

PHD THESIS

Deciphering the Molecular Basis of Specificity in Protein-Glycosaminoglycan Interactions: from All-Atom to Coarse-Grained Models

Faculty of Chemistry
University of Gdańsk

M. Sc. Annemarie Danielsson

Supervisor: prof. dr Sergey A. Samsonov
Co-supervisor: dr Martyna Maszota-Zieleniak

Gdańsk, Poland
2025

ROZPRAWA DOKTORSKA

Rozszyfrowanie molekularnych
podstaw specyficzności
glikozaminoglikanów w interakcjach
z białkami: od modeli
pełnoatomowych do
gruboziarnistych

Wydział Chemii
Uniwersytet Gdański

M. Sc. Annemarie Danielsson

Promotor: prof. dr Sergey A. Samsonov
Promotor pomocniczy: dr Martyna Maszota-Zieleniak

Gdańsk, Polska
2025

Acknowledgements

The results presented in this PhD dissertation were obtained with the support of the National Science Centre (project "Mechanistic insights into the specificity of glycosaminoglycan interactions with regulatory proteins", UMO-2018/31/G/ST4/00246). I gratefully acknowledge Polish high-performance computing infrastructure PLGrid (HPC Center: ACK Cyfronet AGH) for providing computer facilities and support within computational grants no. PLG/2020/014274, PLG/2020/014275, PLG/2022/015663, PLG/2022/015664, PLG/2023/016545, PLG/2023/016581. Calculations were also conducted using the computational resources of the cluster at the Faculty of Chemistry, University of Gdańsk.

Wyniki przedstawione w niniejszej rozprawie doktorskiej zostały uzyskane dzięki wsparciu Narodowego Centrum Nauki (projekt „Mechanistic insights into the specificity of glycosaminoglycan interactions with regulatory proteins”, UMO-2018/31/G/ST4/00246). Pragnę podziękować Polskiej Infrastrukturze PLGrid (Centrum HPC: ACK Cyfronet AGH) za udostępnienie sprzętu komputerowego i wsparcie w ramach grantów obliczeniowych nr. PLG/2020/014274, PLG/2020/014275, PLG/2022/015663, PLG/2022/015664, PLG/2023/016545, PLG/2023/016581. Obliczenia przeprowadzono również z wykorzystaniem zasobów obliczeniowych klastra na Wydziale Chemii Uniwersytetu Gdańskiego.

Annemarie Danielsson

2025

Abstract

Glycosaminoglycans (GAGs) are linear anionic polysaccharides composed of repeating disaccharide units made up of a hexosamine and a uronic acid with varying degrees and patterns of sulfation. They are essential components of the extracellular matrix and are pivotal in numerous cellular processes by interacting with various proteins. Disruptions in these interactions can cause severe health conditions including neurodegenerative and inflammatory diseases.

The specificity of interactions between GAGs and proteins, i.e. the ability of the receptor (protein) or ligand molecule (GAG) to distinguish between similar targets and preferentially form interactions with one or a few selected binding partners, is largely governed by the structural and chemical properties of GAGs, including their sulfation patterns, charge distribution, length, and flexibility, as well as the protein's structural features. Disruptions in binding specificity caused by protein mutations, alterations in GAG structure, or changes in the cellular environment can result in protein dysfunction, dysregulated signaling, and disease. Therefore, understanding binding specificity is crucial for developing GAG mimetics for therapeutic use to recreate, enhance, or modify natural GAG binding.

Isolating structurally homogeneous fragments of unbound GAGs is difficult due to the variations in their chain length, sugar composition, and sulfation pattern. Once GAGs bind to proteins, their increased conformational flexibility and transient interactions become the chief obstacles to structural analysis. The dynamic multivalent nature of protein/GAG interactions, often driven by electrostatics, can result in multiple binding poses with comparable affinities, which complicates the detailed characterization of specificity of individual binding sites *in vitro*. A combination of experimental and computational methods is beneficial to address these challenges and provide a thorough understanding of protein/GAG specificity.

Computational methods elucidate binding mechanisms and the effects of specific molecular features of GAGs at the molecular level. Mapping the electrostatic potential using theoretical models helps to identify protein regions that are likely to interact with GAGs. Molecular docking and molecular dynamics simulations

are used to identify potential binding sites and poses, and to characterize interaction stability and conformational changes within the protein/GAG complex. All-atom simulations provide atomistic detail while coarse-grained simulations are able to capture larger scales over extended timescales owing to the grouping of atoms into "beads". By quantifying the energetic cost or gain of binding, free energy calculations provide quantitative insight into the thermodynamic favorability of protein/GAG binding. Other approaches, such as machine learning, can classify interactions based on the structural properties of both proteins and GAGs, revealing hidden patterns in their binding mechanisms and highlighting key features driving binding specificity.

The research presented in this thesis aimed to explore areas currently beyond experimental reach in the study of protein/GAG binding specificity using computational methods. The employed computational models were integrated with and validated against experimental data to provide a comprehensive, multi-scale understanding of protein/GAG interactions. Using Poisson–Boltzmann calculations, I quantified how different GAGs affect the protein’s near-surface electrostatic field and validated these changes by comparing the computed maps with measurements from a novel paramagnetic-probe NMR technique. Next, I performed atomistic molecular dynamics simulations of protein/GAG and protein/peptide complexes to explore the influence of electrostatic interactions versus other physicochemical properties on binding specificity. Free energy calculations decomposing the binding energy into separate contributions and analysis of ligand flexibility distinguished GAG interactions from acidic peptides. Furthermore, I implemented, fine-tuned, and validated a coarse-grained model of heparin. Finally, I performed MD simulations of GAGs with different sulfation patterns in combination with unsupervised learning to establish links between GAG structural features and binding specificity.

By utilizing computational chemistry techniques, the research presented in this thesis made key contributions to understanding the specificity of protein/GAG interactions. Innovative tools and analysis pipelines were designed for coarse-grained simulations, enhancing the efficiency of protein/GAG interaction modeling. This research demonstrated the complementary nature of computational and experimental methods in studying protein/GAG complexes. The developed computational approaches support large-scale simulations and offer a robust framework for future studies on protein/GAG binding specificity relevant to therapeutic design.

Streszczenie

Glikozaminoglikany (GAGi) to liniowe anionowe polisacharydy zbudowane z powtarzających się jednostek dwucukrowych, składających się z heksozaminy i kwasu uronowego, wykazujących zróżnicowany stopień i wzór sulfatacji. Są istotnymi składnikami macierzy pozakomórkowej i odgrywają kluczową rolę w licznych procesach komórkowych poprzez oddziaływania z różnymi białkami. Specyficzność oddziaływań między GAGami a białkami, czyli zdolność receptora lub ligandu do rozróżniania podobnych do siebie partnerów wiązania i zdolność preferencyjnego tworzenia interakcji z jednym lub kilkoma wybranymi molekułami zależy w dużej mierze od właściwości strukturalnych i chemicznych GAGów, a także od cech strukturalnych samego białka. Zmiany wpływające na specyficzność wiązania mogą powodować zaburzenia w sygnalizacji między- i wewnątrzkomórkowej i wywoływać rozwój chorób takich jak choroby neurodegeneracyjne, nowotworowe, czy autoimmunologiczne. Zrozumienie mechanizmów wiązania ma zatem kluczowe znaczenie dla opracowania mimetyków GAGów do zastosowań terapeutycznych, które mogłyby odtwarzać, wzmacniać lub modyfikować oddziaływania GAGów z białkami.

Złożoność strukturalna GAGów utrudnia eksperymentalną izolację jednorodnych strukturalnie fragmentów, natomiast ich konformacyjna elastyczność i przejściowy charakter oddziaływań z białkami dodatkowo utrudniają analizy strukturalne. Szczegółowa charakterystyka specyficzności poszczególnych miejsc wiązania GAGów *in vitro* jest utrudniona dla oddziaływań białko/GAG z powodu obierania przez cząsteczki GAGów wielu pozycji w przestrzeni o porównywalnych powinowactwach. Połączenie metod eksperymentalnych i obliczeniowych jest więc korzystne dla uzyskania pełniejszego zrozumienia specyficzności interakcji białko/GAG.

Metody obliczeniowe pozwalają na wyjaśnienie mechanizmów wiązania oraz wpływu określonych cech GAGów na poziomie molekularnym. Mapowanie potencjału elektrostatycznego za pomocą modeli teoretycznych pozwala na identyfikację regionów białka, które są prawdopodobnymi miejscami oddziaływań z GAGami z uwagi na dominującą rolę elektrostatyki w ich interakcjach. Dokowanie oraz symulacje dynamiki molekularnej mogą posłużyć do identyfikacji

potencjalnych miejsc wiązania oraz do charakterystyki stabilności oddziaływań i zachodzących w wyniku wiązania zmian konformacyjnych. Podczas gdy atomistyczna dynamika molekularna opisuje ruch i interakcje każdego pojedynczego atomu, podejścia gruboziarniste grupują kilka atomów w reprezentatywne "ziarna", zachowujące kluczowe cechy fizykochemiczne i umożliwiając tym redukcję kosztów obliczeniowych. Powinowactwo wiązania GAGów z białkami można oszacować obliczając energię swobodną towarzyszącą ich oddziaływaniom, co pozwala na poznanie termodynamiki wiązania. Inne podejścia, takie jak uczenie maszynowe, pozwalają na klasyfikowanie interakcji na podstawie cech strukturalnych zarówno białek, jak i GAGów, ujawniając przy tym ukryte wzorce w mechanizmach wiązania.

Badania przedstawione w niniejszej rozprawie miały na celu eksplorację z wykorzystaniem metod obliczeniowych obszarów obecnie niedostępnych dla badań eksperymentalnych w zakresie specyficzności wiązania białko/GAG. Zastosowane modele obliczeniowe zostały zintegrowane z i poddane weryfikacji w oparciu o dane eksperymentalne, aby uzyskać kompleksowy obraz oddziaływań białko/GAG. Wpływ wiązania różnych typów GAGów na potencjał elektrostatyczny powierzchni białka przeanalizowano metodami obliczeniowymi i porównano z wynikami spektroskopii NMR z użyciem sond paramagnetycznych. Następnie przeprowadzone zostały symulacje atomistycznej dynamiki molekularnej kompleksów białko/GAG oraz białko/peptyd w celu zbadania wpływu oddziaływań elektrostatycznych w porównaniu do innych właściwości fizykochemicznych na specyficzność wiązania. Obliczenia energii swobodnej wraz z analizą elastyczności ligandów pozwoliły odróżnić interakcje z GAGami od tych z anionowymi peptydami. Zaimplementowano i sparametryzowano gruboziarnisty model heparyny w celu symulacji struktury i dynamiki łańcuchów heparyny o różnej długości oraz weryfikacji skuteczności modelu w odtwarzaniu biologicznego zachowania tych cząsteczek. Przeprowadzone zostały również symulacje dynamiki molekularnej GAGów o różnych wzorach sulfatacji w połączeniu z nienadzorowanym uczeniem maszynowym, aby powiązać cechy strukturalne GAGów ze specyficznością wiązania.

Badania przedstawione w tej rozprawie miały znaczący wkład w zrozumienie specyficzności oddziaływań białko/GAG. Opracowano innowacyjne narzędzia i schematy analizy dla symulacji gruboziarnistych, zwiększając efektywność modelowania oddziaływań białko/GAG, a zastosowane podejścia obliczeniowe stanowią podstawę dla przyszłych badań nad specyficznością wiązania białko/GAG.

List of publications

1. **Danielsson A**, Kogut MM, Maszota-Zieleniak M, Chopra P, Boons GJ, Samsonov SA. Molecular dynamics-based descriptors of 3-O-Sulfated Heparan sulfate as contributors of protein binding specificity. *Comput Biol Chem.* 2022;99:107716. doi:10.1016/j.compbiolchem.2022.107716.
2. Schulze C*, **Danielsson A***, Liwo A, Huster D, Samsonov SA, Penk A. Ligand binding of interleukin-8: a comparison of glycosaminoglycans and acidic peptides. *Phys Chem Chem Phys.* 2023;25(36):24930–24947. doi:10.1039/d3cp02457a.
3. **Danielsson A**, Samsonov SA, Liwo A, Sieradzan AK. Extension of the SUGRES-1P Coarse-Grained Model of Polysaccharides to Heparin. *J Chem Theory Comput.* 2023;19(17):6023–6036. doi:10.1021/acs.jctc.3c00511.
4. Penk A*, **Danielsson A***, Gaardl s M, Montag C, Sch ler A, Huster D, Samsonov SA, K nze G. Detecting Protein-Ligand Interactions with Nitroxide Based Paramagnetic Cosolutes. *Chemistry.* 2024;30(18):e202303570. doi:10.1002/chem.202303570.
5. **Danielsson A**, Samsonov SA, Sieradzan AK. Implementation of the UNRES/SUGRES-1P Coarse-Grained Model of Heparin for Simulating Protein/Heparin Interactions. *J Chem Theory Comput.* 2024;20(23):10703-10715. doi:10.1021/acs.jctc.4c00575.

* These authors contributed equally.

Contents

Acknowledgments	i
Abstract	iii
Streszczenie	v
List of Publications Comprising the Thesis	vii
1 Introduction	1
1.1 Glycosaminoglycans	2
1.1.1 Hyaluronic Acid	5
1.1.2 Heparan Sulfate and Heparin	5
1.1.3 Chondroitin Sulfate and Dermatan Sulfate	6
1.1.4 Keratan Sulfate	7
1.2 Specificity in Protein/GAG Interactions	7
1.3 Experimental Approaches to Study Protein/GAG Complexes . . .	10
1.3.1 X-ray Crystallography	11
1.3.2 Nuclear Magnetic Resonance	14
1.3.3 Electron Microscopy	16
1.3.4 Mass Spectrometry	16
1.3.5 Capillary Electrophoresis	17
1.3.6 Circular Dichroism	17
1.3.7 Fluorescence Spectroscopy	18
1.3.8 Surface Plasmon Resonance	18
1.3.9 Enzyme-Linked Immunosorbent Assay	19
1.3.10 Biolayer Interferometry	20
1.3.11 Calorimetry	21
1.4 <i>In silico</i> Methods to Study Protein/GAG Complexes	23
1.4.1 Electrostatic Potential Calculations	24
1.4.2 Molecular Docking	29
1.4.3 Molecular Dynamics	33
1.4.4 Binding Free Energy Calculations	45

2	Research Aims and Key Objectives	52
3	Publications Included in the PhD Thesis	54
3.1	Detecting Protein-Ligand Interactions with Nitroxide Based Paramagnetic Cosolutes	55
3.2	Ligand Binding of Interleukin-8: a Comparison of Glycosaminoglycans and Acidic Peptides	57
3.3	Extension of the SUGRES-1P Coarse-Grained Model of Polysaccharides to Heparin	61
3.4	Implementation of the UNRES/SUGRES-1P Coarse-Grained Model of Heparin for Simulating Protein/Heparin Interactions . .	65
3.5	Molecular Dynamics-Based Descriptors of 3-O-Sulfated Heparan Sulfate as Contributors of Protein Binding Specificity	68
3.6	Outlook	71
	Bibliography	72

Abbreviations

Galactose	Gal	All-Atom	AA
Glucosamine	GlcN	Coarse-Grained	CG
N-Acetylgalactosamine	GalNAc	Adaptive Poisson-Boltzmann Solver	APBS
N-Acetylglucosamine	GlcNAc	Poisson-Boltzmann	PB
6-O-Sulfated N-Acetylglucosamine	GlcNAc(6S)	Generalized Born	GB
N-Sulfated Glucosamine	GlcNS	Linear Interaction Energy	LIE
3-O-Sulfated N-Sulfated Glucosamine	GlcNS(3S)	Molecular Mechanics	MM
3-O- and 6-O-Sulfated N-Sulfated Glucosamine	GlcNS(3,6S)	Molecular Dynamics	MD
6-O-Sulfated N-Sulfated Glucosamine	GlcNS(6S)	Replica-Exchange Molecular Dynamics with Repulsive Scaling	RS-REMD
Glucuronic Acid	GlcA	Molecular Mechanics / Poisson-Boltzmann Surface Area	MM/PBSA
Iduronic Acid	IdoA	Molecular Mechanics / Generalized Born Surface Area	MM/GBSA
Chondroitin Sulfate	CS	Quantum Mechanics	QM
Dermatan Sulfate	DS	Quantum Mechanics / Molecular Mechanics	QM/MM
Heparan Sulfate	HS	Periodic Boundary Conditions	PBC
Heparin	HP	Electrostatic Potential	EP, ϕ
Keratan Sulfate	KS	End-to-End Distance	EED
Hyaluronic Acid	HA	Radial Distribution Function	RDF
Glycosaminoglycan	GAG	Root Mean Square Deviation	RMSD
Degree of Polymerization	dp	Root Mean Square Fluctuation	RMSF
Sulfotransferase	ST	Protein Data Bank	PDB
Acidic Fibroblast Growth Factor	aFGF	Chemical Shift Perturbation	CSP
Basic Fibroblast Growth Factor	bFGF	Nuclear Magnetic Resonance	NMR
Bone Morphogenetic Protein-2	BMP-2	X-ray Crystallography	X-ray
Hepatic Growth Factor / Scatter Factor	HGF/SF	Mass Spectrometry	MS
Interleukin-8	IL-8		
Neuropilin-1	Nrp-1		
Receptor for Advanced Glycation End Products	RAGE		
Antithrombin III	AT-III		
Extracellular Matrix	ECM		

List of Figures

1-1	Chemical structures of the six main GAG types.	3
1-2	Experimental approaches to study GAGs and protein/GAG interactions.	11
1-3	IL-8 and HP and the calculated EP isosurface obtained from APBS software, colored by EP isovalue.	25
1-4	Snapshots of an MD trajectory colored according to time step. . .	37
1-5	A CG representation of a protein and ligand complex.	44
3-1	ENS electrostatic potentials determined from the PRE NMR data of IL-8 apo and the IL-8/HP complex.	56
3-2	Electrostatic energy in vacuo, total electrostatic energy, van der Waals energy, and total free energy for the IL-8/ligand complexes determined using MM/GBSA.	58
3-3	Distributions of radius of gyration values for bound and unbound states for HP dp6, p3-, p5-, p7-, and p10-.	60
3-4	The SUGRES-1P model of HP.	62
3-5	Representative conformations of the CG trajectories of HP dp18, dp24, dp30, and dp36, colored according to trajectory time step. .	64
3-6	The coarse-grained UNRES and SUGRES-1P models.	66
3-7	Analysis pipeline for MD-derived descriptors of free GAGs, including correlation analysis, PCA, and hierarchical clustering.	69

List of Tables

1-1	Characteristics of the six major GAG types: charge per disaccharide unit, chemical composition and glycosidic linkage type in the disaccharide units, sulfation group position, typical size of the molecule <i>in vivo</i> expressed in degree of polymerization (dp). . . .	4
1-2	Comparison of Methods for Estimating Binding Free Energies in Computational Chemistry	46

1 Introduction

1.1 Glycosaminoglycans

Glycosaminoglycans (GAGs) are a family of linear anionic polysaccharides composed of repeating disaccharide units typically consisting of an amino sugar and either a uronic acid or galactose [1]. The glycosidic linkages between these sugars confer structural flexibility to GAG chains, a property essential for their biological functions. The unique sulfation patterns of GAGs, i.e., the distinctive arrangements of sulfate groups along their chains, function as a **sulfation code** that selects protein partners and modulates essential biological processes [2]. Together with epimerization and other modifications, this code enhances the functional diversity of GAGs across cell types and tissues [3]. Disruptions in the enzymatic processes that establish and regulate these patterns are associated with various genetic and metabolic disorders [3].

GAGs are ubiquitously produced by all mammalian cells and are either secreted into the extracellular matrix (ECM) or displayed on the cell surface, predominantly in the form of proteoglycans made up of a core protein and associated GAG chains such as aggrecans, small and leucine-rich decorins, and the transmembrane syndecans [1]. Four major classes of GAGs are recognized, each characterized by unique features regarding molecular structure, charge, and size: Hyaluronic Acid (HA), Heparan Sulfate/Heparin (HS/HP), Chondroitin Sulfate/Dermatan Sulfate (CS/DS), and Keratan Sulfate (KS) [1, 4]. These GAGs are synthesized and modified in the Golgi apparatus and endoplasmic reticulum, with the exception of HA, which is synthesized at the plasma or cytoplasmic membrane [5]. Degradation of GAGs primarily occurs in lysosomes, involving specific enzymes that cleave bonds within GAG chains [6]. Mutations in the genes encoding these enzymes can result in mucopolysaccharidoses (MPS), a group of disorders characterized by the accumulation of partially degraded GAGs.

The basic properties and disaccharide unit sequences of the GAG classes are summarized in Table 1-1 and Figure 1-1.

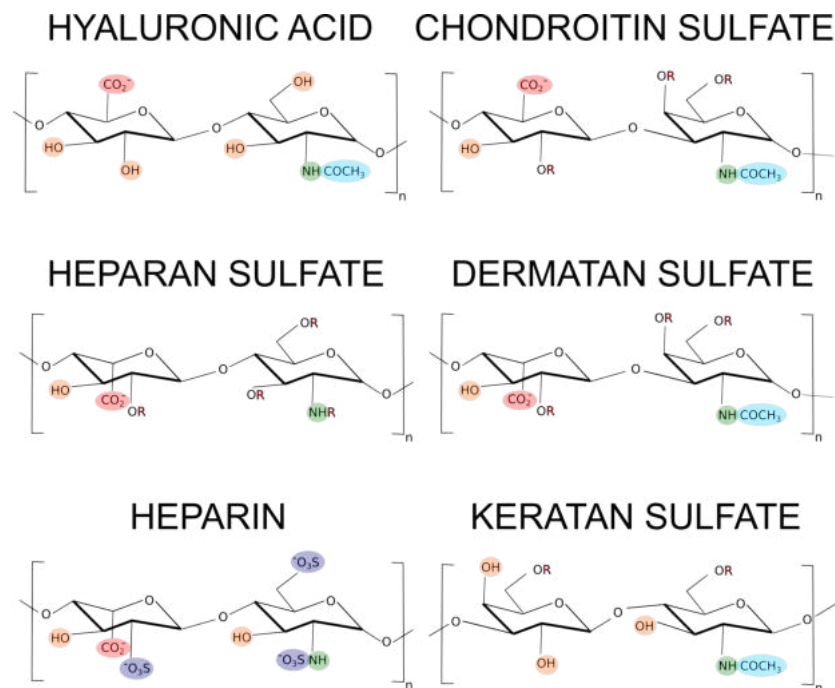


Figure 1-1: Chemical structures of the six major GAG types: hyaluronic acid, chondroitin sulfate, heparan sulfate, dermatan sulfate, heparin, and keratan sulfate. Key functional groups attached to the molecules are highlighted using colors: red for carboxyl groups ($-\text{CO}_2^-$), violet for sulfate groups ($-\text{SO}_3^-$), light blue for acetyl groups ($-\text{COCH}_3$), green for amino groups ($-\text{NH}_2$), and orange for hydroxyl groups ($-\text{OH}$). Red "R" symbols indicate potential sites for sulfate group addition.

Table **1-1**: Characteristics of the six major GAG types: charge per disaccharide unit, chemical composition and glycosidic linkage type in the disaccharide units, sulfation group position, typical size of the molecule *in vivo* expressed in degree of polymerization (dp).

Monosaccharide legend: GlcA – β -D-glucuronic acid; IdoA – α -L-iduronic acid; GlcN – D-glucosamine; GlcNS – N-sulfated D-glucosamine; GlcNAc – N-acetyl D-glucosamine; Gal – D-galactose; GalNAc – N-acetyl D-galactosamine.

