’Ursi e tutto il gruppo del CNR di Milano per avermi iniziata ed accompagnata nei primi passi di questo viaggio.

I would like to thank Professoressa Rebecca Wade for being an incredible mentor, Professoressa Daria Kokh, Stefan Richer and all my incredible colleagues at the Heidelberg Institute for Theoretical Studies.

Ringrazio le mie colleghe Paola Chiodelli e Chiara Urbinati per questi 3 anni insieme.

Ringrazio chi ha condiviso con me questo viaggio.

Ringrazio la mia famiglia, perchè tutto questo è solo grazie a voi.

Infine, ringrazio Alfio.