Name	Charge per disaccharide	Chemical composition	Sulfation groups	Typical size:
Hyaluronic Acid (HA)	-1	GlcNAc ($\beta 1 \rightarrow 4$) GlcA ($\beta 1 \rightarrow 3$)	None	$15\text{-}30 \times 10^4$ dp
Heparan Sulfate (HS)	-1 to -4	GlcN ($\alpha 1 \rightarrow 3$) GlcA/IdoA ($\alpha 1 \rightarrow 4$)	N- & 3-,6- <i>O</i> 2- <i>O</i>	50 - 400 dp
Heparin (HP)	-4	GlcNS ($\alpha 1 \rightarrow 4$) IdoA ($\alpha 1 \rightarrow 4$)	N- & 3-,6- <i>O</i> 2- <i>O</i>	20 - 100 dp
Chondroitin Sulfate (CS)	-2 or -3	GalNAc ($\beta 1 \rightarrow 4$) GlcA ($\beta 1 \rightarrow 3$)	4-,6- <i>O</i> 2- <i>O</i>	80 - 200 dp
Dermatan Sulfate (DS)	-2	GalNAc ($\beta 1 \rightarrow 4$) IdoA ($\alpha 1 \rightarrow 3$)	4- <i>O</i> 2- <i>O</i>	100 - 400 dp
Keratan Sulfate (KS)	-2	GlcNAc ($\beta 1 \rightarrow 4$) Gal ($\beta 1 \rightarrow 4$)	6- <i>O</i>	10 - 70 dp

1.1.1 Hyaluronic Acid

HA is a unique GAG composed of non-sulfated disaccharide units of glucuronic acid (GlcA) and N-acetylglucosamine (GlcNAc), linked by alternating β -1,4 and β -1,3 glycosidic bonds [7]. Unlike other GAGs, HA has a notably high degree of polymerization with considerable variability in chain lengths, and it does not form covalent attachments to proteoglycan core proteins [7]. Under most physiological conditions, HA assumes a stiff random coil conformation, contributing to its stability and functional versatility [7].

HA plays an essential structural role in connective tissues by enhancing the viscoelastic properties of the ECM [8]. In joint synovial fluid, it serves as a lubricant, facilitating smooth movement and minimizing friction in articulating joints [9]. HA also contributes to the resilience of cartilage, allowing it to withstand compressive forces [10].

In addition to its mechanical functions, HA participates in several important biological processes. It interacts with cell surface receptors such as CD44 and the Receptor for Hyaluronan-Mediated Motility (RHAMM), playing roles in cellular signaling, tissue repair, and immune modulation. HA is also involved in regulating cell migration and proliferation through CD44, and contributes to the inflammatory response by interacting with Toll-like receptors (TLR2 and TLR4) [7].

1.1.2 Heparan Sulfate and Heparin

Owing to its immense structural variability, HS can be considered as a family of related GAGs of varying sequence and sulfation patterns which mediate specific binding [11]. HS consists predominantly of repeating units of GlcA or its epimer iduronic acid (IdoA) together with GlcNAc, linked by α -1,4 and β -1,4 glycosidic bonds [1]. HS molecules undergo a variety of modifications, including *N*-deacetylation and *N*-sulfation of GlcNAc residues to *N*-sulfo-glucosamine (GlcNS), resulting in a mixture of *N*-acetylated and *N*-sulfated regions along the HS chain (NA and NS domains, respectively) [12, 13]. The NS domains serve as starting points for further modifications, such as epimerization of GlcA to IdoA and sulfation of IdoA and GlcNAc/GlcNS residues. Sulfate groups can be attached also to oxygen atoms at different positions of IdoA and GlcNS, resulting in 2-*O*-sulfated IdoA (IdoA(2S)), 6-*O*-sulfated GlcNAc/GlcNS (GlcNAc(6S),GlcNS(6S)), and 3-*O*-sulfated GlcNS (GlcNS(3S)), as well as 3-*O*- and 6-*O*-sulfated GlcNS (GlcNS(3,6S)) [3]. Collectively, permutations of these modifications across the two residues yield 48 distinct dp2 sulfation patterns.

Although HS chains are present in almost every animal tissue, their sulfation patterns vary according to spatiotemporal expression differences, lending tissue- and developmental stage-specific biological functions to the molecule [14, 15]. HS is found within cells, on cell surfaces, and in the ECM [16], where it participates in signaling pathways by binding growth factors, cytokines, and morphogens [17]. This influences protein concentrations on cell surfaces and within the ECM, affecting cell adhesion, migration, and various disease processes including cancer and infectious diseases [11, 18].

One of the most heavily sulfated HS family member, heparin (HP) is composed of repeating disaccharides of 2-*O*-sulfated IdoA (IdoA(2S)) and N-sulfated, 6-*O*-sulfated glucosamine (GlcNS(6S)) residues [1]. Compared to most HS molecules, HP is characterized by less structural variability and greater negative charge [19]. Primarily synthesized in mast cells of the connective tissue, HP is distinguished by its potent anticoagulant properties, primarily through interaction with antithrombin III (AT-III) [20]. This interaction accelerates the inactivation of thrombin and factor Xa, key enzymes in the coagulation cascade. Beyond its anticoagulant properties, HP also interacts with various proteins to modulate inflammation, cell proliferation, and metastasis [21, 22, 23].

1.1.3 Chondroitin Sulfate and Dermatan Sulfate

CS is the most abundant GAG, consisting of alternating units of GlcA and N-acetylgalactosamine (GalNAc), connected by β -1,3 and β -1,4 glycosidic bonds [1, 24]. CS is categorized into four types based on variations in the sulfation patterns of the disaccharide units: CS-A (GlcA-GalNAc(4S)), CS-C (GlcA-GalNAc(6S)), CS-D (GlcA(2S)-GalNAc(6S)), and CS-E (GlcA-GalNAc(4,6S)) [1, 24]. These patterns influence interactions with proteins and signaling molecules in the ECM of cartilage tissues, increasing mechanical integrity and resistance to compressive forces [24].

The presence of IdoA linked to GalNAc is characteristic for DS [25, 26]. While CS and DS share many biosynthetic enzymes, specific sulfotransferase (ST) enzymes act on DS to attach sulfate groups to IdoA, preventing the reversible epimerization to GlcA and enhancing its flexibility compared to CS [1]. DS exhibits a high variability of sulfation patterns and is particularly important in the development of blood vessels, skin, and bone, and plays a role in tissue development and wound repair, including participation in cell signaling [25].

1.1.4 Keratan Sulfate

KS is composed of repeating disaccharide units of galactose (Gal) and GlcNAc, connected by alternating β -1,3 and β -1,4 glycosidic bonds [1]. It is divided into three types, each associated with different tissue locations and sulfation patterns. KSI, primarily located in the cornea, includes a mixture of non-sulfated, monosulfated, and disulfated disaccharides. KSII, important in cartilage, shares similar sulfation patterns with KSI but varies in sulfation degree and polysaccharide length [3]. KSIII, found in the brain, features the shortest and least sulfated chains [27].

KS influences cellular proliferation and migration, but its molecular mechanisms remain less well defined than for other GAGs [28, 29]. Its main functions include maintaining corneal hydration [30], and supporting the structural integrity of cartilage and bone through interactions with proteins such as osteoadherin [31], fibromodulin [32], and PRELP [27].

1.2 Specificity in Protein/GAG Interactions

In the context of biomolecular recognition, *specificity* denotes the preferential and selective binding of a ligand to its cognate target on the basis of complementary geometric, electrostatic, and stereochemical features, while simultaneously excluding non-cognate partners and off-target interactions [33]. For GAGs, establishing and deciphering specificity is uniquely challenging because GAG chains are inherently polydisperse, heterogeneous in sequence, and undergo spatial context- and developmental stage-dependent compositional remodeling. The apparent "promiscuity" of many GAGs arises from their high charge density and conformational plasticity, properties that permit multivalent, often degenerate interactions with diverse protein classes (cytokines, growth factors, viral glycoproteins, enzymes). Despite broad electrostatic attraction, distinct sulfation motifs, chain lengths, and higher-order presentations support protein-specific recognition [2]. This contrast is evident for the SARS-CoV-2 trimeric spike glycoprotein: basic patches in its S1 subunit favor highly sulfated HS motifs, whereas the post-fusion S2 core can also engage CS-E fragments [34]. Conversely, the fungal lectin from *Psathyrella velutina* discriminates strictly for HP over other acidic polysaccharides [35], and Protein C inhibitor displays subtype-dependent conformational switching when complexed with CS, DS, or HP [36].

Determinants of Binding Specificity

Sulfation patterns are the primary encoders of GAG specificity. Sequential actions of specialized sulfotransferases generate unique micro-domains such as 2-O-, 6-O-, and 3-O-sulfated motifs in HS that are stringently recognized by partner proteins [37, 38]. The 6-O-sulfotransferases 2 and 3 selectively modify 2-O-sulfated regions to create high-affinity epitopes for fibroblast growth factors, whereas the broader specificity 3-OST-5 targets both N-sulfated and N-unsubstituted GlcNAc residues [39].

Electrostatics underpins every stage of protein/GAG binding. Long-range attraction between the densely negative patches of sulfate and carboxylate and basic amino acid clusters accelerates complex formation, as reflected in the pronounced sensitivity of binding kinetics and affinities to changes in salt concentration [40, 41, 42]. Because counter-ions surround charged GAG groups and charged amino acid side chains, they partially shield the electrostatic interaction during protein/GAG binding, and subsequent short-range interactions refine the initial electrostatic attraction into a specific and sequence-dependent molecular interface [43, 44].

Extracellular sulfatases (e.g. Sulf-1 and Sulf-2) and lyases *remodel the sulfation landscape*, dynamically rewriting binding codes and therefore receptor specificity. For instance, Sulf-2 preferentially excises 6-O-sulfates from trisulfated HS domains, attenuating fibroblast growth factor signalling [45]. Chondroitin AC and B lyases, in turn, exert subtype-selective depolymerisation of CS-A/C and DS, respectively [46].

Beyond sequence, *GAG oligomer length* dictates avidity and topology of the binding interface. VAR2CSA, the placental malaria adhesin, requires CS chains of defined minimum length (around dp24 to dp32) to engage oncofetal CS in tumours [47]. Molecular dynamics and crystallographic studies reveal also that HS adopts extended and kinked conformers whose torsional preferences are modulated by iduronic acid epimerisation and protein docking, as seen for 6-OST-oligosaccharide complexes [48, 49].

The local density and orientation of GAG chains on core proteins (e.g. syndecans, glypicans) further modulate specificity by controlling multivalent clustering and cooperative binding [11, 17]. Recent super-resolution and cryo-electron tomography studies suggest that nanoscale glycocalyx architecture shapes growth-factor gradients, however the underlying rules are only partly defined [50, 51]. Advances in surface plasmon resonance, isothermal titration calorimetry, nuclear magnetic resonance, cryo-electron microscopy, and atomistic simulation now

permit quantitative dissection of affinity versus specificity, kinetics versus thermodynamics, and structural dynamics of protein/GAG complexes. Integration of these techniques is essential for a holistic understanding of protein/GAG interaction specificity.

Understanding the physicochemical characteristics that govern protein/GAG specificity has paved the way for developing GAG mimetics with tailored sulfation patterns, nucleic acid aptamers engineered to discriminate among GAG variants with low-nanomolar binding affinities [52], and self-assembling peptide scaffolds derived from RHAMM that recapitulate HA binding [53]. Synthetic HS analogs with strategically positioned sulfate groups show promise as anticoagulants and anticancer agents with reduced off-target effects [54]. Continued elucidation of protein/GAG specificity thus offers fertile ground for therapeutic innovation.

1.3 Experimental Approaches to Study Protein/GAG Complexes

Several experimental approaches are used to study GAGs and protein/GAG complexes, each limited by GAG flexibility, heterogeneity, and polydispersity. For resolving the structure of protein/GAG complexes, X-ray crystallography provides high-resolution structural information, though crystallizing such complexes can be challenging due to the flexibility and heterogeneity of GAGs. Cryogenic electron microscopy (cryo-EM) has recently emerged as a powerful technique for visualizing large protein/GAG complexes without the need for crystallization, particularly when the complex is too large for crystallographic experiments. Nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS) are the most prominent techniques for revealing atomic-level motion, composition, and sulfation patterns of GAGs, complementing the static structural information obtained by the previously mentioned methods. NMR is particularly useful for resolving the three-dimensional structure and conformational flexibility of GAGs, as well as for studying protein/GAG interactions in solution. MS is used for identifying the monosaccharide composition, sulfation patterns, and sequence of GAGs, especially using techniques like tandem MS and glycomics approaches. Techniques like surface plasmon resonance (SPR) and isothermal titration calorimetry (ITC) are employed for studying the interactions between GAGs and proteins. SPR allows for real-time, label-free measurements of binding kinetics, while ITC provides thermodynamic profiles of the interaction, including binding affinity, enthalpy, and entropy. Other approaches, such as fluorescence polarization, co-immunoprecipitation, and pull-down assays, are used to validate and explore protein/GAG interactions in a more biological context.

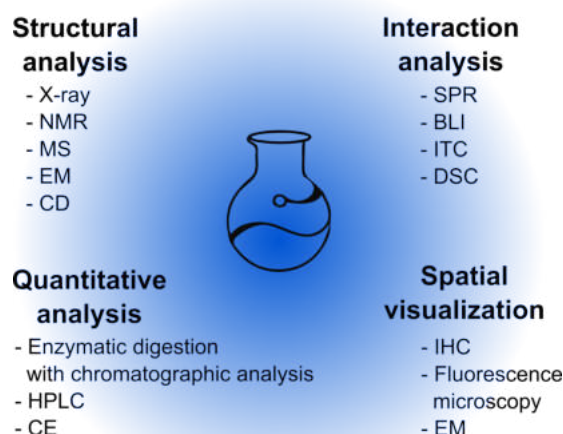


Figure 1-2: Experimental approaches to study GAGs and protein/GAG interactions; structural: X-ray crystallography (X-ray), nuclear magnetic resonance spectroscopy (NMR), mass spectrometry (MS), electron microscopy (EM), and circular dichroism (CD); interaction: surface plasmon resonance (SPR), biolayer interferometry (BLI), isothermal titration calorimetry (ITC), and differential scanning calorimetry (DSC); quantitative: high-performance liquid chromatography (HPLC), capillary electrophoresis (CE), and enzymatic digestion with chromatography; spatial: immunohistochemistry (IHC), fluorescence microscopy, and electron microscopy (EM).

1.3.1 X-ray Crystallography

X-ray crystallography operates by directing X-rays at a crystallized sample of GAGs or protein/GAG complexes to determine their atomic structure [55]. Upon hitting the crystal, the X-rays are diffracted by the electrons in the sample's atoms, producing a diffraction pattern, consisting of spots called reflections, which are captured and used to create an electron density map outlining the positions of atoms within the crystal. The crystal structure is then deduced by fitting atomic models to the electron density [55].

Crystallizing GAGs presents challenges due to their flexibility, heterogeneity, and polydispersity, making it difficult to form a stable, well-ordered crystal lattice. Nevertheless, crystallographic structures of protein/GAG complexes help to map out binding sites, identify critical interactions between specific protein residues and GAG functional groups, and understand the conformational changes that

occur upon binding. The resulting structural data are invaluable for drug design, particularly in identifying how therapeutic molecules might inhibit or enhance GAG-protein interactions. Yet despite its power, X-ray crystallography requires significant preparatory work, such as optimizing crystallization conditions. The crystallization of protein/GAG complexes results in static structures and may not fully capture the dynamic molecular interactions occurring under physiological conditions. In practice, crystallography is frequently paired with solution methods and was successfully employed to obtain detailed atomic-level structural information of specific protein/GAG complexes and identify corresponding binding sites and modes.

X-ray crystallography has been used to provide structural insight into the catalytic mechanisms of proteins interacting with GAGs, particularly in processes such as blood coagulation and fibrinolysis. The structures of thrombin and antithrombin have revealed key aspects of their catalytic mechanisms, including HP-induced conformational changes in antithrombin upon binding [56, 57, 58]. The binding of protease nexin-1 (PN1) and thrombin mediated by HP is accompanied by a substantial conformational shift, facilitating the formation of a highly effective inhibitory complex [59]. Additionally, the high-resolution structure of a thrombin-protein C inhibitor (PCI) complex, bridged by HP, has provided insights into how HP enhances thrombin recognition, thereby influencing PCI's dual role in coagulation [60]. Crystallographic studies have also characterized the calcium-dependent binding of annexin A2 by HP, revealing a new calcium-binding site that forms upon GAG interaction, crucial for annexin A2's role in fibrinolysis [61].

Structural studies on bacterial enzymes involved in GAG depolymerization have elucidated important features of these proteins. For instance, the structure of Heparin Lyase I from *Bacteroides thetaiotaomicron* highlighted an activity-enhancing thumb-like extension that aids in substrate recognition and binding [62]. Similarly, the crystal structure of dimeric Heparinase II from *Pedobacter heparinus* has demonstrated how its independent active sites contribute to substrate recognition in the depolymerization of HP and HS [63].

X-ray crystallography was also employed in the identification and characterization of distinct HP-binding sites in other proteins. The structure of the chemokine CXCL12 revealed distinct HP-binding sites, shedding light on its roles in embryonic development, HIV-1 infection, and cancer metastasis. This work proposed a mechanism in which GAGs sequester and present CXCL12 to its receptor, CXCR4 [64]. Similarly, the crystal structure of HP bound to NK1, a splice variant of hepatocyte growth factor/scatter factor (HGF/SF), identified primary and

secondary HP-binding sites in different domains, suggesting a dual regulatory role of HP in NK1 activity [65]. Additionally, crystallography helped identify two distinct GAG-binding sites in annexin V, which are essential for its high-affinity interactions with HS on cell surfaces [66].

The oligomerization of GAG-binding proteins has also been explored using X-ray crystallography. Liang et al. described the formation of rod-shaped, double-helical oligomers by the chemokine CCL5, and elucidated how GAGs bind through a specific motif, clarifying differences in the GAG-binding mechanisms between CCL5 and CCL3 [67]. Application of X-ray crystallography in combination with other approaches revealed the tetrameric assembly of chemokine CCL11 in complex with the HP derivative fondaparinux [68]. Studies of acidic (aFGF) and basic fibroblast growth factor (bFGF) complexes with their receptors and HP highlighted HP's role in facilitating ternary complex formation, with the 6-*O*-sulfation of HP shown to be critical for FGF signaling [69, 70]. Further crystallographic studies demonstrated the stabilization of protein dimers by fully sulfated HP deca-saccharides, including the dimeric forms of aFGF [71], amyloid precursor-like protein 1 (APLP1) [72], and the reelin-N domain of F-spondin [73].

With 102 experimentally available structures of GAGs and protein/GAG complexes [74], X-ray crystallography is as of today the most popular method for GAG structure determination despite its shortcomings in studies of GAGs. Research employing X-ray crystallography has shed light on molecular processes that have significant implications for disease and inflammatory responses, clarifying molecular steps in coagulation and fibrinolysis with direct relevance to therapeutic targeting.

1.3.2 Nuclear Magnetic Resonance

Nuclear magnetic resonance (NMR) spectroscopy is used to determine the structure, dynamics, and interactions of molecules at the atomic level. By measuring the magnetic properties of atomic nuclei in the presence of a strong magnetic field, NMR provides detailed information about the chemical environment and spatial arrangement of atoms within a molecule [75]. Several types of NMR experiments can be identified based on their level of complexity. One-dimensional (1D) NMR measures the resonance frequencies of nuclei in a single dimension and provides basic chemical shifts and coupling constants. Two-dimensional (2D) NMR experiments, which include COSY (Correlation Spectroscopy), HSQC (Heteronuclear Single Quantum Coherence), and NOESY (Nuclear Overhauser Effect Spectroscopy), measure correlations between two different types of NMR signals in two dimensions. More advanced NMR techniques, designed for studying large biomolecular systems, include three-dimensional (3D) NMR, in particular 3D NOESY-HSQC and 3D TOCSY-HSQC, which combine multiple 2D NMR experiments.

NMR has underpinned discovery and characterization efforts relevant to pharmacology. For instance, HP-like sulfated polysaccharides from snail mucus have been characterized using NMR and chromatography, revealing significant anti-SARS-CoV-2 activity by inhibiting spike protein binding to the ACE2 receptor, thus presenting an alternative HP source that bypasses the traditional porcine and bovine derived HP used in current pharmaceuticals [76]. A study on the slug *Limacus flavus* identified a new GAG with potent heparanase inhibitory activity. NMR analysis of this GAG highlighted the critical roles of chain length and sulfate substitution in determining its bioactivity [77]. A study on artificial ECM coatings used NMR spectroscopy to analyze the structural interactions between GAGs and scaffold materials, finding that GAG-coated scaffolds significantly enhanced bone regeneration and collagen synthesis in rats, highlighting the importance of sulfation degree and GAG type in bone healing [78].

NMR experiments were crucial in elucidating the interactions of GAGs with protein fibrils, e.g. by showing that GAGs bind tightly to amyloid- β ($A\beta$) fibrils, affecting their aggregation and stability [79]. Additionally, NMR characterization of HP tetrasaccharides indicated specific binding to $A\beta$, suggesting their potential as inhibitors of $A\beta$ aggregation [80]. NMR analysis of salmon calcitonin fibrillation revealed that GAGs, particularly HP, significantly accelerate this process at different pH levels [81]. In the context of atherosclerosis, NMR studies demonstrated that HP and epigallocatechin-3-gallate (EGCG) can remodel

apolipoprotein A-I fibrils, transforming them into non-toxic oligomers [82]. Additionally, NMR research on Parkinson's disease highlighted how GAGs influence α -synuclein fibrillation and cytotoxicity [83, 84].

Multiple studies have employed NMR spectroscopy together with other experimental and computational methods in the investigation of protein/GAG binding. Studies on IL-8 with different GAGs assessed the role of lysine residues in GAG binding through a combination of NMR spectroscopy, mutagenesis, and molecular dynamics simulations [85]. A similar approach was used to identify specific anchor points of chemokine CXCL12 for binding sulfated GAGs [86] and to discern GAG binding sites on CXCL14 specific to different GAG types [87]. The importance of particular residues in GAG binding was also demonstrated in a study using NMR spectroscopy to identify a GAG-binding site at the C-terminal end of IL-10, involving specific arginine and lysine residues [88].

To date, only eight NMR structures of GAGs can be found in the PDB [74] as NMR studies face challenges because of the polydispersity and structural heterogeneity of GAGs, causing weak signals and overlapping resonances due to similar repeating units. Despite these difficulties, NMR remains valuable for studying protein/GAG interactions in solution, offering insights beyond the static images provided by X-ray crystallography.

1.3.3 Electron Microscopy

Electron microscopy (EM) offers high-resolution molecular structure analysis due to its shorter electron wavelength compared to visible light, providing significantly higher resolution. A beam of electrons, emitted by a tungsten filament cathode and accelerated by an anode, is used to produce electric potential maps akin to X-ray electron density maps. The main types of EM are transmission electron microscopy (TEM), scanning electron microscopy (SEM), and cryo-EM.

EM has been valuable for structural studies for a variety of GAGs [89, 90]. A combination of cryo-EM and computational approaches enabled the detailed study of substrate selection, polymerization, and transmembrane translocation mechanisms of the glycosyltransferase hyaluronan synthase (HAS), participating in HA synthesis [91]. Using focused ion beam scanning EM, a ring mesh-like structure of GAG chains was shown to form around collagen fibrils in rat tendons, suggesting that multiple GAG chains interact to create a planar network [92]. The interactions of collagen and GAGs were further characterized by EM to reveal a delicate glycoconjugate-rich superficial layer and an underlying network of collagen fibrils bridged by GAGs, which together distribute stress across cartilage tissue [93].

1.3.4 Mass Spectrometry

Mass spectrometry (MS) characterizes GAGs and protein/GAG assemblies, typically via electrospray (ESI) or MALDI ionization [94], where the analyzed sample is ionized by applying high voltage to create an aerosol, or by using a pre-ionized reagent gas that reacts with the sample [94]. MS offers several key advantages over NMR spectroscopy and other experimental techniques, including superior sensitivity, faster processing speed, and the ability to monitor molecular exchanges, which allows for the detection of multimolecular complexes.

MS techniques, particularly ESI and matrix-assisted laser desorption ionization (MALDI), are powerful tools for studying host-guest complexes. A rapid and sensitive method for semi-qualitative and quantitative analyses of CS and DS was presented by Kiselova et al., requiring small sample sizes and providing accurate results in correlating structure to function in different tissues and species [95]. Liquid chromatography tandem mass spectrometry (LC-MS/MS) was employed to measure urinary DS and HS levels in patients with GlcNAc-phosphotransferase deficiency (mucopolidosis II and III), proving to be a sensitive tool for diagnosing the deficiency and potentially monitoring new therapies [96]. These studies underscore the critical role of MS techniques in the precise analysis and diagnosis of various GAG-related disorders.

Hydrogen/deuterium-exchange mass spectrometry (HDX-MS) further broadens MS by tracking the incorporation of deuterium into a protein of interest over time, thus enabling the mapping of interaction sites and conformational changes in solution-phase complexes. In the interleukin-8/CS complex, HDX-MS indicated structural stabilization in specific protein regions, supporting its utility in mapping protein/GAG interaction interfaces [97]. Molecular docking restrained by these data reproduced the binding interface previously characterized by NMR spectroscopy.

1.3.5 Capillary Electrophoresis

Capillary electrophoresis (CE) separates analytes by driving them through capillaries of submillimeter diameter under an applied electric field and is implemented in several complementary modes, including affinity CE (ACE), capillary zone electrophoresis (CZE), capillary electrochromatography (CEC), gel CE (CGE) and frontal analysis continuous CE (FACCE) [98].

In protein/GAG research, each mode has addressed a distinct analytical need. ACE pinpointed the HP-binding amino acid stretch of serum amyloid P component and showed that divalent metal ions modulate this interaction [99]. CZE quantified dissociation constants for HP binding to SAP-derived synthetic peptides [100, 101], whereas FACCE clarified the stoichiometry and affinity of the bovine serum albumin/HP complex [102]. Conventional CE assays have revealed marked differences in HP affinities among cleaved variants of β 2-microglobulin [103]. Together, these examples illustrate how the various CE protocols provide complementary thermodynamic, kinetic, and structural insights into protein/GAG interactions.

1.3.6 Circular Dichroism

Circular dichroism (CD) measures differential absorption of circularly polarized light to investigate their structural properties and conformational changes [104]. It distinguishes protein secondary structures in the far UV range and explores the local tertiary environment of aromatic amino acids in the near UV region. CD spectroscopy has also been used to study the structural features and conformational dynamics of GAGs, such as CS, HA, and HP, revealing their functional group characteristics [105, 106].

The interactions between biomolecules and GAGs have been analyzed using CD for a variety of proteins, e.g. factor Xa-antithrombin [107], and small molecules

like berenil and pentamidine [108, 109, 110]. CD spectroscopy has also been employed in characterizing conformational changes in proteins resulting from GAG binding [111].

1.3.7 Fluorescence Spectroscopy

Fluorescence spectroscopy is used to monitor the functional activity and binding dynamics of GAG-interacting proteins both in the presence and absence of GAGs, providing insights into reaction kinetics and inhibition by distinguishing the different absorption spectra of substrates and products. The addition of the substrate allows for monitoring changes in specific signals, making this assay more sensitive than spectrophotometric methods. However, it may be affected by impurities and the instability of fluorescent compounds under light exposure.

Fluorescence readouts quantify GAG effects on enzymatic activity and provide information about host/guest complex formation through changes in fluorescence intensity. The analysis of different sulfated GAGs using fluorescence spectroscopy showed that fluorescent HP, obtained through interactions of its uronic acid residues with 5-aminofluorescein, retained its anticoagulant activity [112]. A combination of fluorescence and NMR approaches, complemented by computational analyses, demonstrated that increased GAG sulfation, as well as specific sulfation patterns, increased the strength of protein binding [113].

1.3.8 Surface Plasmon Resonance

Surface plasmon resonance (SPR) tracks biomolecular interactions in real time by measuring changes in the refractive index near the surface of a sensor chip when biomolecules bind to immobilized ligands [114]. GAGs or GAG-binding proteins are typically immobilized on the sensor chip through methods like amine coupling or biotin-streptavidin interaction, and the binding events are detected via changes in the angle of reflected light.

SPR has advanced the understanding of protein/GAG interactions, revealing their roles in various biological processes and disease mechanisms. Studies have shown that CS and KS interact with phosphatidylcholine on plasma membranes [115] and that HP binds to low-density lipoprotein, impacting cholesterol accumulation and membrane dynamics [116]. Additionally, it was employed to show that CS-4 regulates cathepsin S activity through mixed-type inhibition [117], as well as the binding of sulfated GAGs to type I collagen at low pH, influencing collagen degradation [118]. Studies on interactions between GAGs and various viral pro-

teins, such as adeno-associated virus [119], SARS-CoV-2 [120, 121], monkeypox virus [122, 123], human respiratory syncytial virus [124], have utilized SPR to demonstrate varying binding affinities and structural dependencies that influence the potential therapeutic use of HP and its derivatives.

Research has further highlighted the specificity and implications of protein/GAG interactions, with structural elements in proteins like PDGF-AA and collagen IX significantly influencing their binding to GAGs [125, 126]. Protein tagging and mutation studies have underscored the importance of protein structure in these interactions [127, 128]. SPR analysis has quantified the high affinity of sclerostin for HP and revealed the effects of salt and metal ion concentrations on these interactions [129]. Additionally, studies on chemokine CXCL10 and prion proteins have explored how structural modifications affect GAG binding and cellular functions [130], while HA-mimicking glycopolymers have shown potential as tools for studying HA functions and developing new therapies [131].

1.3.9 Enzyme-Linked Immunosorbent Assay

The enzyme-linked immunosorbent assay (ELISA) is designed for high-throughput detection and analysis of GAGs and GAG-binding proteins, utilizing specific antibodies for recognizing GAG molecules or their complexes. Sandwich ELISA is commonly employed for GAG detection, in which two antibodies bind different parts of a GAG. Alternatively, competitive ELISA may be used, where labeled GAGs compete with sample GAGs for antibody binding. Both methods produce a measurable, enzyme-driven colorimetric signal indicating GAG concentration.

ELISA assays have been developed for use in clinical diagnostics to monitor diseases that alter GAG metabolism or degradation, such as cancer [132], osteoarthritis [133, 134], systemic sclerosis [135], and acute pancreatitis [136]. Additionally, studies involving ELISA experiments detail the role of GAGs in ECM organization in embryonic development [137, 138] and the modulation of GAG production by enzymes and drugs [139, 140].

ELISA provides specific and quantitative analysis of GAGs and can be adapted to high-throughput formats to handle multiple samples. It is less labor-intensive compared to traditional biochemical assays, however, the heterogeneity of GAG sequences and sulfation patterns makes it difficult to create universally effective antibodies.

1.3.10 Biolayer Interferometry

Biolayer interferometry (BLI) involves the immobilization of molecules on a biosensor, which is subsequently submerged in a solution with an analyte. Interference patterns caused by the binding of the immobilized molecule with the molecules of interest are measured, which provides the researcher with information on rate constants, reaction rates, or binding strength.

Key studies involving BLI include the discovery that HS proteoglycans are crucial for the intoxication process by the bacterial toxin CNF γ from *Yersinia pseudotuberculosis* [141], and detailed analysis of HS-derived oligosaccharides to understand protein/GAG interactions [142]. Additionally, investigations of the structure of the Amyloid Precursor Protein (APP) in the presence and absence of HP [143] and the binding efficiencies of different HP variants against histones demonstrated the diverse functional implications of GAGs in pathological conditions, such as vascular diseases related to elevated histone levels [144], and their potential in modulating chemokine signaling, as seen with the glycoprotein CXCL17 [145].

1.3.11 Calorimetry

Calorimetry quantifies heat changes associated with molecular interactions and chemical reactions, providing insights into the thermodynamics of biological processes [146]. In Isothermal Titration Calorimetry (ITC), heat change (i.e., consumption or release) is detected as the ligand molecule is titrated into a solution containing the receptor molecule, usually a protein, allowing for the determination of binding affinity, stoichiometry, enthalpy, and entropy using equation 1-1:

$$\Delta Q = n \times V \times \Delta H \times \left(\frac{[L]_t + [M]_t + K_d - \sqrt{([L]_t + [M]_t + K_d)^2 - 4[L]_t[M]_t}}{2[L]_t} \right), \quad (1-1)$$

where ΔQ is the heat released or absorbed, n is the stoichiometry of the interaction, V is the volume of the calorimeter cell containing the samples, ΔH is the enthalpy change, $[L]_t$ and $[M]_t$ are the total concentrations of the ligand and the receptor macromolecule, respectively, and K_d is the dissociation constant. This equation models the binding curve and helps determine the binding parameters of interest (K_d , ΔH , n) as well as the entropy change ΔS , using the relationships between enthalpy, entropy, and Gibbs free energy change ΔG , quantifying the overall change in energy during a process: $\Delta G = \Delta H - T\Delta S$ and $\Delta G = -RT \ln K_a$, where K_a is the association constant (inverse of K_d).

Differential Scanning Calorimetry (DSC) measures the difference in the amount of heat required to increase the temperature of a sample and reference as a function of temperature and is typically used to study thermal transitions such as protein denaturation and conformational changes upon heating. The methods are particularly useful in studying complex biological processes involving multiple binding partners, also because they do not require labeling or immobilization of the molecules studied. Moreover, ITC is able to provide a complete thermodynamic profile (ΔH , ΔS , ΔG , and K_d). However, these methods require a high sample concentration and the interpretation of results may be challenging when multiple binding sites and/or poses are involved in the binding process, which is particularly relevant for GAG-containing systems.

Both ITC and DSC have been used extensively in studies of GAGs and GAG-containing systems [147, 148, 149, 150, 151, 152]. ITC has been instrumental in research on protein/GAG complexes [153, 154, 155, 43] and novel GAG-based biomaterials [156, 157, 158, 159]. Calorimetric analyses were used to study the role of GAGs in fibrillogenesis and aggregation as a result of interactions with proteins [101, 160, 161]. ITC was successfully employed to characterize the GAG-binding

site on proteins [152] and assess the differences in binding affinity for different GAG types [162].

The principal use of DSC in the study of GAGs is the analysis of the thermal stability and characteristics of natural [163, 164] and synthetic GAG-containing biomaterials [165, 166, 167, 168, 169, 170, 171, 172]. DSC has also been employed to study the influence of GAGs on the structure of proteins [173, 174] as well as the interactions of GAGs and protein/GAG complexes with lipids [175, 176], providing comprehensive insights into the thermodynamics of protein/GAG interactions and advancing the development of new therapeutic strategies.

1.4 *In silico* Methods to Study Protein/GAG Complexes

The structural characterization of GAGs presents substantial challenges due to their inherent diversity, large molecular size, conformational flexibility, and negative charge. Experimental techniques can struggle to resolve GAG structures at high resolution. The flexibility, rapid conformational dynamics, and high negative charge of GAG chains impede crystal formation for X-ray crystallography, complicate the interpretation of NMR experiments, and reduce cryo-EM resolution. Computational methods therefore play a central role, but still need refinement to capture GAG physicochemistry and to predict affinities and selectivities.

Computational modeling is complicated by heterogeneity in GAG disaccharide compositions, sulfation, and epimerization. Sparse high-quality data limit force-field parameterization. Additionally, rapid conformational dynamics and flexible glycosidic linkages exceed the capabilities of standard MD simulations, necessitating computationally intensive enhanced sampling techniques. The high negative charge density of GAGs introduces strong electrostatic interactions and complex solvent and counterion effects, further complicating accurate modeling. GAG hydrophilicity requires explicit solvent models, significantly increasing computational demands, as implicit solvent approaches inadequately capture hydration effects.

Predicting protein/GAG binding affinities and selectivities is further complicated by non-canonical, multivalent, or allosteric interaction mechanisms and the interplay of hydrophobic, hydrogen-bonding, and electrostatic forces. Current force fields may inadequately represent sulfate groups or capture the behavior of large protein/GAG complexes with sufficient fidelity, impeding reliable energy calculations. Integrating experimental data (e.g., NMR or SAXS) with computational methods can refine predictions, yet the complexity in GAG structures often introduces uncertainties into hybrid models. Lastly, limited sequence-level data and small interaction datasets slow method development and validation.

1.4.1 Electrostatic Potential Calculations

Electrostatic Potential (EP, ϕ , eq. 1-2) is defined as the work of bringing a unit positive charge from infinity to a position in space and signifies how energetically costly or favorable it is for a unit charge to be located at a given position inside an electric field [177, 178]. Positive EP values reflect repulsion of a positive test charge (unfavorable, higher energy), and negative values reflect attraction (favorable, lower energy) [177, 178].

Mathematically, for a system of N discrete point charges, the electrostatic potential at a given point in space $\mathbf{r}$ is calculated by summing the potential from each charge using Coulomb's law:

$$\phi(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \sum_i \frac{q_i}{|\mathbf{r} - \mathbf{r}_i|}, \quad (1-2)$$

where ϕ_r is the electrostatic potential at position $\mathbf{r}$, q_i is the charge of the i -th particle located at $\mathbf{r}_i$, $|\mathbf{r} - \mathbf{r}_i|$ is the Euclidean distance between the observation point and that charge, and $4\pi\epsilon_0$ contains the vacuum permittivity ϵ_0 , which scales the strength of the interaction in SI units. The electric field $\vec{E}$ is the negative spatial gradient of the electrostatic potential: $\vec{E} = -\vec{\nabla}\phi$ [178].

EP mediates the interactions among charged molecular entities, including ions, amino acid side chains, DNA phosphates, and sulfated regions of GAGs. Mapping ϕ onto a protein surface highlights patches that will attract or repel other charged molecules, and helps predict binding sites, catalytic hotspots, or protonation equilibria [179, 180].

EP isosurface maps depict the spatial distribution of EP around biomolecules, using color gradients to mark regions of positive and negative charge (fig. **1-3**). These maps help pinpoint areas on the protein surface that are electrostatically favorable for ligand binding.

EP maps help locate binding hotspots, rationalize pK_a shifts, and guide drug or biomaterial design, i.e. tasks that would be prohibitively slow or impossible to probe experimentally at atomic detail [181, 182]. Three tiers of methods are routinely used: quantum mechanical (QM) calculations that solve the electronic structure explicitly and give the most accurate EP for small systems, semi-empirical and molecular mechanics approaches that assign pre-parameterized atomic charges to handle whole proteins efficiently, and continuum electrostatics models such as Poisson-Boltzmann or Generalized Born, which treat solvent and ions as a dielectric medium and are fast enough for scanning large biomolecular datasets.

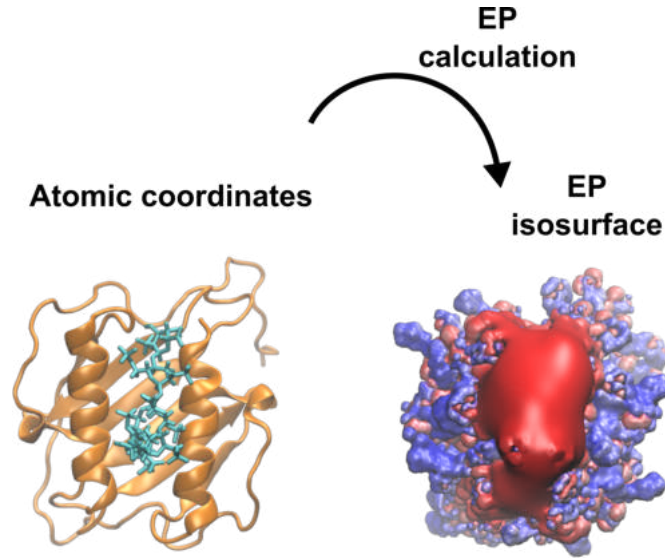


Figure **1-3**: IL-8 and HP in orange cartoon and cyan licorice representation, respectively, and the calculated EP isosurface obtained from APBS software, colored by EP isovalue, with red color indicating negative and blue indicating positive values.

QM methods calculate the electrostatic potential by explicitly solving how electrons are distributed around nuclei [183, 184]. *Ab initio* (wave-function) methods solve the time-independent Schrödinger equation for the entire many-electron wave function:

$$\hat{H} \Psi = E \Psi, \quad (1-3)$$

where $\hat{H}$ is the electronic Hamiltonian, Ψ is the many-electron wave function and E is its eigen-energy. Density-functional theory (DFT) reduces the many-electron problem to the electron density, and in its Kohn–Sham approach this is achieved by iteratively solving a self-consistent set of effective single-electron equations that reproduce that density:

$$\left[-\frac{1}{2} \nabla^2 + V_{\text{ext}}(\mathbf{r}) + V_{\text{H}}[\rho](\mathbf{r}) + V_{\text{XC}}[\rho](\mathbf{r}) \right] \phi_i(\mathbf{r}) = \varepsilon_i \phi_i(\mathbf{r}), \quad (1-4)$$

with the density recovered as $\rho(\mathbf{r}) = \sum_i |\phi_i(\mathbf{r})|^2$, where V_{ext} is the nuclear (external) potential, V_{H} the classical (Hartree) electron-electron term and V_{XC} the exchange-correlation potential that contains all many-body effects.

Because the cost of either approach rises steeply with system size, QM calculations

on protein/GAG complexes are normally confined to chemically critical fragments (catalytic residues, metal cofactors, or short GAG motifs). The resulting high-fidelity EP is then used to benchmark or parameterize faster semi-empirical or molecular-mechanics approaches [183, 184, 185, 186, 187].

While QM calculations give the most chemically detailed picture, their cost makes them impractical for whole protein/GAG assemblies. This gap is bridged by semi-empirical and classical molecular-mechanics (MM) methods, which approximate the electron cloud instead of solving for it exactly. Semi-empirical schemes keep a simplified quantum framework but plug in experimentally fitted parameters, trimming computer time enough to handle short GAG oligomers or active site snapshots with reasonable accuracy [188]. MM goes further by replacing electrons with fixed partial charges taken from a force fields such as AMBER or CHARMM, allowing entire proteins and long GAG chains to be simulated thousands of times faster [189, 190]. These faster approaches can therefore generate EP maps for full complexes, screen mutations, or follow binding events in MD trajectories [191, 192]. The trade-off is that subtle charge rearrangements, e.g. sulfate polarization, cation- π interactions, may be smoothed out, so MM results are often calibrated against the more precise (but smaller scale) QM or semi-empirical calculations.

Continuum electrostatics approaches treat the aqueous environment and dissolved ions that surround a biomolecule as smooth "backgrounds" instead of individual particles, making their mathematical foundation tractable for systems as large as an entire protein/GAG complex. The starting point is the Poisson equation, which links the electrostatic potential ϕ_r to the fixed charge distribution $\rho_e(r)$ [193]:

$$-\nabla \cdot (\epsilon(\mathbf{r}) \nabla \phi(\mathbf{r})) = 4\pi \rho_e(\mathbf{r}), \quad (1-5)$$

where $\epsilon(r)$ switches from a low value inside the solute (ca. 2–4) to the high dielectric constant of water (ca. 80). To account for mobile salt ions, crucial for the highly charged sulfates and carboxylates on GAGs, the equation gains a Boltzmann term, giving the Poisson-Boltzmann (PB) equation:

$$-\nabla \cdot (\epsilon \nabla \phi) = 4\pi \rho_e + \kappa^2 \sinh(\phi k_B T / e), \quad (1-6)$$

where κ (the inverse Debye length) encodes ionic strength. Finite-difference solvers such as APBS or the PBSA module in AMBER discretize this equation on a 3D grid. The resulting potential can be projected onto a molecular surface to highlight the positively charged amino acid patches that attract the negative GAG chain. Because PB still resolves a full grid, one calculation takes

seconds to minutes, which is fast enough for mutation scans or post-processing MD snapshots.

To maximize computational efficiency, the Generalized Born (GB) model replaces the grid with an analytical pairwise formula that mimics PB screening [191]:

$$E_{GB} = -\frac{1}{2} \sum_{i,j} \frac{q_i q_j}{f_{GB}(r_{ij}, \alpha_i, \alpha_j)}, \quad (1-7)$$

where f_{GB} uses Born radii (α_i) to capture how deeply each atom is buried. GB retains the key physics, i.e. reduced electrostatic coupling in a high dielectric solvent, yet scales almost like a simple Coulomb sum, letting full protein/GAG complexes run for microseconds in implicit solvent MD.

Electrostatic Potential Surface Calculations for GAGs

Electrostatic complementarity between clusters of basic residues on proteins and the highly sulfated, carboxylated GAG chains is a fundamental driver of specific molecular recognition. However, accurately modeling the EP of protein/GAG complexes presents significant challenges. These include the pronounced heterogeneity of GAG sulfation patterns, their intrinsic conformational flexibility, which often necessitates ensemble sampling, and the extreme negative charge density of GAGs, which strains the assumptions underlying continuum electrostatics and complicates dielectric boundary definitions.

Experimental techniques such as site-directed mutagenesis, NMR spectroscopy, and ITC shed light on electrostatics yet suffer, respectively, from structural perturbations, size limits, and bulk-only thermodynamics. Atomistic EP maps overcome these weaknesses, pinpointing charge hotspots and rationalizing GAG-binding affinities for targeted design.

Poisson–Boltzmann-based EP calculations using APBS have been applied successfully to several protein/GAG complexes. For example, mapping the EP of antithrombin bound to HP dp5 revealed a concentrated basic patch comprising lysine and arginine residues that dominated the favorable Coulombic interaction energy [194]. Mutagenesis of these residues to uncharged analogs results in dramatic reductions in binding affinity as measured by fluorescence and ITC [195]. In another example, EP surfaces computed for both monomeric and dimeric forms of the chemokine IL-8 were used to identify two distinct positive regions in the N-terminal loop and the C-terminal helix [196, 197, 85, 198]. These regions correspond closely to chemical shift perturbations observed by NMR during heparin titration, and targeted substitutions at these residues raise the dissociation con-

stant by 5 to 20-fold in SPR assays [196, 197, 85, 198].

Similarly, in the case of human serum albumin, continuum electrostatics and computational docking studies with HA and CS-6 highlight several shallow EP-positive grooves across subdomains IB, IIA, and IIIA [182]. These regions correspond to weak affinities validated through sedimentation and calorimetry, consistent with albumin’s role in low-affinity, multivalent GAG binding.

Together, these examples underscore the critical role of EP calculations in dissecting and predicting protein–GAG interactions, especially in systems where experimental resolution is limited or ensemble behavior is significant.

1.4.2 Molecular Docking

Molecular docking is a computational technique utilized to predict favorable binding conformations of ligands within protein targets, estimate the associated binding affinities, and identify promising binders through computational ranking. Its conceptual roots lie in classical "lock-and-key" and "induced-fit" models of molecular recognition, where favorable complementarity in shape, electrostatics, hydrogen bonding, and hydrophobic packing drives binding. Although proteins are intrinsically flexible and aqueous solvent imposes significant entropic and desolvation penalties, benchmarks show near-native poses are often recoverable with careful sampling and scoring. Modern docking approaches therefore balance physical realism with computational expediency, using transferable force field terms, empirical or knowledge-based scoring to generate testable hypotheses that guide biochemical experimentation and rational drug design.

A practical docking protocol begins with system preparation, where protein protonation states, missing atoms, and bound waters or cofactors are curated and a chemically consistent ligand geometry is generated. This is followed by search space definition, for example, a three-dimensional grid centred on a known or predicted binding pocket. Next, a sampling phase explores the ligand's rigid-body translations and rotations plus its internal torsions using deterministic searches, Monte Carlo or genetic algorithms to generate candidate poses. Each pose then enters the scoring stage, where empirical, knowledge-based, or physics-derived, functions approximate the binding free energy, often followed by consensus or machine learning rescoring to improve rank ordering. Finally, a post-processing and validation step refines the top poses via energy minimization, short MD simulations, or MM/GBSA calculations.

The core divergence among docking approaches lies in how they explore conformational space and account for flexibility through conformational sampling. At one end, rigid-body and soft docking treat both partners as static, relying on exhaustive or grid-based searches that are fast but miss induced-fit effects. Progressively richer schemes permit ligand torsions and side-chain rotamers, typically sampled with Monte Carlo or Genetic Algorithms (GAs) that stochastically hop across high-dimensional landscapes. Ensemble docking widens receptor flexibility by docking against pre-generated protein snapshots from MD or NMR, while induced-fit protocols iteratively alternate docking and local minimization to adapt pocket residues *in situ*. Deterministic, systematic searches guarantee coverage for small systems, while stochastic Monte Carlo and GA methods balance breadth and speed, fragment growth and incremental construction reduce

torsional burden for large ligands, and emerging reinforcement learning and diffusion-based generative models learn to propose low-energy poses directly. Each step up the flexibility ladder improves realism but inflates search space exponentially.

Once poses are generated, docking algorithms proceed with scoring and ranking. Traditional functions fall into three families: empirical (e.g., ChemScore, GlideScore), which regress weighted van der Waals, electrostatic, H-bond, and desolvation terms against experimental data; knowledge-based potentials, which transform atom-pair frequencies observed in crystallographic complexes into statistical potentials of mean force, and physics-based force field scores, which sum bonded and nonbonded energies from molecular mechanics parameter sets. Because each type of scoring function excels on different molecule types, modern docking workflows may combine them in a consensus approach, then apply post-docking refinement such as MM/GBSA, thermodynamic integration, or fast free energy perturbation to the top-scoring poses. Machine learning models employing graph neural networks [199, 200, 201, 202], as well as protein/ligand "language" models [203, 204], have enabled rapid rescoring and show promising results relative to conventional scoring approaches. Nonetheless, all scoring approaches face challenges due to inaccurate solvent treatment and entropy estimates.

Molecular Docking of GAGs

Protein/GAG complexes pose unique challenges in molecular docking studies stemming from the physicochemical characteristics of GAGs. The GAG backbone is very flexible as two glycosidic torsions plus ring puckering at every GAG residue create a plethora of low-energy shapes even for small or moderately sized GAG fragments. Their repetitive sequence means a larger GAG molecule can bind in several similar or identical sequence fragments, multiplying pose permutations even when keeping the GAG backbone conformation fixed. Lastly, the scarcity of high-resolution protein/GAG complex structural data hinders both parameterization and validation of scoring functions.

Dynamic Molecular Docking (DMD) addresses these issues by combining docking and steered MD (SMD), guiding a GAG molecule toward known binding sites. Although computationally expensive, DMD yields insights when prior binding site knowledge is available [205].

Fragment-based docking takes a different approach, decomposing GAG polysac-

charides into short conformationally pre-screened oligosaccharides that can be sampled exhaustively. Each fragment is first guided to basic patches on the protein surface by electrostatic potential maps, then subjected to a rapid local relaxation so that glycosidic torsions and ring puckers adapt without creating a search space explosion. Low-resolution experimental data are integrated early in the docking protocol to discard spurious placements. Once potential poses are obtained, spatially compatible neighbors are merged together into progressively longer chains whose linkages satisfy distance, dihedral torsions, and sulfation pattern criteria. Subsequent MD relaxation optimizes local conformations and supplies a physics-based pose ranking. Benchmark studies confirmed that fragment-based docking can reliably provide accurate atomic models of protein/GAG complexes [206].

Replica Exchange MD with Repulsive Scaling (RS-REMD) docking, in contrast, directly handles conformational plasticity of GAGs by running parallel MD replicas in which the van der Waals radii of all atoms are uniformly scaled [207, 208, 209]. Enlarged radii in high-radius replicas sterically push both the GAG and nearby protein side chains out of local minima. Replica exchanges move configurations that have escaped local traps in the enlarged-radius replicas down to replicas with unmodified van der Waals radii, producing binding poses that account for the adaptability of both ligand and protein.

For highly charged flexible ligands, RS-REMD goes beyond the static electrostatic placement of fragment-based approaches by explicitly sampling long-range charge-driven steering together with the accompanying protein side chain rearrangements [210]. The sampling performance of RS-REMD derives from the scaling factor rather than an exhaustive enumeration of rotamers, rendering the method largely insensitive to GAG length. With no prior assumptions about the binding site or GAG conformation, RS-REMD is unbiased with respect to the starting structure of the complex. The use of explicit solvent preserves the water-mediated interactions crucial to protein/GAG binding while still allowing the docking protocol to run efficiently. Its principal drawback is cost, as hundreds of nanoseconds of MD per replica across multiple replicas are typically required, yet benchmarks on multiple protein/GAG complexes have shown RS-REMD to recover native-like poses in the majority of cases where classical docking fails [211, 212].

Fragment-based assembly and RS-REMD stand as two representative ways that occupy opposite ends of a broader spectrum from rapid modular assembly to exhaustive enhanced sampling. Their usefulness depends on the size of the GAG ligand, available experimental restraints, and the desired balance between speed

and accuracy. Because protein/GAG complexes often demand customized docking workflows, several computational studies have demonstrated their ability to translate theoretical predictions into experimentally confirmed binding insights. However, current studies indicate that while fragment-based methods are reliable for shorter HP fragments, RS-REMD, though costlier, can converge on native-like poses for chains longer than dp10. Gandhi et al. combined surface conservation analysis with fragment-based docking of HP dp6 onto bone morphogenetic protein 2, pinpointing two histidine-rich patches whose mutagenesis abolished binding, thereby validating the electrostatic "hot strips" predicted *in silico* [213]. Sapay et al. demonstrated a hybrid workflow in which trisaccharide fragments of HS were first placed with AutoDock4, annealed to explore backbone registers, and then fed into explicit-solvent MD, the refined poses reproduced NMR CSPs for FGF2, showcasing the power of post-docking dynamics [214].

A systematic assessment was provided by Uciechowska-Kaczmarzyk et al. who benchmarked fourteen docking methods against twenty-eight protein/GAG complexes and found that even the top performer (AutoDock3) achieved less than or equal to 3 Å RMSD in just 43 % of cases, underscoring the need for GAG-specific scoring and sampling improvements [215].

1.4.3 Molecular Dynamics

Molecular dynamics (MD) is a computational simulation technique for investigating the physical motion of atoms and molecules. It relies on the principles of Newtonian mechanics to model the behavior of molecular systems by solving the equations of motion that govern interacting particles. This enables the tracking of how a molecular system evolves throughout a simulation. MD simulations are an essential tool for understanding the fundamental processes at the atomic level as they can reveal important information about molecular conformations, interactions, and reactions that are often difficult to capture through experimental techniques alone.

The development of MD traces back to the mid-20th century when the advent of digital computers enabled the practical implementation of computational simulations. The pioneering work by scientists like Berni Alder and Thomas Wainwright in the late 1950s laid the groundwork by using MD to study the properties of simple liquids [216]. Advances in computational power and algorithms, along with the development of increasingly sophisticated force fields, further propelled the progress of MD simulations. Force fields, in this context, are mathematical frameworks consisting of parameterized equations that define the potential energy landscape of molecular systems based on atomic positions, accounting for both bonded (e.g., bond stretching, angle bending, and dihedral torsions) and nonbonded (e.g., van der Waals, electrostatic) interactions. Continuous improvements such as parallel computing and enhanced sampling techniques have expanded the scope and scale of MD simulations, enabling the study of complex biological systems and materials with detail and accuracy.

MD simulations are based on Newtonian mechanics, which relate atomic motion to the forces acting on each particle used to compute time-dependent trajectories [217] (eq. 1-8):

$$a_i = \frac{d^2 \mathbf{r}_i}{dt^2} = \frac{1}{m_i} \mathbf{F}_i, \quad (1-8)$$

where a_i is the acceleration of particle i with mass m_i and position $\mathbf{r}_i$ resulting from the force $\mathbf{F}_i$ acting on it. By knowing the forces acting on each particle, one can compute the acceleration, and subsequently the velocity and position of each particle. The iterative process of computing the forces acting on, and the displacement of simulated particles, repeated over many small time steps, produces a trajectory of the particles over time showing the molecular system's dynamic behavior.

In most force fields the nonbonded terms are cast in explicit analytic forms: van

der Waals interactions can be approximated with a 12-6 Lennard-Jones potential, while electrostatic forces obey Coulomb's law. These nonbonded terms, together with the bonded contributions (bonds, angles, torsional angles) are summed to yield the total potential energy. An accurate representation of the structural, thermodynamic, and dynamic properties of simulated molecules requires parametrizing the potential energy functions, which involves fitting the force field to experimental and/or quantum mechanical data.

In MD simulation tools, the equations of motion are solved using numerical integration algorithms. Among the most widely used algorithms are the Verlet, velocity Verlet, and leapfrog Verlet methods [218]. The Verlet algorithm calculates new positions of the simulated particles based on their current and previous positions, offering excellent energy conservation properties without providing velocities directly [217]. The leapfrog algorithm addresses the lack of direct velocity calculation by computing velocities at half-time steps, "leaping" over positions, which allows for efficient updates of both positions and velocities [219]. The velocity Verlet algorithm combines the strengths of both, updating positions, velocities, and accelerations in a single step, while maintaining the energy conservation benefits of the original Verlet algorithm [220, 221].

Atomistic (all-atom, AA) and coarse-grained (CG) MD simulations are two fundamental approaches used to study the behavior of molecular systems. AA MD provides a detailed, high-resolution view by modeling each atom explicitly, making it ideal for examining specific molecular interactions and conformational changes. In contrast, CG MD simplifies the system by grouping atoms into larger units or "beads," reducing computational complexity and enabling the simulation of larger systems and longer timescales. Together, these approaches offer complementary insights, with AA MD capturing fine details and CG MD efficiently exploring large-scale dynamics, thus broadening the scope of molecular simulations.

Atomistic Approaches

In atomistic MD, molecules are represented as collections of atoms, each with distinct positions and properties: mass, charge, radius, and bonding characteristics [222]. Each atom's movement is calculated iteratively, taking into account the forces exerted by neighboring atoms, i.e., those located within a defined interaction distance cutoff. This level of detail enables the observation of intricate aspects of MD, such as hydrogen bonding patterns, van der Waals interactions, and polarization effects, providing valuable insights into the structural and functional properties of molecular systems. Running atomistic

MD simulations involve using specialized software to perform the complex calculations. Popular MD software includes GROMACS (GRONingen Machine for Chemical Simulations) [223], AMBER (Assisted Model Building with Energy Refinement) [189], CHARMM (Chemistry at HARvard Molecular Mechanics) [224], NAMD (Nanoscale Molecular Dynamics) [225], LAMMPS (Large-scale Atomic/Molecular Massively Parallel Simulator) [226], and TINKER-HP [227], each characterized by features that make them applicable to specific biomolecule types and/or computational resources available. These tools provide a comprehensive suite of functionalities, including energy minimization, equilibration, and production runs, allowing detailed and accurate simulations.

Setting up an AA MD simulation involves selecting or constructing the molecular structure of interest, often sourced from databases like the Protein Data Bank (PDB) or modeled *de novo* when experimental data are unavailable. Once the structure is obtained, it may need refinement to correct stereochemistry, optimize the geometry, and adjust protonation states to ensure the structure accurately reflects physiological conditions.

Periodic boundary conditions (PBC) are employed in MD simulations to approximate bulk-phase behavior by simulating a small, representative segment (simulation box) of the system [228]. The simulation box is virtually replicated in all directions, allowing particles that exit one face to re-enter from the opposite side. This eliminates artificial edge effects and maintains a continuous, uniform environment throughout the system. The simulated molecule is placed in a simulation box that should be large enough to prevent interactions between the molecule and its periodic images, i.e., the virtual copies of the system generated by PBC, and solvated to mimic the natural aqueous environment of biological molecules. PBC are particularly crucial when simulations include explicit solvent, where solvent molecules are represented individually to capture solvation dynamics with high accuracy. Without PBC, solvent at the box edges would experience unrealistic conditions, such as vacuum exposure, leading to artifacts that distort both structural and thermodynamic properties [228].

Several explicit water models are used to realistically mimic the aqueous environment of biomolecular processes, each differing in complexity, accuracy, and computational efficiency. Among the most commonly employed are TIP3P (Transferable Intermolecular Potential with 3 Points) [229] and SPC (Simple Point Charge) [230], both of which represent water as rigid three-site models with fixed bond angles and partial charges. TIP4P and its variants, such as TIP4P-Ew, TIP4P/2005, and TIP4P-D, introduce a fourth interaction site to better approximate the location of the negative charge and improve the repre-

sensation of properties like density, diffusion, and dielectric constant. SPC/E (Extended SPC) adds polarizability corrections to enhance energetic accuracy, while TIP5P includes two additional virtual sites to capture the lone pairs on the oxygen atom, offering better agreement with the experimental density and radial distribution functions of water. More advanced models, such as POL3 and SWM4-NDP, incorporate polarizability to more accurately represent the electronic response of water molecules to local electrostatic fields, although at a higher computational cost. The choice of water model depends on the simulation goals and required balance between accuracy and performance, with TIP3P and TIP4P/2005 among the most commonly used in biomolecular MD due to their compatibility with major force fields and favorable trade-offs between realism and efficiency. To maintain overall electrostatic neutrality of the system, counterions (e.g., Na^+ , Cl^-) are added after solvation.

Prior to the production run, in which simulation data are collected for analysis, the system undergoes energy minimization to eliminate steric clashes and high-energy configurations. This is followed by equilibration steps, during which temperature and pressure are gradually brought to target values, ensuring the system reaches a stable, physiologically relevant state before data acquisition begins.

Proper temperature and pressure control are essential for the stability and accuracy of MD simulations, influencing the reliability of the simulated molecular behaviors. During the MD simulation, the thermal conditions of the simulated system are controlled by thermostats, which regulate the kinetic energy of particles to achieve and sustain a target temperature. The two most common thermostats are the Berendsen and Langevin thermostats. The Berendsen thermostat scales the velocities of particles towards the target temperature using a first-order differential equation, providing smooth temperature control, however sometimes resulting in non-physical temperature distributions [231]. The Langevin thermostat, on the other hand, adds a frictional force and a random force to the system's particles, simulating the effect of a heat bath, offering a more accurate temperature control and the maintaining of the correct thermodynamic ensemble [232].

Pressure control in MD simulations is managed by barostats. The Berendsen barostat adjusts the system's volume by scaling the coordinates of particles, driving the pressure towards a target value [233]. However, similar to the Berendsen thermostat, it can lead to non-physical volume fluctuations. The Langevin barostat incorporates stochastic forces and friction, akin to the Langevin thermostat, ensuring more realistic pressure regulation [234].

The choice of time step in MD simulations is critical for ensuring accurate and stable integration of the equations of motion. Typically, a time step of 1 to 2 femtoseconds (fs) is used in atomistic MD to capture the fast vibrational motions of bonds, especially hydrogen bonds. Using time steps larger than this can lead to integration errors and unrealistic behavior. However, longer time steps can be employed if constraints like SHAKE [235] or LINCS [236] are applied to bond lengths involving hydrogen atoms, which exhibit the fastest vibrational motions, allowing the simulation to run faster without losing accuracy. The total simulation time depends on the scientific question being addressed and the available computational resources. For instance, studies of protein folding or large conformational changes may require simulations extending to microseconds or longer. Given the computational intensity, these extended simulations often necessitate access to high-performance computing clusters or specialized hardware, like GPUs, to be feasible within reasonable time frames.

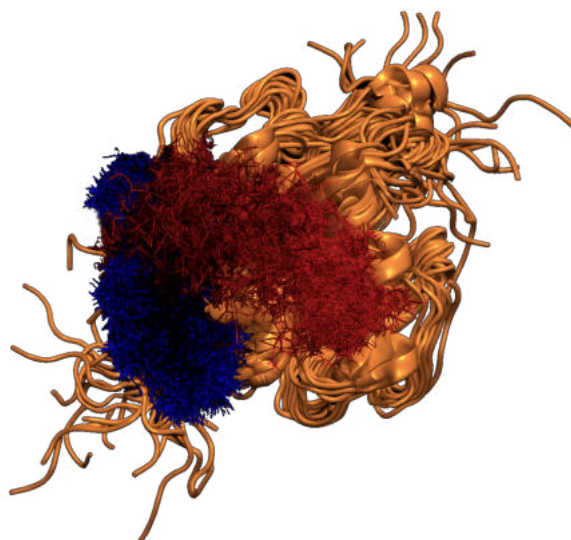


Figure 1-4: Snapshots of an MD trajectory; the protein, IL-8, is shown in cartoon representation, while the ligand (HP dp6) is shown in sticks representation colored according to time step, with red color corresponding to the initial position of the ligand in the MD trajectory and blue corresponding to the final position.

Post-simulation, detailed analysis of trajectories and derived properties is conducted (fig. 1-4). Trajectory visualization is one of the first steps in this process, where software like VMD (Visual Molecular Dynamics) [237], PyMOL [238], or Chimera [239] is used to visualize the resulting trajectories of atoms over time,

helping to identify significant conformational changes, interactions, and overall system behavior. Beyond visualization, quantitative analyses are performed to extract detailed information about the system, typically include root mean square deviation (RMSD) to measure the average deviation of atomic positions from a reference structure over time, and root mean square fluctuation (RMSF) to assess the flexibility of individual residues or atoms, highlighting regions with high mobility, which might be functionally important or prone to interactions. Radial distribution functions (RDF) describe how particle density varies as a function of distance from a reference particle and are particularly useful for understanding the spatial distribution and interactions of atoms or molecules in the system, such as solvation shells around ions or the organization of lipid bilayers. Other important properties often analyzed include hydrogen-bond analysis to track the formation and stability of hydrogen bonds, and secondary structure analysis, to monitor changes in protein folding or unfolding.

AA MD Simulations of GAGs

Fully atomic structures of GAGs and protein/GAG complexes can be obtained from established databases of experimentally determined structural data, e.g., the PDB or carbohydrate-focused databases such as Glyco3D [240] or GAG-DB [74]. If experimental structures are not available, specialized tools can be used to build and refine three-dimensional GAG models, including the GAGBuilder of GLYCAM-Web [241], CHARMM-GUI Glycan Modeler [242], PRODRG [243], LigParGen [244], or the tleap module of the AmberTools suite [245]. For protein/GAG complexes, initial models can be generated through molecular docking techniques, followed by refinement to resolve any steric clashes or unrealistic conformations.

Force fields tailored for carbohydrate simulations enable precise modeling of GAGs, reflecting the unique physicochemical characteristics and sulfation variations among GAG classes. GLYCAM (Glycoprotein and Carbohydrate Molecular Mechanics) is one of the most frequently used force fields for various carbohydrate types, offering specific parameters for monosaccharides and linkages found in GAGs [246]. Another option, the CHARMM (Chemistry at HARvard Macromolecular Mechanics) force field, is comprehensive and versatile, supporting a broad spectrum of biological molecules, including carbohydrates, and is suitable for GAG simulations [190]. These force fields are integrated into MD software such as AMBER [189] to streamline the simulation process.

The parametrization process for force fields like GLYCAM and CHARMM involves deriving parameters that accurately describe the potential energy surfaces of molecules. This process typically involves quantum mechanical (QM) calculations on model compounds or molecular fragments to determine the optimal bond lengths, angles, dihedral torsions, and nonbonded interactions [247, 248]. These QM results are then used to fit the parameters of the force field to reproduce experimental data, including crystal structures and NMR spectroscopy results, ensuring accurate representation of energetics and geometries, particularly the conformational properties of glycosidic linkages [247].

Simulating protein/GAG complexes requires the integration of force fields that can accurately represent both proteins and GAGs within the same MD simulation. Commonly, a combination of the GLYCAM force field for GAGs with AMBER or CHARMM force fields for proteins is employed, which leverages the strengths of each force field: GLYCAM's detailed parametrization for carbohydrates and AMBER or CHARMM's comprehensive parameters for proteins, including peptide backbones, side chains, and their interactions. A comprehensive study evaluating the effectiveness of the CHARMM, GROMOS, and AMBER/GLYCAM force fields in modeling protein/carbohydrate interactions highlighted significant divergences in their predictions of structural descriptors and unbinding free energies [249]. Although certain structural descriptors like intermolecular hydrogen bonding were generally consistent across the tested force fields, differences were observed in specific amino acid/carbohydrate contact patterns. While GROMOS predicted a higher frequency of contacts, CHARMM and GLYCAM reported fewer but more intense interactions, with CHARMM showing the best correlation to experimental data albeit underestimating unbinding energies, and GLYCAM overestimating them. Additionally, particular attention has to be paid to CH- π interactions in protein/carbohydrate binding, which are often not adequately represented in force fields [249, 250].

CG Approaches

Advances in computational power, efficient sampling techniques, and accurate force field parameterization have significantly broadened the scope of AA molecular simulations [251, 252], allowing efficient sampling of conformational space [253, 254] and enabling explicit simulations of solvated molecular systems, thus benefiting biochemical research. However, processes involving large biomolecular assemblies or long-timescale conformational transitions often remain beyond the practical reach of AA simulations. Coarse-grained (CG) approaches address this limitation by grouping atoms into larger units, known as pseudoatoms or "beads," thereby reducing the degrees of freedom, enhancing computational efficiency, and facilitating the study of longer timescales and larger biomolecular systems, including protein folding [255, 256], conformational transitions [257, 258, 259, 260], and macromolecular interactions. Although CG MD sacrifices some atomic-level detail, it captures essential system behaviors, providing valuable insights into the mechanisms and functions of molecular systems while maintaining computational tractability. Developing accurate and transferable force fields for CG models remains challenging, given the need to represent complex molecular interactions with fewer parameters, often leading to biases toward reference configurations [261, 262].

Early CG models were developed to simplify the representation of proteins and nucleic acids, enabling studies of large-scale dynamics and interactions. The level of CG is determined by the definition of pseudoatoms, with common approaches involving one pseudoatom per residue or two per residue to account for side-chain packing and specificity [263].

Elastic network models (ENMs) represent systems as beads connected by elastic springs, which have been useful for analyzing protein behavior, characterizing principal modes of motion, fitting atomic structures to electron density maps [264], and refining low-resolution structural data [265, 266]. Applications include studies of diphtheria toxin [267] and HIV-1 protease [268]. The Gō model, originally developed for protein folding simulations, represents proteins as chains of one-bead amino acids with interactions biased toward the native structure, effectively modeling folding thermodynamics and kinetics [262, 269, 270]. To address the limitations of the Gō model in describing intermediate metastable states, additional energy terms have been introduced. These modifications have improved its ability to model complex folding landscapes [271], enhanced the representation of residue-level specificity [272, 273], and enabled more realistic simulations of biomolecular interactions such as protein-protein binding [274]. ENM and Gō

models are inherently biased toward reference configurations, which limits their transferability compared to early CG models developed by Levitt that employed more transferable, knowledge-based parameterizations [275].

Modern one-bead models are largely derived from Gō models, incorporating advanced potential energy terms but still depending partially on reference configurations. Introducing a second bead at the side chain centroid improves local interaction specificity [256], while further refinements include adding energy terms [276] and fitting free energy functions derived from AA simulations [277, 278].

More complex CG models combine multiple residues into larger pseudoatoms, such as the MARTINI force field, which maps four heavy atoms to a single pseudoatom [279], or employ shape-based CG models to preserve the molecular shape [280]. These models have been employed in simulations of viral capsid dynamics, protein-induced membrane bending [281], and large liposome systems [282]. In four-bead models, the side chain is represented by a single bead while the backbone's three heavy atoms are modeled explicitly, enabling detailed descriptions of hydrogen bonding, which are useful for applications such as *de novo* protein structure prediction [283] and protein folding studies [284].

Coarser CG descriptions can be achieved using ENMs with rigid-block decomposition [285, 286], Brownian dynamics for electrostatics-driven diffusional interactions [287, 288], and continuum representations via density fields [289]. These approaches offer simplified yet insightful methods to study protein-protein docking [290, 291] and the competition between protein folding and aggregation [289]. The interactions between CG pseudoatoms are described by the effective CG energy function, which can be derived from AA MD simulations, experimental data, or theoretical models and statistical analyses of structural databases. In practice, many CG potentials combine these sources to balance accuracy, transferability, and computational efficiency. CG models based on AA MD aim to reproduce configurational probabilities and thermodynamic properties of atomistic systems, employing methods such as inverse Monte Carlo [292] and force-matching [293]. Knowledge-based CG potentials derived from experimental structural data are based on assumptions like the Boltzmann distribution of the number of particular residue pairs interacting in native structures [263]. Initially developed for protein folding [255], these potentials are also used to investigate protein binding mechanisms, with recent developments enhancing their accuracy and applicability [294].

Most rigorous CG MD simulations depend on specific molecular systems and environmental conditions, which limits their transferability and often results in overly collapsed states compared to experimental observations. Recent efforts to

improve parameterization have included the use of a CG polarizable water model [295] and the renormalization of CG multibody interactions [296].

Although no single CG model is universally optimal, the combined use of different CG models alongside other computational techniques can provide a more comprehensive understanding of experimental results and help discriminate between competing hypotheses regarding molecular mechanisms, structural conformations, or interaction pathways.

CG MD Simulations of GAGs

Despite their biological importance and widespread presence, only a limited number of CG models have been developed for GAGs, likely due to the complexity of their sequences and structures [297]. The initial CG model for GAGs was introduced by Bathe et al. in 2005 [298], specifically for CS and HA. This model represented each sugar residue with five CG beads: two for the carbon atoms and one for the oxygen atom to simulate the glycosidic linkage, an interaction site for the center of mass to model steric interactions, and an interaction site for the center of charge to account for electrostatic interactions. This model accurately reproduced the conformations and titration characteristics of the studied GAGs [298]. Sattelle et al. [299, 300] created a CG model for HS, simplifying the sugar residue representation to two interaction centers: one for the sugar ring and another for the glycosidic linkage oxygen atom. This model was effectively used for studying large HS molecules [300] and HS-containing proteoglycans [299].

Samsonov et al. proposed a more intricate model encompassing 28 pseudoatoms to simulate 17 different GAG types [301]. These pseudoatoms corresponded to various functional groups of GAGs, such as the oxygens in glycosidic linkages, sulfate groups, carboxylate and N-acetyl groups, the centers of mass of pyranose rings, and the CH₂OH group of GlcNAc and GalNAc. Depending on the type and sulfation pattern of the sugar residue (GalNAc/GlcNAc and GlcA/IdoA) and its position in the GAG chain (internal/terminal), each repeating unit was represented by 2 to 5 CG pseudoatoms. Geometrical parameters for virtual bonds, bond angles, and torsional angles were derived from all-atom MD simulations of various GAGs, while nonbonded interaction parameters were obtained using steered MD. Pseudoatoms representing carboxyl and sulfate groups were given a charge of -1, with all other pseudoatoms being neutral. This CG model effectively simulated the global and local properties of GAG chains of varying lengths.

Another CG model of GAGs is derived from the SUGRES-1P model for carbohydrates, part of the UNified COarse gRaiNed (UNICORN) model for biomacro-

molecules [302, 303]. The UNICORN model employs a physics-based approach, reducing each repeating unit to one or two CG sites depending on the macromolecule type and the system's effective energy function [304]. This model ensures transferability across different systems by decomposing the system's potential of mean force into contributions from individual CG sites, pairs of CG sites, and groups of CG sites. The UNICORN model has successfully predicted protein structures [305], folding kinetics [306], conformational changes [307], and RNA and DNA structure and dynamics [308]. The implementation of the SUGRES-1P model for HP and protein/HP interactions was performed as part of this thesis and is further detailed in Sections 3.3 and 3.4.

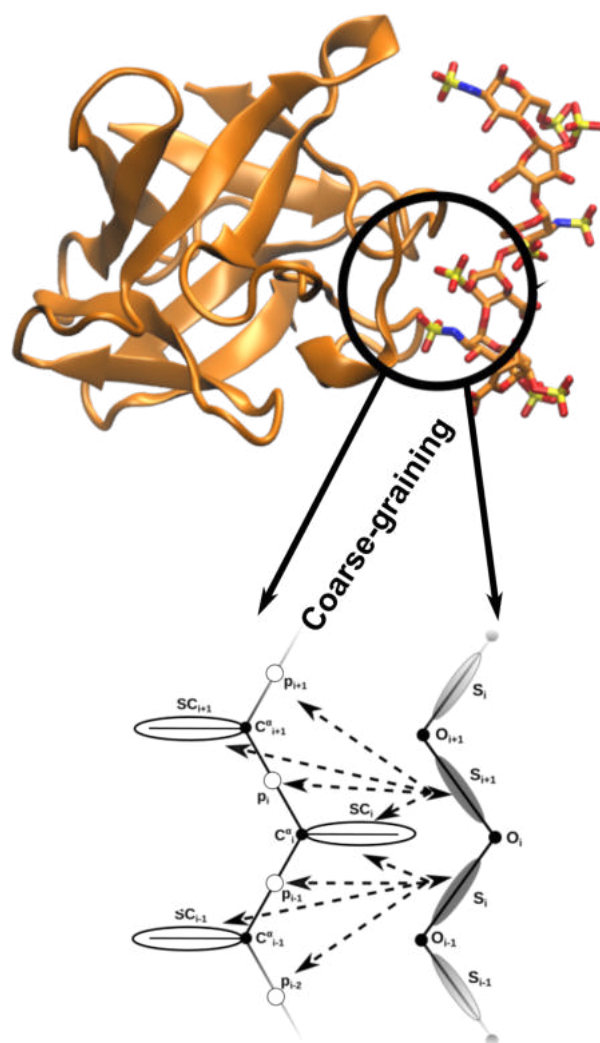


Figure **1-5**: A protein and ligand complex shown in orange cartoon representation (protein) and sticks representation colored according to atom name (ligand). Below are the CG representations of fragments of the same complex, with unified CG beads corresponding to groups of atoms and dashed arrow lines indicating interactions between the CG interaction sites of the protein and ligand.

1.4.4 Binding Free Energy Calculations

Many biological processes rely on thermodynamic properties of the receptor/ligand binding, rendering accurate binding affinity calculations essential in computational biophysics and drug discovery. Understanding binding strength and the role of individual residues in this process enhances insights into biological mechanisms. The intricate nature of binding involves factors such as direct receptor/ligand interactions, desolvation, energetic strain from adopting specific binding conformations, and changes in configurational entropy. Binding free energy calculations complement docking and MD simulations by providing insights into binding energetics, which is useful in identifying optimal binding poses of ligands (Table 1-2). Per-residue decomposition of the binding free energy highlights residues exerting the strongest favorable or unfavorable contributions, thereby guiding targeted mutagenesis by mapping these residues onto most populated MD conformations or high-resolution crystal/NMR structures. Free energy is a thermodynamic quantity representing the amount of work a system can perform, a quantity crucial in binding studies as it quantifies how tightly a ligand binds to its receptor. Binding free energy is a specific case of Gibbs free energy (G), a general concept used to predict the feasibility of any thermodynamic process at constant pressure and temperature. The change in its value (ΔG) is used to determine if binding occurs spontaneously, depending on the enthalpy change (ΔH), the entropy change (ΔS) and the temperature T :

$$\Delta G = \Delta H - T\Delta S, \quad (1-9)$$

A negative ΔG indicates favorable binding, moving the system toward equilibrium, i.e., a state of balanced forward and reverse binding rates. Understanding these changes helps predict the affinity and stability of molecular interactions, directly impacting drug design.

Several strategies exist for calculating binding free energy, each with different levels of accuracy and computational requirements. Among these, fast end-point techniques are computationally efficient methods that estimate binding free energy between two molecules by focusing typically only on the bound and unbound states. These methods, including Molecular Mechanics/Poisson-Boltzmann Surface Area (MM/PBSA), Molecular Mechanics/Generalized Born Surface Area (MM/GBSA), Linear Interaction Energy (LIE), end-point Free Energy Perturbation (FEP), and Quantum Mechanics/Molecular Mechanics (QM/MM) are frequently used for virtual screening due to their speed and simplicity. Despite their efficiency, fast end-point techniques may not capture

all free energy contributions, particularly for flexible or complex systems, and some depend heavily on empirical parameters. Potential of Mean Force (PMF) methods, on the other hand, allow a more detailed exploration of the binding landscape by calculating free energy as a function of a distance or along pathway between interacting molecules.

More rigorous approaches like thermodynamic integration (TI) and alchemical FEP require extensive sampling through MD simulations, offering greater accuracy through statistically rigorous post-processing. The high computational costs arise from the explicit solvent treatment and the need to determine free energy differences at multiple intermediate states. TI calculates free energy differences by integrating the gradient of the free energy along a reaction coordinate, while alchemical FEP estimates changes by gradually transforming one molecule into another, allowing precise energy calculations.

Table 1-2: Comparison of Methods for Estimating Binding Free Energies in Computational Chemistry

Method	Description	Advantages	Disadvantages
Alchemical FEP	Calculates free energy difference by simulating a gradual transformation between two states	High accuracy for small perturbations; handles complex systems; can account for entropy changes	Computationally expensive; requires significant sampling; convergence issues in large transformations
TI	Integrates the derivative of the free energy with respect to a coupling parameter over a series of simulations	High precision for moderate perturbations; straightforward error analysis	Computationally expensive; requires careful placement of intermediate coupling-parameter states; sensitive to sampling

Method	Description	Advantages	Disadvantages
MM/PBSA or MM/GBSA	Combine molecular mechanics energies with continuum-solvent free energies from either a Poisson–Boltzmann (PB) or Generalized Born (GB) implicit-solvent model	Fast post-processing of MD; no alchemical transformations; GB variant even quicker	Accuracy depends on force field and implicit model; solvent treated implicitly; limited entropy treatment; GB typically less accurate than explicit solvent or the PB variant
LIE	Estimates binding free energy using a linear approximation of van der Waals and electrostatic interaction energies between ligand and solvent	Simple and computationally efficient; requires fewer simulations	Empirical parameters needed; less accurate for complex systems; limited to specific systems
QM/MM	Combines quantum mechanics for the active site or ligand with molecular mechanics for the rest of the system	Accurate for systems with electronic effects; captures polarization and charge transfer	Computationally intensive; requires careful QM region selection

Compared to TI and alchemical FEP, MM/PBSA, MM/GBSA, and LIE offer a practical balance of accuracy and computational efficiency, making them particularly suitable for studying large, flexible molecules like GAGs, where extensive sampling is often impractical. Their ability to approximate binding energetics without requiring multiple intermediate states provides valuable insights into GAG interactions while remaining computationally feasible.

MM/PBSA is used for estimating the binding free energy of biomolecular complexes by combining molecular mechanics (MM) calculations with continuum solvation models. It estimates free energy by decomposing the system into MM energy, solvation energy, and entropic contributions. The MM energy includes bonded interactions (bond stretching, angle bending, dihedral torsions) and nonbonded interactions (van der Waals and electrostatic interactions). Solvation

energy is split into polar contributions, calculated using the Poisson-Boltzmann equation (PBE, eq. 1-10), and nonpolar contributions, estimated using the solvent-accessible surface area (SASA).

$$-\nabla \cdot (\epsilon(\mathbf{r})\nabla\phi(\mathbf{r})) = \rho_f(\mathbf{r}) + \rho_m(\mathbf{r}), \quad (1-10)$$

where $\epsilon(\mathbf{r})$ is the dielectric permittivity at position $\mathbf{r}$, $\phi(\mathbf{r})$ is the electrostatic potential at position $\mathbf{r}$, $\rho_f(\mathbf{r})$ is the fixed charge density at position $\mathbf{r}$, and $\rho_m(\mathbf{r})$ is the mobile charge density at position $\mathbf{r}$. Entropy is typically estimated through normal mode analysis (NM) or quasiharmonic approximation (QH), though it is often omitted due to computational cost.

The MM/PBSA workflow involves an MD simulation on the complex to generate an ensemble of configurations, followed by computing the total free energy of the complex, ligand, and receptor from snapshots extracted from the MD trajectory. Binding free energy ($\Delta G_{binding}$) is then calculated as:

$$\Delta G_{binding} = \langle G_{complex} \rangle - \langle G_{ligand} \rangle - \langle G_{receptor} \rangle, \quad (1-11)$$

where $\langle G \rangle$ represents the average over MD snapshots.

MM/PBSA is computationally efficient and relatively accessible for application as a post-processing method to MD simulations using common software tools like AMBER or GROMACS, providing flexibility for analyzing different parts of the trajectory. However, the use of an implicit solvent model may fail to accurately capture specific solvent interactions like hydrogen bonding and ion effects compared to explicit solvent models. As entropy estimation is computationally expensive and therefore often approximated or omitted, this additionally reduces accuracy. Since MM/PBSA evaluates binding energies based on discrete MD snapshots, its results are inherently limited by the conformations sampled in the trajectory.

The MM/GBSA method is another computational technique used for estimating binding free energies of biomolecular complexes as a sum of molecular mechanics energy (bonded and nonbonded interactions), solvation energy, and an entropic contribution. It is similar in concept to MM/PBSA but uses the Generalized Born (GB) model (eq. 1-12) instead of the PBE to calculate the polar solvation energy, allowing for faster computation while retaining a reasonable level of accuracy.

$$\Delta G_{solv}^{polar} = -\frac{1}{2} \sum_{i,j} \frac{q_i q_j}{f_{GB}(r_{ij})}, \quad (1-12)$$

where q_i and q_j are the charges of atoms i and j , r_{ij} is the distance between atoms i and j , $f_{GB}(r_{ij})$ is the Generalized Born function, which depends on the effective Born radii and the solvent.

The GB model estimates the electrostatic solvation energy that would be obtained by solving the PBE, doing so under the assumptions of a near-spherical solute and a continuum dielectric solvent. It is faster than the full PB equation, making it suitable for simulations requiring many energy evaluations. MM/GBSA is faster and more computationally efficient than MM/PBSA due to the analytical GB model, making it suitable for large datasets, but may not capture electrostatic interactions as well as PB, especially in highly charged systems.

An alternative computational approach for estimating binding free energies is LIE, which is based on a linear approximation of the interaction energies between the ligand and its environment (protein and solvent), derived from MD, or Monte Carlo (MC) simulations. The method uses empirical scaling factors to establish a quantitative relationship between interaction energies and binding free energy, adjusting these factors to fit experimental data for more accurate predictions.

The LIE method estimates $\Delta G_{binding}$ by using a linear combination of van der Waals and electrostatic interaction energies for the ligand in both the bound and free states. To achieve this, independent MD simulations are conducted for the receptor/ligand complex and for the ligand in solution, followed by the computation of interaction energies. The binding free energy is then determined using the LIE equation (eq. 1-13) and empirical scaling factors, resulting in a linear approximation that aligns with experimental observations:

$$\Delta G_{binding} = \alpha \langle E_{vdW}^{bound} \rangle + \beta \langle E_{elec}^{bound} \rangle - (\alpha \langle E_{vdW}^{free} \rangle + \beta \langle E_{elec}^{free} \rangle) + \gamma, \quad (1-13)$$

where $\langle E_{vdW}^{bound} \rangle$ and $\langle E_{elec}^{bound} \rangle$ are the average van der Waals and electrostatic interaction energies between the ligand and its environment in the bound state, $\langle E_{vdW}^{free} \rangle$ and $\langle E_{elec}^{free} \rangle$ are the corresponding interaction energies in the solvated free state, α and β are empirical scaling factors for van der Waals and electrostatic interactions, respectively, and γ is an empirical constant to account for other contributions (e.g., entropy, desolvation effects).

Moving beyond empirical methods, PMF offers a more detailed exploration of the energy profile along a reaction coordinate, making it particularly effective for studying binding pathways, understanding energy barriers, and identifying intermediate states in receptor/ligand interactions. PMF calculations often use techniques like Umbrella Sampling (US). US applies restraining potentials

to maintain the system at predefined positions along the reaction coordinate, enhancing sampling in low-probability areas and generating the PMF curve. SMD, on the other hand, uses external forces to pull a ligand along the reaction coordinate, sampling different configurations to produce a free energy profile. Both methods use biasing techniques to efficiently sample the reaction coordinate and provide a comprehensive energy landscape. PMF provides a detailed free energy landscape along a chosen pathway, useful for understanding complex binding mechanisms, but it is computationally intensive and requires well-defined reaction coordinates.

Application of Free Energy Calculation to GAGs

Computational approaches for estimating binding free energies in GAGs require judicious method selection, given the substantial flexibility, dense charge distribution, and extensive solvation effects inherent to these molecules. Alchemical FEP and TI are among the most accurate methods, although they are also computationally demanding. A notable example demonstrated that FEP could predict the binding affinities of HP to AT-III with an accuracy within ± 2.3 kcal/mol in comparison to experimental values, emphasizing that explicitly accounting for solvation dynamics is critical in capturing the strong electrostatic and hydration components of protein/GAG interactions [309]. TI has been employed to refine CG models such as the Martini force field by calculating partition coefficients of GAG disaccharides between water and octanol phases, thereby improving parameterization of GAG simulations and enabling more realistic replication of GAG behavior in biological contexts [310].

Despite the high precision achievable with alchemical methods, computational efficiency often necessitates alternative strategies such as MM/PBSA and MM/GBSA. These methods approximate binding free energies by summing intramolecular energy terms (computed *in vacuo*) with a solvation term derived from implicit-solvent models. Although MM/PBSA and MM/GBSA have provided satisfactory rankings of GAG-binding affinities, they can lack the accuracy seen in fully explicit methods, particularly for highly charged systems [44, 311]. One study combined MM/PBSA with MD simulations to analyze the energetics of HP binding to PECAM-1 and annexin A2, revealing that longer HP fragments bind more strongly due to enhanced electrostatic interactions in specific protein domains, but also showing that binding free energies can be slightly overestimated relative to experimental data [44]. MM/GBSA has also been applied

to protein/GAG complexes, illustrating that the flexibility of the iduronic acid ring, especially in its 2-O-sulfated form, significantly affects computed binding free energies, underscoring the need to incorporate ring conformational variability into GAG force fields [311]. Another study using MM/GBSA found that non-enzymatic protein/GAG systems typically exhibit stronger electrostatic contributions to binding than enzymatic complexes, which rely more heavily on van der Waals interactions [179]. Further analyses focused on APRIL, a protein that interacts with various GAGs, used MM/GBSA to elucidate the roles of key N-terminal lysine residues, GAG chain length, and sulfation patterns in modulating binding affinities. These studies consistently identified electrostatic forces as dominant, particularly in the case of HP, and showed that full-length APRIL binds more effectively than its truncated forms [312].

For research questions demanding more detailed pathways or mechanistic insights, techniques such as LIE, SMD, and US can be used. LIE is relatively inexpensive computationally, although its accuracy depends heavily on calibration against extensive experimental datasets, which can then enable robust scoring of protein/GAG interactions [313, 314]. Methods that determine the potential of mean force, whether by US or the Jarzyński equation in SMD simulations, can yield both binding free energy and kinetic parameters [313, 314, 315, 316]. However, these approaches often require substantial simulation time to achieve convergence, making them more challenging for routine quantitative prediction. Hybrid quantum mechanics/molecular mechanics (QM/MM) calculations, or fully QM approaches, have also been proposed for further improving the treatment of electrostatic interactions, but their application to GAGs is relatively limited due to high computational cost and complexity [317, 318].

2 Research Aims and Key Objectives

Protein/GAG interactions fall into three categories: nonspecific (the binding is determined only by the GAG net charge), broadly specific (where a protein binds a limited range of GAG types), or highly specific (requiring an exact GAG sequence). The underlying factors shaping these interactions remain not fully understood and are often attributed to the *sulfation code* (GAG sulfation pattern and degree) completely determining the physicochemical properties of GAGs.

The primary aim of my research was to investigate the factors affecting protein/GAG interactions through computational chemistry techniques, alongside developing novel tools and analysis pipelines for analyzing these systems. The employed theoretical models were validated and refined using experimental data from NMR spectroscopy and glycan microarray experiments. The presented work is structured into four distinct sections, each focusing on the use of different computational tools to study GAGs and protein/GAG complexes.

1. Complementarity of electrostatic potential isosurfaces as a determinant of protein/GAG binding specificity:
 - *Characterization of the near-surface electrostatic potential of proteins and protein/GAG complexes;*
 - *Validation of theoretical electrostatic potential surface calculations using a novel NMR technique;*
 - *Development of an analysis pipeline integrating experimental and computational data for electrostatic potential predictions.*
2. Significance of electrostatic interactions in protein/GAG binding specificity:
 - *Comparison of the binding behavior of HP and de novo designed acidic peptides;*

- *Investigation of the structural, dynamic, and energetic features distinguishing HP from anionic peptides using computational and complementary experimental data.*

3. Development and optimization of GAG-specific computational tools:

- *Implementation of a coarse-grained model of HP for MD simulations;*
- *Optimization of the corresponding parameters, including interaction energy term weights;*
- *Critical assessment of the method against experimental data.*

4. Investigation of the influence of GAG sulfation pattern on binding specificity:

- *Analysis of physicochemical properties of unbound 3-O-sulfated HS;*
- *Proposal of MD-based descriptors to predict specificity and affinity of HS binding to different proteins.*

3 Publications Included in the PhD Thesis

The following section provides brief summaries of the publications included in this PhD thesis, with each subsection focusing on a distinct study. The full publications can be found at the end of this section.

3.1 Detecting Protein-Ligand Interactions with Nitroxide Based Paramagnetic Cosolutes

NMR spectroscopy provides valuable insights into protein/ligand binding affinities and conformational changes [319]. Paramagnetic NMR broadens this scope, enabling sensitive screening of fragment binding events, even when affinities are very low, even for weakly binding fragments, providing structural restraints to aid computational chemistry especially when crystallography is challenging [320]. Solvent distance-dependent paramagnetic relaxation enhancement (sPRE) effects, induced by soluble paramagnetic probes, map solvent-accessible surfaces without altering receptor or ligand molecules [320] and can be used to characterize solvent-accessible surfaces [321, 322], predict structures [323, 324], detect transient conformations [325], and identify binding interfaces [326, 327, 328].

A recently developed method using PROXYL cosolutes allows determination of the protein near-surface electrostatic potential (ϕ_{ENS}), with NMR data correlating well with Poisson-Boltzmann theory predictions [329, 330, 331]. Ionic PROXYL probes are useful for detecting interactions with charged ligands like GAGs, which significantly alter the electrostatic potential at their binding site. The presented study used the PROXYL method to examine IL-8 binding to GAGs and the Grb2 SH2 domain interaction with phosphotyrosine peptides, mapping ligand binding sites and changes in ϕ_{ENS} . Experimental NMR data aligned with theoretical models from atomistic MD simulations and ϕ_{ENS} predictions via the Adaptive Poisson-Boltzmann Solver (APBS) software [193, 332]. The comparison provided deeper insights into how ligand binding alters protein electrostatics, emphasizing the importance of considering both protein and ligand flexibility in interaction predictions. Free IL-8 was characterized by a strong positive ϕ_{ENS} , particularly in GAG-binding regions, which aligned well with computational predictions, especially in the structured regions of the protein. This agreement validated the theoretical model and highlighted the importance of IL-8's positive charge in GAG interactions.

Upon binding to IL-8, HP caused widespread electrostatic potential alterations that extended beyond the binding interface. In contrast, HA left the electrostatic profile of the protein largely unaltered, owing to its lower affinity. HP adopted an extended conformation between the C-terminal helix and N-terminal loop of IL-8, consistent with previous theoretical and experimental studies, while HA bound energetically less favorably in a groove between the C-terminal helices of the IL-8 dimer. The results imply that the impact of HP on electrostatic

potential may be crucial for IL-8's role in forming chemokine gradients during inflammation, whereas HA likely plays a more transient or context-dependent role, contributing to less stable or secondary interactions.

The study demonstrated that combining computational and experimental methods improves understanding of protein/ligand interactions by providing accurate predictions and validation of ϕ_{ENS} and ligand binding poses. While agreement between experimental and computational ϕ_{ENS} values for IL-8 and Grb2 SH2 supported the models, discrepancies revealed areas needing further refinement or where additional factors, such as protein flexibility or solvent effects, should be considered.

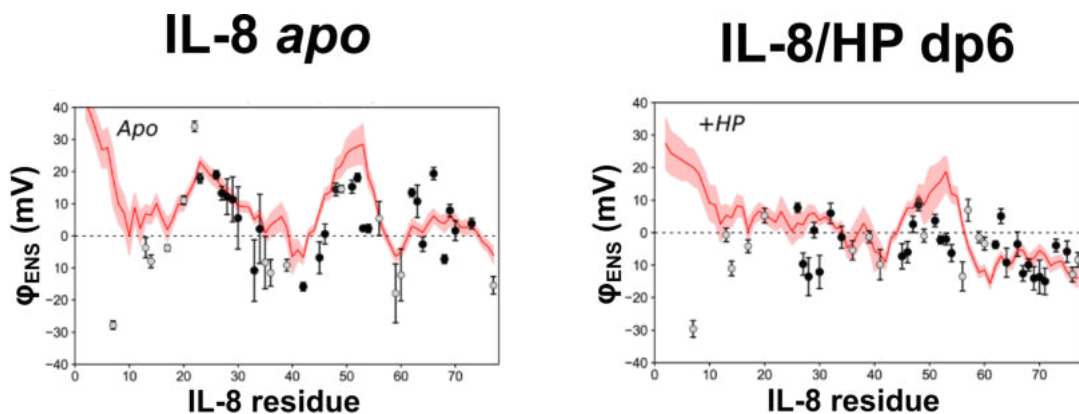


Figure **3-1**: ENS electrostatic potentials (ϕ_{ENS}) determined from the PRE NMR data of IL-8 *apo* (top) and the IL-8/HP complex (bottom). Experimentally determined ϕ_{ENS} (black points for secondary structure residues and gray points for loop regions) and predicted ϕ_{ENS} data (red line) are plotted along the IL-8 residue number. The predicted ϕ_{ENS} values represent the mean of the ENS potentials calculated from a 100 ns MD simulation, the red shaded regions indicate the standard deviation of the ϕ_{ENS} predictions. Figure adapted with permission of the publisher from Penk A, et al. *Detecting Protein-Ligand Interactions with Nitroxide Based Paramagnetic Cosolutes*. Chemistry. 2024;30(18):e202303570. Author rights retained.

3.2 Ligand Binding of Interleukin-8: a Comparison of Glycosaminoglycans and Acidic Peptides

As previously observed, the alteration of the near-surface electrostatic potential of IL-8 upon GAG binding was greater for the more negatively charged HP than HA. The ϕ_{ENS} changes extended beyond the HP binding site, underscoring the role of electrostatic interactions in IL-8/HP binding and suggesting that electrostatic complementarity may influence binding specificity and affinity.

The nature of protein/GAG binding spans a continuum of mechanisms in terms of specificity. On one hand, certain complexes show highly specific recognition, as with AT-III binding a defined HP pentasaccharide [333], driven by precise sulfation patterns and 3D conformation. On the other hand, electrostatically dominated interactions, exemplified by chemokines (e.g., CXCL12 [86], CXCL14 [87]) and proteins like BMP-2 [334] or TGF- β 1 [335], depend heavily on net negative charge but also reflect sulfation positions, chain length, and protein surface topology. Even these less specific cases often exhibit sulfation or conformational preferences, highlighting the multifactorial nature of GAG recognition. Many protein/GAG interactions lie somewhere between these extremes, resisting simple classification as purely specific or electrostatic. Although NMR, X-ray crystallography, and computational methods have enriched our understanding, many molecular details of protein/GAG binding remain to be clarified.

This study aimed to evaluate the role of electrostatics in IL-8/HP binding and contrast it with the influence of structural properties of HP. Using NMR and molecular modeling, IL-8 interactions with HP dp6 and a series of synthetic acidic decapeptides of comparable length and charge, designated as p3-, p5-, p7-, and p10-, labeled according to their incremental negative charges, were analyzed to better understand how electrostatics and structural differences between GAGs and peptides affect binding specificity. The distinct chemical nature and conformational preferences of GAGs and peptides, including differences in the number of degrees of freedom, charge distribution, and solvent-mediated interactions, likely lead to differences in their recognition descriptors and impact on the specificity of their interactions with proteins.

Molecular docking and atomistic MD simulations of IL-8/ligand complexes identified probable binding sites and key binding residues. Subsequent free energy calculations allowed to determine binding energy contributions. The frequency

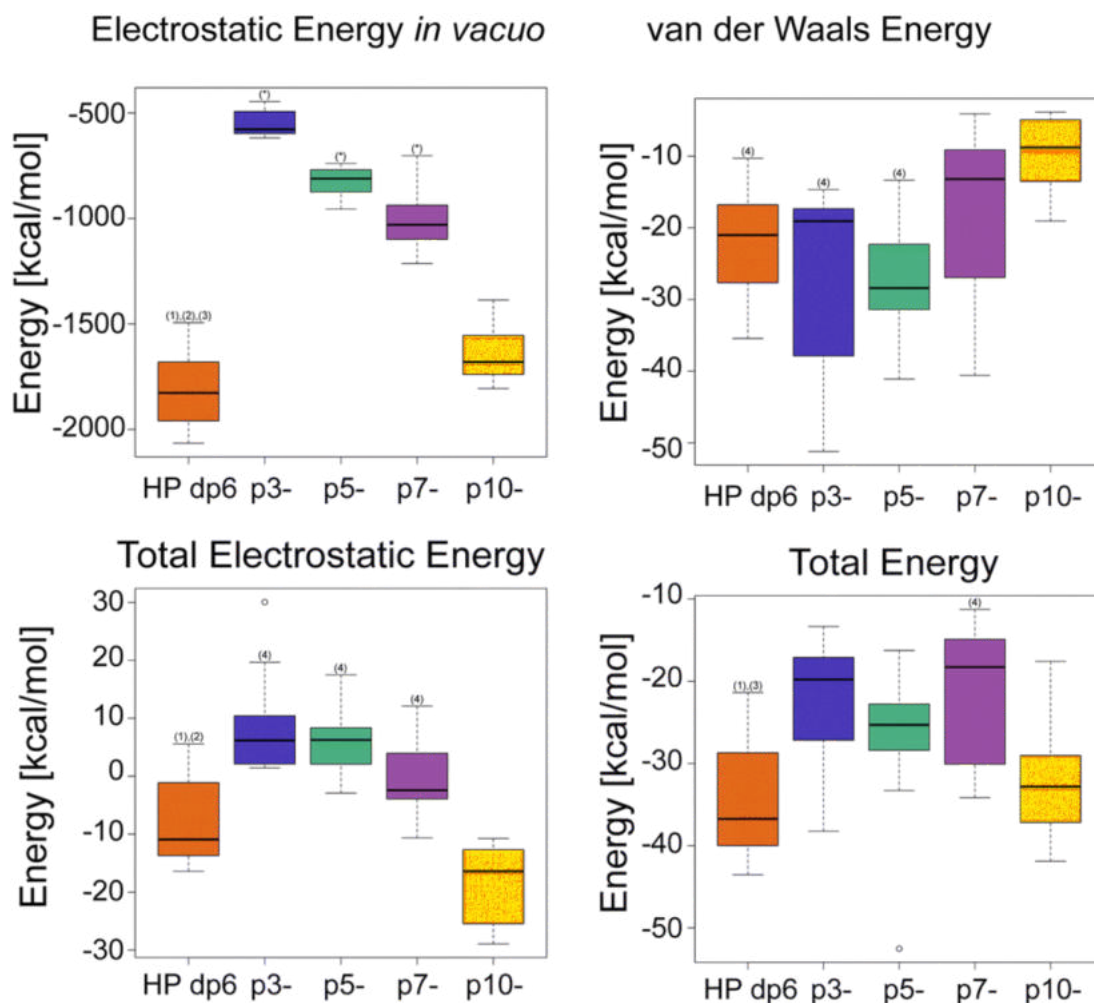


Figure 3-2: Electrostatic energy *in vacuo* (top left), total electrostatic energy (bottom left), van der Waals energy (top right) and total free energy (bottom right) for the IL-8/ligand complexes determined using MM/GBSA. Statistical significance of differences is shown as (*) for statistical significance when compared to all other groups, (1) statistical significance when compared to p3-, (2) statistical significance when compared to p5-, (3) statistical significance when compared to p7-, (4) statistical significance when compared to p10-. Figure adapted with permission of the publisher from Schulze C, et al. *Ligand binding of interleukin-8: a comparison of glycosaminoglycans and acidic peptides*. Phys Chem Chem Phys. 2023;25(36):24930-24947. Author rights retained.

and stability of H-bonds between IL-8 and ligands were also analyzed. Additionally, long MD simulations of unbound ligands were conducted to assess their flexibility compared to IL-8-bound ligands using radius of gyration (Rg) and RMSD, indirectly measuring entropy loss upon binding. Independent of the computational analyses, NMR spectroscopy quantified HP and peptide binding affinities and mapped ligand binding sites via chemical shift perturbations (CSPs).

The extensive computational analysis confirmed the importance of electrostatics in binding, though structural and energetic differences between HP and peptides suggested additional factors at play. Docking simulations showed both HP and peptides bind parallel to the α -helices of IL-8, however HP exhibited a more geometrically defined binding mode, while peptides were characterized by more dispersed poses, indicating reduced specificity. In the MD simulations, peptides formed less stable hydrogen bonds, leading to weaker binding affinity compared to HP, which exhibited stronger binding and more favorable total binding free energy. While electrostatic contributions were similar for HP and the most charged peptide p10-, peptides had weaker overall binding due to less favorable van der Waals and solvation energies. HP retained its conformation upon IL-8 binding, showing minimal change in radius of gyration and suggesting it was already in the optimal binding state. Peptides, particularly those of lower charge, instead displayed substantial conformational rearrangements with associated greater entropic loss.

The study revealed that while electrostatics is crucial in IL-8/ligand interactions, it alone cannot account for GAGs' higher affinity for IL-8 over acidic peptides. Other factors, such as ligand flexibility, conformational changes, hydrogen bonding, solvent effects, and the specific arrangement of charged groups, form a complex network of parameters driving protein/GAG specificity. These findings enhance the understanding of protein/GAG interactions and offer insights for designing more effective GAG mimetics.

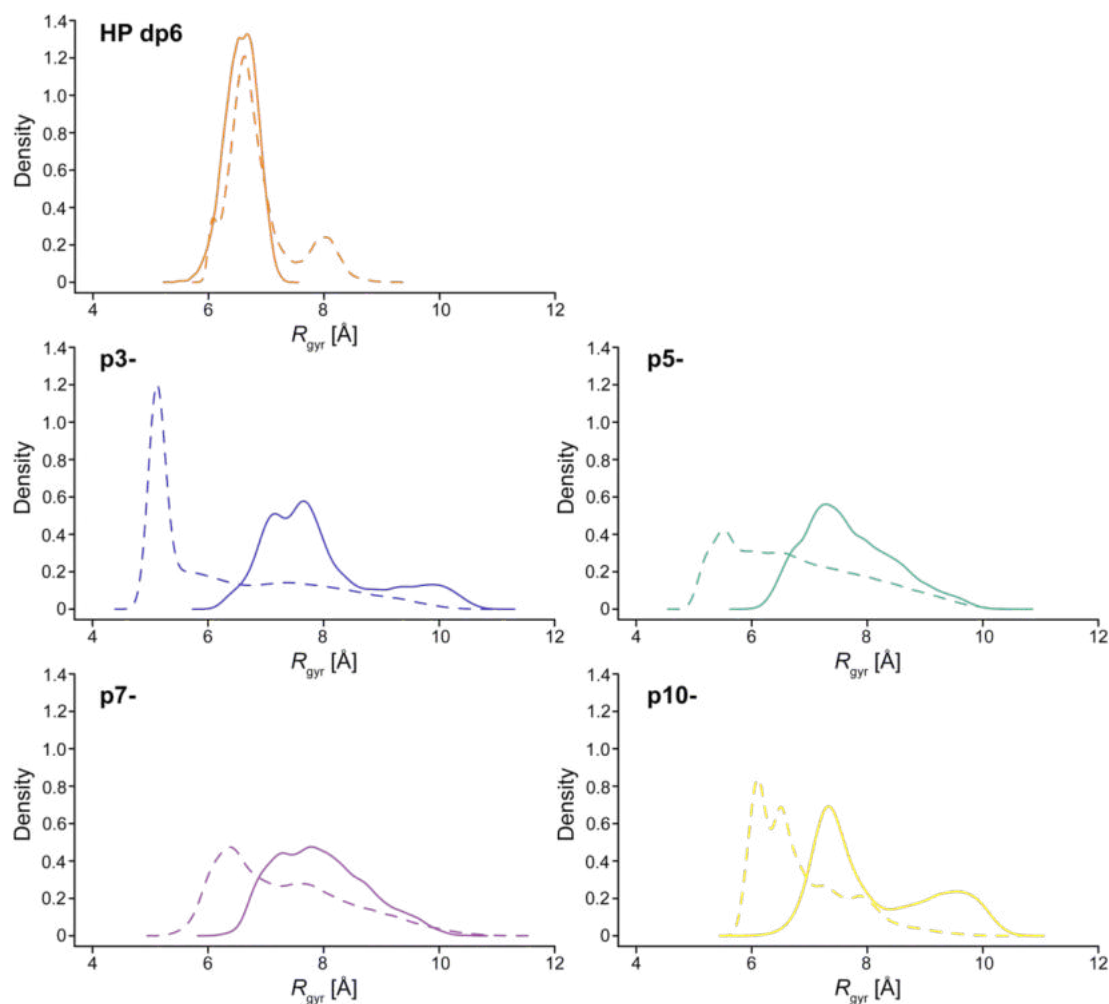


Figure **3-3**: Distributions of radius of gyration (R_{gyr}) values (kernel density estimation) for bound and unbound states in continuous and dashed lines, respectively, for HP dp6, p3-, p5-, p7-, and p10-. Figure adapted with permission of the publisher from Schulze C, et al. *Ligand binding of interleukin-8: a comparison of glycosaminoglycans and acidic peptides*. Phys Chem Chem Phys. 2023;25(36):24930-24947. Author rights retained.

3.3 Extension of the SUGRES-1P Coarse-Grained Model of Polysaccharides to Heparin

Studying the structure and dynamics of HP through classical AA MD simulations is resource-intensive. The ability of long HP chains to form large multiprotein assemblies further complicates MD simulations by necessitating longer timescales to accurately capture structural changes. CG methods present a more computationally efficient alternative by grouping atoms into beads, thereby reducing complexity and computational cost while retaining critical molecular properties. CG approaches effectively address challenges arising from the periodicity and diversity of GAGs [298]. Despite the biological importance of GAGs, only a limited number of specialized CG models have been developed for these molecules. Many earlier models have either been overly complex or excessively simplified, limiting their ability to effectively simulate long HP structures [298, 300, 299]. One CG model that achieves a suitable balance between simplicity and accuracy is SUGRES-1P, which provides sufficient precision to capture the global structural properties of HP, potentially useful for elucidating its roles in regulating blood coagulation, maintaining protein gradients essential for cell signaling, and molecular transport [336, 337, 338, 339], and mediating structural processes such as collagen reorganization [340, 341] and amyloid formation [342, 343].

The work presented in this publication aimed to extend the SUGRES-1P CG model to HP molecules and validate it through comparisons with experimental data. As part of the UNICORN framework, SUGRES-1P represents polysaccharide chains using one CG site per sugar residue, positioned between glycosidic linkage oxygen atoms that serve as geometrical anchor points [303]. The model was now refined by representing each CG bead as an ellipsoid (fig. **3-6**) and incorporating two interaction sites per bead to account for the off-center charge distribution in HP residues [344].

The SUGRES-1P force field is physics-based. By deriving all bonded and non-bonded interaction potentials from first-principles calculations, it provides an accurate model of atomic and molecular behavior. This approach enables the accurate representation of molecular interactions and, as a result, better predictions of energy states and conformational changes over time. Although the UNICORN framework utilizes solely an implicit solvent model, Debye-Hückel screening is applied to electrostatic interactions to account for ionic strength, thereby capturing the influence of factors such as salt concentration on electrostatic forces.

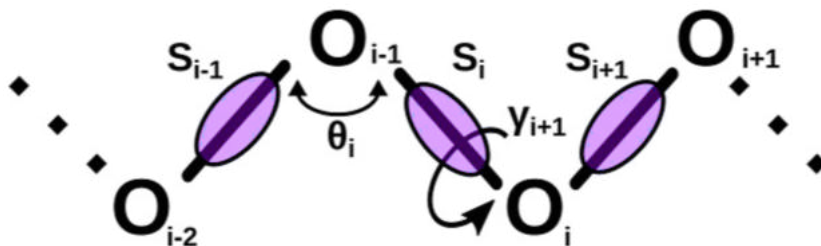


Figure **3-4**: The SUGRES-1P model of HP. The interaction sites correspond to united sugar rings, represented by violet ellipsoids, located halfway between glycosidic oxygen atoms (shown as white spheres) that serve to define the geometry of the polysaccharide molecule. The virtual bonds connecting the oxygen atoms are shown as thick black lines. The geometry of the polysaccharide chain is defined by the virtual bond angles θ_i and torsional angles γ_i .

Following the computational implementation of the CG model, I fine-tuned the force field parameters, including energy terms and electrostatic screening factors, to ensure accurate simulations of the structure and dynamics of highly charged HP molecules. An empirical adjustment was conducted on HP chains of three lengths (dp6, dp12, dp24) to find the optimal parameter set across different oligosaccharide lengths. MD simulations were used to collect data on end-to-end distance (EED) and radii of gyration (Rg) of free HP chains to guide the modifications of the Debye-Hückel parameter κ , which accounts for differing ion concentrations, as well as of energy term weights governing bonded and nonbonded interactions. Higher κ values as well as increased torsional-energy weights promoted coiling of the HP chains, resulting in more compact conformations. Conversely, increasing the electrostatic term weight led to extended HP chains due to stronger repulsion within the chain. Notably, lower κ values (i.e., lower ion concentrations) and higher electrostatic energy weights provided the best fit for shorter HP chains, while longer chains required higher κ and adjustments to bond-stretching and torsional energy weights to match experimental data.

After optimizing the parameters, the performance of SUGRES-1P in modeling the structure of free HP chains was evaluated through extended CG MD simulations across a broad range of chain lengths (dp6 to dp68), followed by comparison of the obtained results to experimental data on EED and Rg [345, 346]. The optimized parameters accurately captured the overall shape, flexibility, and global structural characteristics of HP chains (fig. **3-5**). Simulated EED and Rg values closely matched experimental data, although shorter chains exhibited slightly

more extended conformations, while longer chains tended to adopt more coiled structures compared to their experimentally determined counterparts. Notably, the simulations successfully replicated a structural kink observed experimentally in chains longer than dp14 [345]. Visual comparisons confirmed the accuracy of the model, with CG chains closely resembling experimental structures. Additionally, RMSD analysis confirmed the model's capability to reproduce the dynamic flexibility of HP within experimentally observed ranges.

The SUGRES-1P model demonstrated remarkable accuracy in simulating the structure and dynamics of HP molecules, establishing itself as a reliable tool for future GAG simulations. Its success in modeling free HP chains provided a solid foundation for applications in studying protein/HP interactions. Furthermore, the integration of SUGRES-1P with the UNICORN framework enables the simulation of larger, more complex biomolecular systems, including HP, proteins, lipids, and ions.

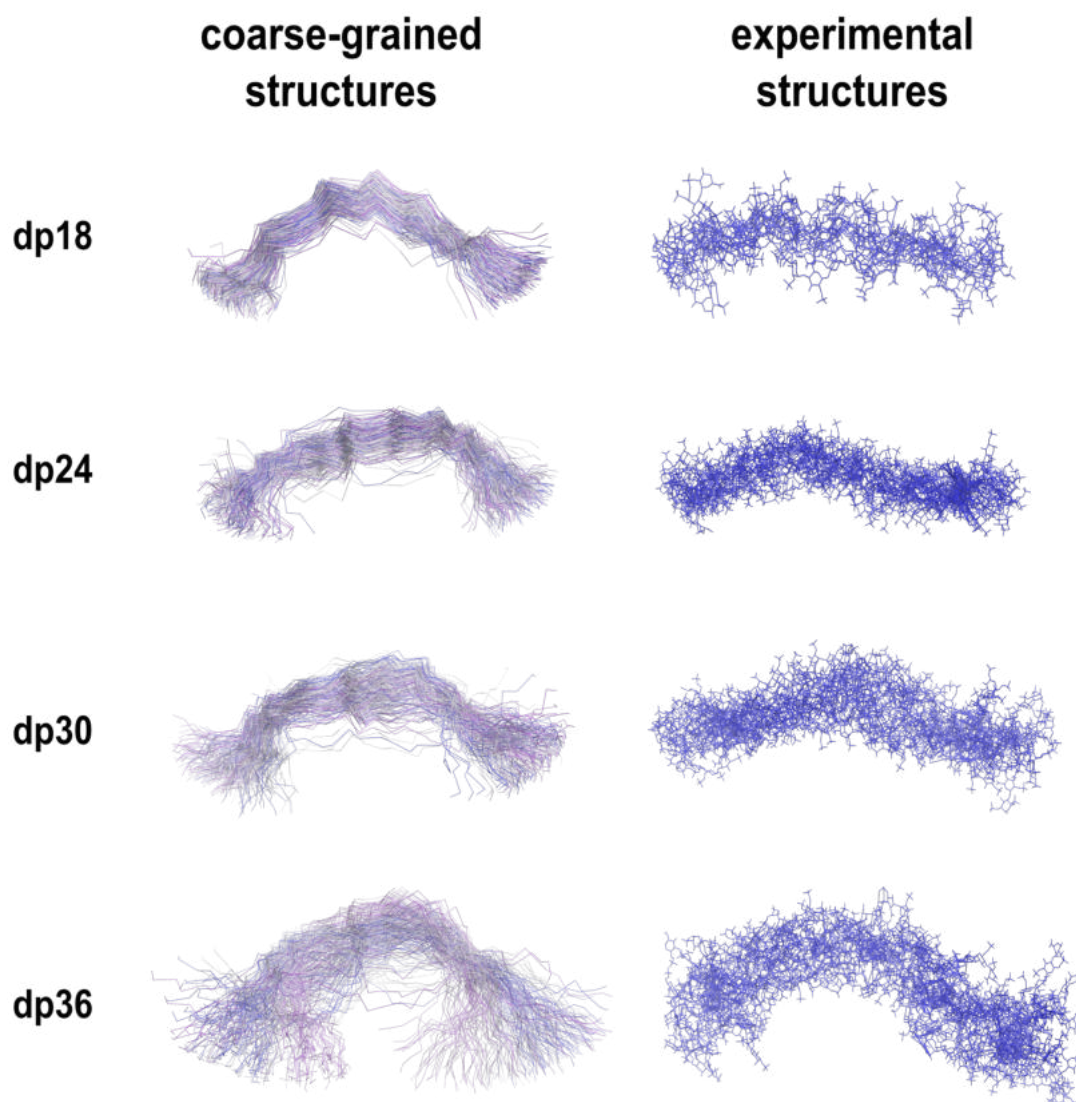


Figure **3-5**: Representative conformations of the CG trajectories of HP dp18, dp24, dp30, and dp36 visualized in trace representations, colored according to trajectory time step (left); experimentally determined conformation of HP dp18 (PDB ID: 1IRI), dp24 (1IRJ), dp30 (1IRK), dp36 (1IRL) in licorice representation (right). Figure adapted with permission of the publisher from Danielsson A, et al. *Extension of the SUGRES-1P Coarse-Grained Model of Polysaccharides to Heparin*. J Chem Theory Comput. 2023;19(17):6023-6036. Author rights retained.

3.4 Implementation of the UNRES/SUGRES-1P Coarse-Grained Model of Heparin for Simulating Protein/Heparin Interactions

Following the implementation and validation of the SUGRES-1P model for free HP molecules, the next step was an extensive and thorough analysis of its performance for protein/HP complexes. By combining the CG UNRES representation for proteins with SUGRES-1P for HP, a computationally efficient framework was established for accurately modeling protein/HP interactions. The SUGRES-1P model simplifies each HP residue into one CG bead composed of two interaction sites, while UNRES-1P represents amino acids with two CG beads, one for the side chain and one for the backbone peptide group. The combined CG energy function included three components modeling protein/protein, HP/HP, and protein/HP interactions. Force field parameters were calibrated and optimized to accurately model protein/HP binding interactions.

CG MD simulations of three protein/HP complexes, each spanning approximately 1 μ s of all-atom time, were conducted to study the binding dynamics and protein stability influenced by HP. The proteins studied were basic and acidic Fibroblast Growth Factor (bFGF, aFGF), and the NK1 splicing variant of Heparatocyte Growth Factor/Scatter Factor (HGF/SF), all complexed with HP dp6, while aFGF was also complexed with HP dp8 and dp10 to assess the effect of HP chain length on aFGF dimer stability. After energy minimization, ten replicates of canonical CG MD simulations were run for each system to sample a broad range of conformations. Protein/HP interactions were analyzed by clustering the CG structures to identify the most probable conformations and protein/HP contact maps. Structural deviation from the experimentally determined crystal structures and flexibility of the interacting molecules were assessed using RMSD and RMSF calculations, while Principal Component Analysis (PCA) was used to capture major modes of motion, highlighting conformational shifts during the MD simulations. The results were validated against crystallographic data to assess the model's predictive accuracy.

The results for the bFGF/HP dp6 complex demonstrated that the UNRES/SUGRES-1P CG model accurately captured the interaction, validating its predictive capability. The complex remained stable across all 10 CG MD replicates, with no dissociation of HP from bFGF. RMSD values showed effective modeling of structural dynamics, with 4.6 Å for bFGF and 11.7 Å for HP compared to experimental data. The CG simulations enabled the successful identification of

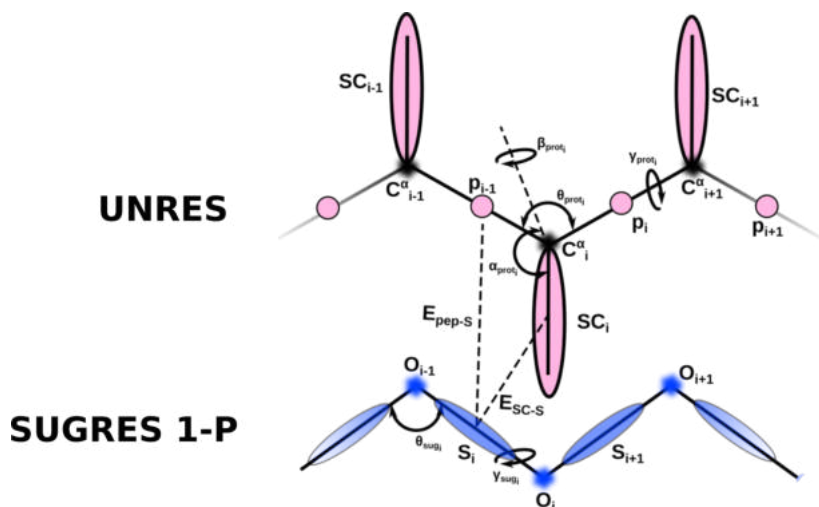


Figure 3-6: The coarse-grained UNRES and SUGRES-1P models. The unified CG peptide groups (p) are represented by a pink filled circle, the unified side chain beads (SC) by a pink filled ellipsoid. The unified CG sugar rings of the polysaccharide (S) are shown as blue filled ellipsoids. The C α atoms of the polypeptide and the O $_4$ atoms of the polysaccharide are used to define the geometry of the polymer chains and do not function as interaction sites. The virtual bonds connecting the CG beads to the C α and O atoms are shown as thick black lines. Figure adapted with permission of the publisher from Danielsson A, et al. *Implementation of the UNRES/SUGRES-1P Coarse-Grained Model of Heparin for Simulating Protein/Heparin Interactions*. J Chem Theory Comput. 2024;20(23):10703-10715. Author rights retained.

HP-binding residues, which aligned closely with experimental data, particularly Arg-82, Lys-120, Gln-124, Lys-126, and Lys-136, thereby confirming accurate binding behavior [347, 348, 349]. The orientation of HP within the binding site varied slightly across the CG clusters, highlighting the adaptability of the binding site. PCA revealed conformational changes in bFGF upon HP binding. Notably, HP stabilized the bFGF dimer, as simulations of free bFGF dimers showed intermittent dissociation and reformation of the dimer, whereas in the presence of HP, the dimer remained consistently stable throughout the simulation.

The UNRES/SUGRES-1P CG model also effectively captured the dynamics and interactions between NK1 and HP dp6. Although NK1 remained bound to HP in only 5 out of 10 simulation replicates, the HP-binding site aligned well with experimentally known sites. Key HP-binding residues were accurately identified

and consistently formed contacts with HP, despite notable flexibility in binding poses across simulations. PCA revealed limited motion of NK1 when bound to HP, indicating complex stability, though increased mobility was observed in the flexible N-terminal region. RMSF analysis confirmed this finding, showing that HP-binding regions were stable, while the N-terminal was the most flexible.

In the aFGF/HP system, the combined UNRES/SUGRES-1P model captured the effects of varying HP chain lengths on HP-mediated aFGF dimer stability. The aFGF/HP complex remained stable in most simulations, with no dissociation in 8 of 10 replicates for HP dp6, and 6 of 10 for both dp8 and dp10. Interestingly, for dp10 one replicate showed reassociation towards the end of the simulation, indicating that longer HP chains may promote dynamic binding to aFGF. The simulations also revealed that in complex with longer HP chains (dp8 and dp10), the aFGF dimer adopted a more compact structure, likely due to stronger HP-mediated interactions between aFGF monomers. Key HP-binding residues on aFGF were identified and consistently matched experimentally determined binding sites. Contact frequency between aFGF and HP increased with chain length, with longer chains forming more stable contacts. PCA indicated a twisting and squeezing motion of the aFGF dimer around the HP chain, with monomer rotation decreasing as HP length increased, suggesting that longer chains limit dimer flexibility.

The SUGRES-1P model has proven accurate in simulating the structure and dynamics of HP molecules, making it a reliable tool for future protein/GAG simulations. Furthermore, its integration with the UNICORN framework allows for the simulation of larger, more complex biomolecular systems involving proteins, lipids, nucleic acids, and ions, significantly expanding its potential in advanced biological research.

3.5 Molecular Dynamics-Based Descriptors of 3-O-Sulfated Heparan Sulfate as Contributors of Protein Binding Specificity

HS molecules play critical roles in biological processes such as coagulation, viral infection, and cell signaling, primarily through their interactions with various proteins. However, the complexity and variability in HS sulfation patterns and their template-less synthesis make understanding these binding mechanisms difficult. Research suggests that protein/GAG binding is subject to conformational selection, where the GAG fragments adopt protein-binding conformations even in their free state, reducing the need for significant structural changes during binding, particularly in electrostatically driven interactions [299, 350, 351]. For instance, Guglier et al. found that the inherent flexibility of HP allows it to meet FGF2's binding requirements without needing further conformational shifts [352]. The sequence-dependent flexibility of HS, shaped by specific sulfation patterns, plays an important role in adopting these pre-binding conformations, which are crucial for the binding specificity of proteins like FGF receptors [353]. Nonetheless, some studies argue that while the free-state conformation is important, final adjustments occur upon protein binding [354].

In their study, Chopra et al. developed a modular synthetic approach to generate a library of 27 structurally diverse HS hexasaccharides, systematically varying sulfation patterns and backbone structures [355]. This library provided a powerful platform for elucidating the binding requirements of HS-binding proteins, revealing distinct preferences for different sulfation motifs in physiological and pathological processes. The study included a diverse set of HS-binding proteins to explore the ligand requirements of 3-O-sulfated HS. Antithrombin-III (AT-III) and heparin cofactor II (HC-II) were chosen for their roles in blood coagulation and dependence on specific sulfation patterns, while fibroblast growth factors (FGF-7, FGF-9) and FGFR1 were included to investigate their sulfation-specific binding in cell signaling and development. Proteins like neuropilin-1 (Nrp-1) and receptor for advanced glycation end products (RAGE) were studied for their involvement in angiogenesis and inflammatory diseases, respectively, and stabilin-2 was selected for its role in scavenging and clearance processes. Bone morphogenetic protein-2 (BMP-2) was examined for its therapeutic potential in bone regeneration.

Based on this unique library of 27 synthetic HS hexasaccharides and the results of Chopra et al. [355], the presented study aimed to elucidate the specificity of

protein/HS interactions, with a focus on the rare 3-O-sulfation of HS molecules, through the application of computational techniques and unsupervised machine learning algorithms. Molecular descriptors of the HS molecules were identified to explain protein/HS binding, based on the structural properties of unbound HS molecules observed in all-atom MD simulations. All-atom MD simulations of the unbound HS molecules were conducted using AMBER16 and the GLYCAM06-j force field with the explicit TIP3P water model. A range of physicochemical descriptors were extracted from the MD trajectories, including intramolecular H-bonds, end-to-end distance (EED), root-mean-square fluctuation (RMSF), radii of gyration (Rg), dipole moments, glycosidic dihedral angles and percentage in minimum-energy conformations, ring puckers, total free energy and individual energy components (van der Waals, electrostatic, solvation), and configurational entropy. These descriptors were used to assess the stability, conformational freedom and flexibility of the HS molecules, and uncover associations between these descriptors and the HS molecules' binding potential.

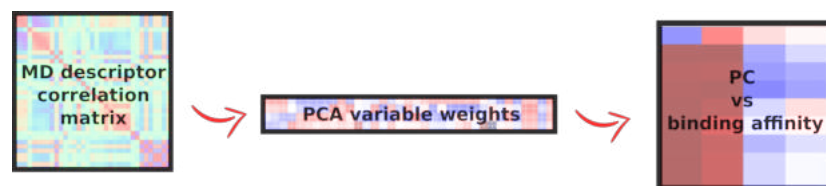


Figure 3-7: The correlation between MD-derived descriptors of free GAGs was analyzed using Pearson correlation coefficients and plotted as a correlation matrix. The standardized data was subjected to PCA to uncover underlying patterns and construct Principal Components, i.e. new dimensions that are each a linear combination of the original MD-derived descriptors. The original data set was transformed into the new PC-dimensionality, and the values corresponding to each of the HS molecules in this new dimension were correlated with binding affinity data obtained from Chopra et al. [355]. Figure adapted with permission of the publisher from Danielsson A, et al. *Molecular dynamics-based descriptors of 3-O-Sulfated Heparan sulfate as contributors of protein binding specificity*. Comput Biol Chem. 2022;99:107716. Author rights retained.

Principal Component Analysis (PCA) was applied to the descriptor dataset to reduce dimensionality and extract the most important patterns from the data. PCA transformed the descriptors into new uncorrelated variables, or PCs, capturing the main variability in the dataset. Pearson correlation analysis was

then used to assess the relationship between these descriptors and protein/HS binding affinity, identifying those most relevant to binding specificity. Linear regression models evaluated how well the PCA-derived components explained binding affinity. Hierarchical clustering of the correlation matrix further revealed patterns and subgroups, grouping proteins based on their correlation with HS molecular features, offering insights into binding affinity determinants.

The MD simulations revealed significant structural variability among the HS molecules. RMSD, Rg, and EED indicated that most HS molecules were flexible, adopting multiple conformations throughout the simulation. Certain HS sequences were observed to form compact stable structures toward the simulation's end, with increased intramolecular H-bonds correlating to greater rigidity, potentially affecting protein interactions. Sugar ring puckering was also found to contribute to increased flexibility in some of the HS molecules. Pearson correlation analysis revealed that specific descriptors strongly correlated with binding affinity for certain proteins. Following PCA, linear regression models confirmed that descriptors from unbound HS molecules were reliable indicators of binding strength. Clustering based on protein correlations with PCs identified three distinct protein groups: group 1 (BMP-2, FGFR1, Stabilin-2), where binding affinity was influenced by H-bonds, ring puckering, and molecular flexibility; group 2 (RAGE, Nrp-1, FGF-7, FGF-9), which preferred highly flexible HS molecules; and AT-III, which formed a separate cluster, as its binding affinity was not well explained by the selected descriptors, indicating that other factors might contribute to its specificity.

The study confirmed that different proteins exhibit distinct preferences for specific structural and physicochemical features of HS molecules, strongly indicating that the conformational space of the unbound HS plays a role in determining protein binding specificity. Descriptors from MD simulations successfully explained much of the binding affinity, although further research incorporating protein descriptors and more complex interactions is necessary for a more comprehensive understanding of interaction specificity.

3.6 Outlook

This thesis advances the understanding of protein/GAG interactions by providing new insights into the roles of electrostatics, sulfation patterns, and molecular flexibility in binding specificity. The studies presented integrate experimental data with multiscale computational modeling, establishing a versatile framework for future investigations of protein/GAG complexes.

While this thesis furthers the understanding of protein/GAG binding specificity, several limitations remain. The computational approaches employed necessarily rely on approximations, simplified solvent models, parameter optimization from limited benchmarks, and restricted sampling of large biomolecular systems, thus enabling the exploration of complexes otherwise inaccessible to experimental approaches.

Future work should focus on continued force field refinement, rigorous validation, and expansion of datasets to improve the predictive accuracy of computational approaches. The extension of the SUGRES-1P and UNRES/SUGRES-1P frameworks to diverse GAG libraries and multivalent protein assemblies will enable large-scale, high-throughput simulations. Furthermore, developments in GPU acceleration and machine-learning-based analysis of structural descriptors and free energy landscapes offer the potential to model extracellular protein/GAG networks at physiologically relevant scales, bridging atomistic detail with mesoscale organization.

Although the biological motivation for this thesis lies in understanding GAG function, its primary contribution is methodological. The workflows presented here offer a generalizable platform for studying molecular recognition across chemically complex systems. The work presented in this thesis demonstrates how computational chemistry can complement experimental studies and accelerate hypothesis testing.

By integrating computational simulations with experimental validation, this work establishes computational and experimental frameworks that deepen understanding of protein/GAG interactions and provide a foundation for future advances in molecular modeling.

Bibliography

- [1] Jeffrey D. Esko, Koji Kimata, and Ulf Lindahl. *Proteoglycans and Sulfated Glycosaminoglycans*. 2009.
- [2] Cristal I. Gama, Sarah E. Tully, Naoki Sotogaku, Peter M. Clark, Manish Rawat, Nagarajan Vaidehi, William A. Goddard, Akinori Nishi, and Linda C. Hsieh-Wilson. Sulfation patterns of glycosaminoglycans encode molecular recognition and activity. *Nature Chemical Biology*, 2(9):467–473, 2006.
- [3] Diana Soares da Costa, Rui L. Reis, and Iva Pashkuleva. Sulfation of Glycosaminoglycans and Its Implications in Human Health and Disorders. *Annual Review of Biomedical Engineering*, 19:1–26, 2017.
- [4] Vince Hascall and Jeffrey D. Esko. Hyaluronan. 2009.
- [5] Run Fang, Qifei Jiang, Yukun Guan, Pengfei Gao, Rui Zhang, Zhen Zhao, and Zhengfan Jiang. Golgi apparatus-synthesized sulfated glycosaminoglycans mediate polymerization and activation of the cGAMP sensor STING. *Immunity*, 54(5):962–975.e8, 2021.
- [6] Maria Fuller, Peter J. Meikle, and John J. Hopwood. Glycosaminoglycan degradation fragments in mucopolysaccharidosis I. *Glycobiology*, 14(5):443–450, 2004.
- [7] Melanie Simpson, Liliana Schaefer, Vincent Hascall, and Jeffrey D. Esko. Hyaluronan. 2022.
- [8] Beatriz Lázaro, Prudencio Alonso, Andrea Rodriguez, Manuel La Nuez, Florencio Marzo, and Julio G. Prieto. Characterization of the visco-elastic properties of hyaluronic acid. *Biorheology*, 55(1):41–50, 2018.
- [9] Simon J. Haward. Characterization of hyaluronic acid and synovial fluid in stagnation point elongational flow. *Biopolymers*, 101(3):287–305, 2014.

-
- [10] Xu Li, Bingyang Dai, Jiabin Guo, Lizhen Zheng, Quanyi Guo, Jiang Peng, Jiankun Xu, and Ling Qin. Nanoparticle–Cartilage Interaction: Pathology-Based Intra-articular Drug Delivery for Osteoarthritis Therapy. *Nano-Micro Letters*, 13:149, 2021.
- [11] J.-P. Li and M. Kusche-Gullberg. Heparan Sulfate: Biosynthesis, Structure, and Function. *Int Rev Cell Mol Biol*, 325:215–273, 2016.
- [12] N. K. Karamanos, P. Vanky, G. N. Tzanakakis, T. Tseggenidis, and A. Hjerpe. Ion-pair high-performance liquid chromatography for determining disaccharide composition in heparin and heparan sulphate. *Journal of Chromatography. A*, 765(2):169–179, 1997.
- [13] Victor Schultz, Mathew Suffita, Xinyue Liu, Xing Zhang, Yanlei Yu, Lingyun Li, Dixy E. Green, Yongmei Xu, Fuming Zhang, Paul L. DeAngelis, Jian Liu, and Robert J. Linhardt. Heparan Sulfate Domains Required for Fibroblast Growth Factor 1 and 2 Signaling through Fibroblast Growth Factor Receptor 1c. *The Journal of Biological Chemistry*, 292(6):2495–2509, 2017.
- [14] Adam B. Cadwallader and H. Joseph Yost. Combinatorial expression patterns of heparan sulfate sulfotransferases in zebrafish: I. The 3-O-sulfotransferase family. *Developmental Dynamics*, 235(12):3423–3431, 2006.
- [15] Preeti Ravindra Bhoge, Rakesh Raigawali, Sandhya Mardhekar, Saurabh Anand, and Raghavendra Kikkeri. Synergistic interplay of uronic acid and sulfation composition of heparan sulfate on molecular recognition to activity. *Carbohydrate Research*, 532:108919, 2023.
- [16] Shao-Chen Lee, Hong-Hsiang Guan, Chia-Hui Wang, Wei-Ning Huang, Siu-Cin Tjong, Chun-Jung Chen, and Wen-guey Wu. Structural Basis of Citrate-dependent and Heparan Sulfate-mediated Cell Surface Retention of Cobra Cardiotoxin A3*. *Journal of Biological Chemistry*, 280(10):9567–9577, 2005.
- [17] Samuel G. Holmes, Balaji Nagarajan, and Umesh R. Desai. 3-O-Sulfation induces sequence-specific compact topologies in heparan sulfate that encode a dynamic sulfation code. *Computational and Structural Biotechnology Journal*, 20:3884–3898, 2022.

- [18] Ivan O. Potapenko, Vilde D. Haakensen, Torben Lüders, Aslaug Helland, Ida Bukholm, Therese Sørli, Vessela N. Kristensen, Ole C. Lingjaerde, and Anne-Lise Børresen-Dale. Glycan gene expression signatures in normal and malignant breast tissue; possible role in diagnosis and progression. *Molecular Oncology*, 4(2):98–118, 2010.
- [19] Christina J. Malavaki, Achilleas D. Theocharis, Fotini N. Lamari, Ioannis Kanakis, Theodore Tsegenidis, George N. Tzanakakis, and Nikos K. Karamanos. Heparan sulfate: biological significance, tools for biochemical analysis and structural characterization. *Biomedical chromatography: BMC*, 25(1-2):11–20, 2011.
- [20] R. D. Rosenberg. Role of heparin and heparinlike molecules in thrombosis and atherosclerosis. *Federation Proceedings*, 44(2):404–409, 1985.
- [21] D. J. Tyrrell, A. P. Horne, K. R. Holme, J. M. Preuss, and C. P. Page. Heparin in inflammation: potential therapeutic applications beyond anticoagulation. *Advances in Pharmacology (San Diego, Calif.)*, 46:151–208, 1999.
- [22] Alok A. Khorana, Abha Sahni, Owen D. Altland, and Charles W. Francis. Heparin inhibition of endothelial cell proliferation and organization is dependent on molecular weight. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 23(11):2110–2115, 2003.
- [23] Juliana M. Motta, Kayene V. A. Micheli, Carlos Roberto-Fernandes, Michelle Hermsdorff-Brandt, Alessandra L. Guedes, Flávia S. Frattani, Paulo A. S. Mourão, and Mariana S. Pereira. A low-anticoagulant heparin suppresses metastatic dissemination through the inhibition of tumor cell-platelets association. *Biomedicine & Pharmacotherapy = Biomedecine & Pharmacotherapie*, 171:116108, 2024.
- [24] L. Djerbal, H. Lortat-Jacob, and Jef Kwok. Chondroitin sulfates and their binding molecules in the central nervous system. *Glycoconjugate Journal*, 34(3):363–376, 2017.
- [25] Anders Malmström, Barbara Bartolini, Martin A. Thelin, Benny Pacheco, and Marco Maccarana. Iduronic Acid in Chondroitin/Dermatan Sulfate. *Journal of Histochemistry and Cytochemistry*, 60(12):916–925, 2012.

- [26] Kristen R. Taylor, Jennifer A. Rudisill, and Richard L. Gallo. Structural and Sequence Motifs in Dermatan Sulfate for Promoting Fibroblast Growth Factor-2 (FGF-2) and FGF-7 Activity*. *Journal of Biological Chemistry*, 280(7):5300–5306, 2005.
- [27] James L. Funderburgh. Keratan Sulfate Biosynthesis. *IUBMB life*, 54(4):187–194, 2002.
- [28] Neeraj Vij, Luke Roberts, Sarah Joyce, and Shukti Chakravarti. Lumican suppresses cell proliferation and aids Fas-Fas ligand mediated apoptosis: implications in the cornea. *Experimental Eye Research*, 78(5):957–971, 2004.
- [29] Mariusz Malinowski, Katarzyna Pietraszek, Corinne Perreau, Mateusz Boguslawski, Véronique Decot, Jean-François Stoltz, Laurent Vallar, Jolanta Niewiarowska, Czesław Cierniewski, François-Xavier Maquart, Yanusz Wegrowski, and Stéphane Brézillon. Effect of lumican on the migration of human mesenchymal stem cells and endothelial progenitor cells: involvement of matrix metalloproteinase-14. *PloS One*, 7(12):e50709, 2012.
- [30] Leona T.Y. Ho, Anthony M. Harris, Hidetoshi Tanioka, Naoto Yagi, Shigeru Kinoshita, Bruce Caterson, Andrew J. Quantock, Robert D. Young, and Keith M. Meek. A comparison of glycosaminoglycan distributions, keratan sulphate sulphation patterns and collagen fibril architecture from central to peripheral regions of the bovine cornea. *Matrix Biology*, 38:59–68, 2014.
- [31] Y. Sommarin, M. Wendel, Z. Shen, U. Hellman, and D. Heinegård. Osteoadherin, a cell-binding keratan sulfate proteoglycan in bone, belongs to the family of leucine-rich repeat proteins of the extracellular matrix. *The Journal of Biological Chemistry*, 273(27):16723–16729, 1998.
- [32] R. M. Lauder, T. N. Huckerby, and I. A. Nieduszynski. The structure of the keratan sulphate chains attached to fibromodulin from human articular cartilage. *Glycoconjugate Journal*, 14(5):651–660, 1997.
- [33] Christian Schulze, Annemarie Danielsson, Adam Liwo, Daniel Huster, Sergey A. Samsonov, and Anja Penk. Ligand binding of interleukin-8: a comparison of glycosaminoglycans and acidic peptides. *Physical chemistry chemical physics: PCCP*, 25(36):24930–24947, 2023.

- [34] Tomoko Watanabe, Ko Takeda, Keiko Hiemori, Toshikazu Minamisawa, and Hiroaki Tateno. A glycosaminoglycan microarray identifies the binding of SARS-CoV-2 spike protein to chondroitin sulfate E. *FEBS letters*, 595(18):2341–2349, 2021.
- [35] H. Ueda, T. Saitoh, K. Kojima, and H. Ogawa. Multi-specificity of a *Psathyrella velutina* mushroom lectin: heparin/pectin binding occurs at a site different from the N-acetylglucosamine/N-acetylneuraminic acid-specific site. *Journal of Biochemistry*, 126(3):530–537, 1999.
- [36] G. J. Taylor, S. C. Yorke, and D. R. Harding. Glycosaminoglycan specificity of a heparin-binding peptide. *Peptide Research*, 8(5):286–293, 1995.
- [37] Catherine L. R. Merry, Ulf Lindahl, John Couchman, and Jeffrey D. Esko. Proteoglycans and Sulfated Glycosaminoglycans. 2022.
- [38] Per Jemth, Emanuel Smeds, Anh-Tri Do, Hiroko Habuchi, Koji Kimata, Ulf Lindahl, and Marion Kusche-Gullberg. Oligosaccharide library-based assessment of heparan sulfate 6-O-sulfotransferase substrate specificity. *The Journal of Biological Chemistry*, 278(27):24371–24376, 2003.
- [39] Jinghua Chen, Michael B. Duncan, Kevin Carrick, R. Marshall Pope, and Jian Liu. Biosynthesis of 3-O-sulfated heparan sulfate: unique substrate specificity of heparan sulfate 3-O-sulfotransferase isoform 5. *Glycobiology*, 13(11):785–794, 2003.
- [40] M. P. Van Damme, J. M. Moss, W. H. Murphy, and B. N. Preston. Binding properties of glycosaminoglycans to lysozyme—effect of salt and molecular weight. *Arch Biochem Biophys*, 310(1):16–24, 1994.
- [41] Jonathan M. Moss, Marie-Paule I. Van Damme, William H. Murphy, and Barry N. Preston. Dependence of Salt Concentration on Glycosaminoglycan–Lysozyme Interactions in Cartilage. *Archives of Biochemistry and Biophysics*, 348(1):49–55, 1997.
- [42] Suman Samantray, Olujide O. Olubiyi, and Birgit Strodel. The Influences of Sulphation, Salt Type, and Salt Concentration on the Structural Heterogeneity of Glycosaminoglycans. *Int J Mol Sci*, 22(21):11529, 2021.
- [43] Jacek Janusz Walkowiak, Matthias Ballauff, Ralf Zimmermann, Uwe Freudenberg, and Carsten Werner. Thermodynamic Analysis of the Inter-

- action of Heparin with Lysozyme. *Biomacromolecules*, 21(11):4615–4625, 2020.
- [44] Neha S. Gandhi and Ricardo L. Mancera. Free energy calculations of glycosaminoglycan-protein interactions. *Glycobiology*, 19(10):1103–1115, 2009.
- [45] Elizabeth H. Pempe, Yongmei Xu, Sandhya Gopalakrishnan, Jian Liu, and Edward N. Harris. Probing structural selectivity of synthetic heparin binding to Stabilin protein receptors. *The Journal of Biological Chemistry*, 287(25):20774–20783, 2012.
- [46] K. Gu, R. J. Linhardt, M. Laliberté, K. Gu, and J. Zimmermann. Purification, characterization and specificity of chondroitin lyases and glycuronidase from *Flavobacterium heparinum*. *The Biochemical Journal*, 312 (Pt 2)(Pt 2):569–577, 1995.
- [47] Charlotte B. Spliid, Alejandro Gomez Toledo, Patience Sanderson, Yang Mao, Francesco Gatto, Tobias Gustavsson, Swati Choudhary, Ana L. Saldanha, Rasmus P. Vogelsang, Ismail Gögenur, Thor G. Theander, Franklin E. Leach, I. Jonathan Amster, Jeffrey D. Esko, Ali Salanti, and Thomas Mandel Clausen. The specificity of the malarial VAR2CSA protein for chondroitin sulfate depends on 4-O-sulfation and ligand accessibility. *The Journal of Biological Chemistry*, 297(6):101391, 2021.
- [48] Rahul Raman, Ganesh Venkataraman, Steffen Ernst, V. Sasisekharan, and Ram Sasisekharan. Structural specificity of heparin binding in the fibroblast growth factor family of proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 100(5):2357–2362, 2003.
- [49] Yongmei Xu, Andrea F. Moon, Shuqin Xu, Juno M. Krahn, Jian Liu, and Lars C. Pedersen. Structure Based Substrate Specificity Analysis of Heparan Sulfate 6-O-Sulfotransferases. *ACS chemical biology*, 12(1):73–82, 2017.
- [50] Leonhard Möckl, Kayvon Pedram, Anish R. Roy, Venkatesh Krishnan, Anna-Karin Gustavsson, Oliver Dorigo, Carolyn R. Bertozzi, and W. E. Moerner. Quantitative super-resolution microscopy of the mammalian glycolyx. *Dev Cell*, 50(1):57–72.e6, 2019.

- [51] Mansi Gupta, Thomas Kurth, Fabian Heinemann, Petra Schwille, Sebastian Keil, Franziska Knopf, and Michael Brand. Fine-tuning of Fgf8 morphogen gradient by heparan sulfate proteoglycans in the extracellular matrix. *Biophys J*, 124(6):996–1010, 2025.
- [52] Megan Kizer, Peiqin Li, Brady F. Cress, Lei Lin, Tom T. Jing, Xing Zhang, Ke Xia, Robert J. Linhardt, and Xing Wang. RNA Aptamers with Specificity for Heparosan and Chondroitin Glycosaminoglycans. *ACS omega*, 3(10):13667–13675, 2018.
- [53] Kyle M. Koss, Terrance J. Sereda, Vlad K. Kumirov, and Jason A. Wertheim. A class of peptides designed to replicate and enhance the Receptor for Hyaluronic Acid Mediated Motility binding domain. *Acta Biomaterialia*, 167:293–308, 2023.
- [54] Craig Freeman, Ligong Liu, Martin G. Banwell, Kathryn J. Brown, Anna Bezos, Vito Ferro, and Christopher R. Parish. Use of sulfated linked cyclitols as heparan sulfate mimetics to probe the heparin/heparan sulfate binding specificity of proteins. *The Journal of Biological Chemistry*, 280(10):8842–8849, 2005.
- [55] Laurent Maveyraud and Lionel Mourey. Protein X-ray Crystallography and Drug Discovery. *Molecules (Basel, Switzerland)*, 25(5):1030, 2020.
- [56] Wendy J. Carter, Evis Cama, and James A. Huntington. Crystal Structure of Thrombin Bound to Heparin*. *Journal of Biological Chemistry*, 280(4):2745–2749, 2005.
- [57] L. Jin, J. P. Abrahams, R. Skinner, M. Petitou, R. N. Pike, and R. W. Carrell. The anticoagulant activation of antithrombin by heparin. *Proceedings of the National Academy of Sciences of the United States of America*, 94(26):14683–14688, 1997.
- [58] Jonathan Langdown, Klara J. Belzar, Wendy J. Savory, Trevor P. Baglin, and James A. Huntington. The critical role of hinge-region expulsion in the induced-fit heparin binding mechanism of antithrombin. *Journal of Molecular Biology*, 386(5):1278–1289, 2009.
- [59] Wei Li and James A. Huntington. Crystal structures of protease nexin-1 in complex with heparin and thrombin suggest a 2-step recognition mechanism. *Blood*, 120(2):459–467, 2012.

- [60] Wei Li, Ty E. Adams, Jyoti Nangalia, Charles T. Esmon, and James A. Huntington. Molecular basis of thrombin recognition by protein C inhibitor revealed by the 1.6-Å structure of the heparin-bridged complex. *Proceedings of the National Academy of Sciences of the United States of America*, 105(12):4661–4666, 2008.
- [61] Chenghua Shao, Fuming Zhang, Melissa M. Kemp, Robert J. Linhardt, David M. Waisman, James F. Head, and Barbara A. Seaton. Crystallographic analysis of calcium-dependent heparin binding to annexin A2. *The Journal of Biological Chemistry*, 281(42):31689–31695, 2006.
- [62] Young-Hyun Han, Marie-Line Garron, Hye-Yeon Kim, Wan-Seok Kim, Zhenqing Zhang, Kyeong-Seok Ryu, David Shaya, Zhongping Xiao, Chae-joon Cheong, Yeong Shik Kim, Robert J. Linhardt, Young Ho Jeon, and Miroslaw Cygler. Structural Snapshots of Heparin Depolymerization by Heparin Lyase I*. *Journal of Biological Chemistry*, 284(49):34019–34027, 2009.
- [63] David Shaya, Ante Tocilj, Yunge Li, James Myette, Ganesh Venkataraman, Ram Sasisekharan, and Miroslaw Cygler. Crystal Structure of Heparinase II from *Pedobacter heparinus* and Its Complex with a Disaccharide Product*. *Journal of Biological Chemistry*, 281(22):15525–15535, 2006.
- [64] James W. Murphy, Yoonsang Cho, Aristidis Sachpatzidis, Chengpeng Fan, Michael E. Hodsdon, and Elias Lolis. Structural and Functional Basis of CXCL12 (Stromal Cell-derived Factor-1 α) Binding to Heparin*. *Journal of Biological Chemistry*, 282(13):10018–10027, 2007.
- [65] D. Lietha, D. Y. Chirgadze, B. Mulloy, T. L. Blundell, and E. Gherardi. Crystal structures of NK1-heparin complexes reveal the basis for NK1 activity and enable engineering of potent agonists of the MET receptor. *The EMBO journal*, 20(20):5543–5555, 2001.
- [66] Ishan Capila, Mariá J Hernáiz, Y. D Mo, Tanya R Mealy, B Campos, John R Dedman, Robert J Linhardt, and Barbara A Seaton. Annexin V–Heparin Oligosaccharide Complex Suggests Heparan Sulfate–Mediated Assembly on Cell Surfaces. *Structure*, 9(1):57–64, 2001.
- [67] Wenguang G. Liang, Catherine G. Triandafillou, Teng-Yi Huang, Medel Manuel L. Zulueta, Shiladitya Banerjee, Aaron R. Dinner, Shang-Cheng

- Hung, and Wei-Jen Tang. Structural basis for oligomerization and glycosaminoglycan binding of CCL5 and CCL3. *Proceedings of the National Academy of Sciences of the United States of America*, 113(18):5000–5005, 2016.
- [68] Andrew B. Dykstra, Matt D. Sweeney, and Julie A. Leary. Structural Evidence for the Tetrameric Assembly of Chemokine CCL11 and the Glycosaminoglycan ArixtraTM. *Biomolecules*, 3(4):905–922, 2013.
- [69] L. Pellegrini, D. F. Burke, F. von Delft, B. Mulloy, and T. L. Blundell. Crystal structure of fibroblast growth factor receptor ectodomain bound to ligand and heparin. *Nature*, 407(6807):1029–1034, 2000.
- [70] J. Schlessinger, A. N. Plotnikov, O. A. Ibrahimi, A. V. Eliseenkova, B. K. Yeh, A. Yaron, R. J. Linhardt, and M. Mohammadi. Crystal structure of a ternary FGF-FGFR-heparin complex reveals a dual role for heparin in FGFR binding and dimerization. *Molecular Cell*, 6(3):743–750, 2000.
- [71] A. D. DiGabriele, I. Lax, D. I. Chen, C. M. Svahn, M. Jaye, J. Schlessinger, and W. A. Hendrickson. Structure of a heparin-linked biologically active dimer of fibroblast growth factor. *Nature*, 393(6687):812–817, 1998.
- [72] Yi Xue, Sangwon Lee, and Ya Ha. Crystal structure of amyloid precursor-like protein 1 and heparin complex suggests a dual role of heparin in E2 dimerization. *Proceedings of the National Academy of Sciences of the United States of America*, 108(39):16229–16234, 2011.
- [73] Kemin Tan, Mark Duquette, Jin-huan Liu, Jack Lawler, and Jia-huai Wang. The crystal structure of the heparin-binding reelin-N domain of f-spondin. *Journal of Molecular Biology*, 381(5):1213–1223, 2008.
- [74] Serge Pérez, François Bonnardel, Frédérique Lisacek, Anne Imberty, Sylvie Ricard Blum, and Olga Makshakova. GAG-DB, the New Interface of the Three-Dimensional Landscape of Glycosaminoglycans. *Biomolecules*, 10(12):1660, 2020.
- [75] Richard R. Ernst, Geoffrey Bodenhausen, Alexander Wokaun, Richard R. Ernst, Geoffrey Bodenhausen, and Alexander Wokaun. *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*. 1990.

-
- [76] Kanchanok Kodchakorn, Tawan Chokeypaichitkool, and Prachya Kongtawert. Purification and characterisation of heparin-like sulfated polysaccharides with potent anti-SARS-CoV-2 activity from snail mucus of *Achatina fulica*. *Carbohydrate Research*, 529:108832, 2023.
- [77] Zhicheng He, Lutan Zhou, Lisha Lin, Ronghua Yin, and Jinhua Zhao. Structure and heparanase inhibitory activity of a new glycosaminoglycan from the slug *Limacus flavus*. *Carbohydrate Polymers*, 220:176–184, 2019.
- [78] Yvonne Förster, Sabine Schulze, Anja Penk, Christin Neuber, Stephanie Möller, Vera Hintze, Dieter Scharnweber, Matthias Schnabelrauch, Jens Pietzsch, Daniel Huster, and Stefan Rammelt. The influence of different artificial extracellular matrix implant coatings on the regeneration of a critical size femur defect in rats. *Materials Science & Engineering. C, Materials for Biological Applications*, 116:111157, 2020.
- [79] Katie L. Stewart, Eleri Hughes, Edwin A. Yates, Geoffrey R. Akien, Teng-Yi Huang, Marcelo A. Lima, Timothy R. Rudd, Marco Guerrini, Shang-Cheng Hung, Sheena E. Radford, and David A. Middleton. Atomic Details of the Interactions of Glycosaminoglycans with Amyloid- β Fibrils. *Journal of the American Chemical Society*, 138(27):8328–8331, 2016.
- [80] Xiang Zhou, Yuanyuan Wang, Wei Zheng, Guangxiu Deng, Fuyi Wang, and Lan Jin. Characterizing Heparin Tetrasaccharides Binding to Amyloid-Beta Peptide. *Frontiers in Molecular Biosciences*, 9:824146, 2022.
- [81] Kirsten Gade Malmos, Morten Bjerring, Christian Moestrup Jessen, Erik Holm Toustrup Nielsen, Ebbe T. Poulsen, Gunna Christiansen, Thomas Vosegaard, Troels Skrydstrup, Jan J. Enghild, Jan Skov Pedersen, and Daniel E. Otzen. How Glycosaminoglycans Promote Fibrillation of Salmon Calcitonin. *The Journal of Biological Chemistry*, 291(32):16849–16862, 2016.
- [82] David Townsend, Eleri Hughes, Geoffrey Akien, Katie L. Stewart, Sheena E. Radford, David Rochester, and David A. Middleton. Epigallocatechin-3-gallate remodels apolipoprotein A-I amyloid fibrils into soluble oligomers in the presence of heparin. *The Journal of Biological Chemistry*, 293(33):12877–12893, 2018.
- [83] Surabhi Mehra, Dhiman Ghosh, Rakesh Kumar, Mrityunjoy Mondal, Laxmikant G. Gadhe, Subhadeep Das, Arunagiri Anoop, Narendra N. Jha,

- Reeba S. Jacob, Debdeep Chatterjee, Soumik Ray, Nitu Singh, Ashutosh Kumar, and Samir K. Maji. Glycosaminoglycans have variable effects on α -synuclein aggregation and differentially affect the activities of the resulting amyloid fibrils. *The Journal of Biological Chemistry*, 293(34):12975–12991, 2018.
- [84] Line K. Skaanning, Angelo Santoro, Thomas Skamris, Jacob Hertz Martinsen, Anna Maria D’Ursi, Saskia Bucciarelli, Bente Vestergaard, Katrine Bugge, Annette Eva Langkilde, and Birthe B. Kragelund. The Non-Fibrillating N-Terminal of α -Synuclein Binds and Co-Fibrillates with Heparin. *Biomolecules*, 10(8):1192, 2020.
- [85] Kristin Möbius, Karoline Nordsieck, Annelie Pichert, Sergey A. Samsonov, Lars Thomas, Jürgen Schiller, Stefan Kalkhof, M. Teresa Pisabarro, Annette G. Beck-Sickinger, and Daniel Huster. Investigation of lysine side chain interactions of interleukin-8 with heparin and other glycosaminoglycans studied by a methylation-NMR approach. *Glycobiology*, 23(11):1260–1269, 2013.
- [86] Nydia Panitz, Stephan Theisgen, Sergey A. Samsonov, Jan-Philip Gehrcke, Lars Baumann, Kathrin Bellmann-Sickert, Sebastian Köhling, M. Teresa Pisabarro, Jörg Rademann, Daniel Huster, and Annette G. Beck-Sickinger. The structural investigation of glycosaminoglycan binding to CXCL12 displays distinct interaction sites. *Glycobiology*, 26(11):1209–1221, 2016.
- [87] Anja Penk, Lars Baumann, Daniel Huster, and Sergey A. Samsonov. NMR and molecular modeling reveal specificity of the interactions between CXCL14 and glycosaminoglycans. *Glycobiology*, 29(10):715–725, 2019.
- [88] Georg Künze, Sebastian Köhling, Alexander Vogel, Jörg Rademann, and Daniel Huster. Identification of the Glycosaminoglycan Binding Site of Interleukin-10 by NMR Spectroscopy. *The Journal of Biological Chemistry*, 291(6):3100–3113, 2016.
- [89] John E. Scott, Christine Cummings, Helmut Greiling, Helmut W. Stuhlsatz, John D. Gregory, and Shridhar P. Damle. Examination of corneal proteoglycans and glycosaminoglycans by rotary shadowing and electron microscopy. *International Journal of Biological Macromolecules*, 12(3):180–184, 1990.
- [90] Christos Koutsakis, Marco Franchi, Anastasia-Gerasimoula Tavianatou, Valentina Masola, and Nikos K. Karamanos. Studying the Effects of Gly-

- cosaminoglycans in Cell Morphological Aspect with Scanning Electron Microscopy. pages 99–106. 2023.
- [91] Finn P. Maloney, Jeremi Kuklewicz, Robin A. Corey, Yunchen Bi, Ruoya Ho, Lukasz Mateusiak, Els Pardon, Jan Steyaert, Phillip J. Stansfeld, and Jochen Zimmer. Structure, substrate recognition and initiation of hyaluronan synthase. *Nature*, 604(7904):195–201, 2022.
- [92] Takafumi Watanabe, Kiyokazu Kametani, Yoh-ichi Koyama, Daisuke Suzuki, Yasutada Imamura, Kazushige Takehana, and Kohzy Hiramatsu. Ring-Mesh Model of Proteoglycan Glycosaminoglycan Chains in Tendon based on Three-dimensional Reconstruction by Focused Ion Beam Scanning Electron Microscopy*. *Journal of Biological Chemistry*, 291(45):23704–23708, 2016.
- [93] Mario Raspanti, Marina Protasoni, Piero Antonio Zecca, and Marcella Reguzzoni. Slippery when wet: The free surface of the articular cartilage. *Microscopy Research and Technique*, 84(6):1257–1264, 2021.
- [94] C. Przybylski, F. Gonnet, Y. Hersant, D. Bonnaffé, H. Lortat-Jacob, and R. Daniel. Desorption electrospray ionization mass spectrometry of glycosaminoglycans and their protein noncovalent complex. *Analytical Chemistry*, 82(22):9225–9233, 2010.
- [95] Nadezda Kiselova, Tabea Dierker, Dorothe Spillmann, and Margareta Ramström. An automated mass spectrometry-based screening method for analysis of sulfated glycosaminoglycans. *Biochemical and Biophysical Research Communications*, 450(1):598–603, 2014.
- [96] Nataniel Ludwig, Francyne Kubaski, Fernanda Sperb-Ludwig, Maira Graeff Burin, Thiago Oliveira Silva, Fabiano de Oliveira Poswar, Carolina Fischinger Moura Souza, Roberto Giugliani, and Ida V. D. Schwartz. Quantification of glycosaminoglycans by liquid chromatography tandem mass spectrometry is a useful tool for screening of GlcNAc-phosphotransferase deficient patients. *Molecular Genetics and Metabolism*, 129(2):S102, 2020.
- [97] Tommy Hofmann, Sergey A. Samsonov, Annelie Pichert, Katharina Lemm-nitzer, Jürgen Schiller, Daniel Huster, M. Teresa Pisabarro, Martin von Bergen, and Stefan Kalkhof. Structural analysis of the interleukin-8/glycosaminoglycan interactions by amide hydrogen/deuterium exchange mass spectrometry. *Methods*, 89:45–53, 2015.

- [98] Aiye Liang and Umesh Desai. Advances in Studying Glycosaminoglycan-ProteinProteinsInteractions Using Capillary ElectrophoresisCapillary electrophoresis (CE). pages 365–387. 2022.
- [99] Niels H. H. Heegaard, Hanne D. Mortensen, and Peter Roepstorff. Demonstration of a heparin-binding site in serum amyloid P component using affinity capillary electrophoresis as an adjunct technique. *Journal of Chromatography A*, 717(1):83–90, 1995.
- [100] Niels H. H. Heegaard and Frank A. Robey. Use of capillary zone electrophoresis to evaluate the binding of anionic carbohydrates to synthetic peptides derived from human serum amyloid P component. *Analytical Chemistry*, 64(21):2479–2482, 1992.
- [101] Maria J. Hernaiz, Laurie A. LeBrun, Yi Wu, Jette W. Sen, Robert J. Linhardt, and Niels H. H. Heegaard. Characterization of heparin binding by a peptide from amyloid P component using capillary electrophoresis, surface plasmon resonance and isothermal titration calorimetry. *European Journal of Biochemistry*, 269(12):2860–2867, 2002.
- [102] Toshiaki Hattori, Kozue Kimura, Emek Seyrek, and Paul L. Dubin. Binding of Bovine Serum Albumin to Heparin Determined by Turbidimetric Titration and Frontal Analysis Continuous Capillary Electrophoresis. *Analytical Biochemistry*, 295(2):158–167, 2001.
- [103] Niels H. H. Heegaard, Peter Roepstorff, Steen G. Melberg, and Mogens H. Nissen. Cleaved β 2-Microglobulin Partially Attains a Conformation That Has Amyloidogenic Features*. *Journal of Biological Chemistry*, 277(13):11184–11189, 2002.
- [104] A. J. Miles, Robert W. Janes, and B. A. Wallace. Tools and methods for circular dichroism spectroscopy of proteins: a tutorial review. *Chemical Society Reviews*, 50(15):8400–8413, 2021.
- [105] Koichi Matsuo and Kunihiro Gekko. Synchrotron-Radiation Vacuum-Ultraviolet Circular-Dichroism Spectroscopy for Characterizing the Structure of Saccharides. pages 101–117. 2018.
- [106] Bireswar Chakrabarti. Carboxyl and Amide Transitions in the Circular Dichroism of Glycosaminoglycans. volume 150, pages 275–292. 1981.

-
- [107] Audrey L. Stone, David Beeler, Gary Oosta, and Robert D. Rosenberg. Circular dichroism spectroscopy of heparin-antithrombin interactions. *Proceedings of the National Academy of Sciences*, 79(23):7190–7194, 1982.
- [108] Ferenc Zsila. Glycosaminoglycans are potential pharmacological targets for classic DNA minor groove binder drugs berenil and pentamidine. *Physical Chemistry Chemical Physics*, 17(38):24560–24565, 2015.
- [109] Ferenc Zsila, Sergey A. Samsonov, and Martyna Maszota-Zieleniak. Mind Your Dye: The Amyloid Sensor Thioflavin T Interacts with Sulfated Glycosaminoglycans Used To Induce Cross- β -Sheet Motifs. *The Journal of Physical Chemistry B*, 124(51):11625–11633, 2020.
- [110] Martyna Maszota-Zieleniak, Ferenc Zsila, and Sergey A. Samsonov. Computational insights into heparin-small molecule interactions: Evaluation of the balance between stacking and non-stacking binding modes. *Carbohydrate Research*, 507:108390, 2021.
- [111] Lakkoji Satish, Santanu Santra, Mikhail V. Tsurkan, Carsten Werner, Madhurima Jana, and Harekrushna Sahoo. Conformational changes of GDNF-derived peptide induced by heparin, heparan sulfate, and sulfated hyaluronic acid – Analysis by circular dichroism spectroscopy and molecular dynamics simulation. *International Journal of Biological Macromolecules*, 182:2144–2150, 2021.
- [112] Akira Ogamo, Kimiyo Matsuzaki, Hideki Uchiyama, and Kinzo Nagasawa. Preparation and properties of fluorescent glycosamino-glycuronans labeled with 5-aminofluorescein. *Carbohydrate Research*, 105(1):69–85, 1982.
- [113] Annelie Pichert, Sergey A. Samsonov, Stephan Theisgen, Lars Thomas, Lars Baumann, Jürgen Schiller, Annette G. Beck-Sickinger, Daniel Huster, and M. Teresa Pisabarro. Characterization of the interaction of interleukin-8 with hyaluronan, chondroitin sulfate, dermatan sulfate and their sulfated derivatives by spectroscopy and molecular modeling. *Glycobiology*, 22(1):134–145, 2012.
- [114] Ray Bakhtiar. Surface plasmon resonance spectroscopy: A versatile technique in a biochemist’s toolbox. *Journal of Chemical Education*, 90(2):203–209, 2013.

- [115] A. Satoh, T. Toida, K. Yoshida, K. Kojima, and I. Matsumoto. New role of glycosaminoglycans on the plasma membrane proposed by their interaction with phosphatidylcholine. *FEBS letters*, 477(3):249–252, 2000.
- [116] K. Gaus and E. A. Hall. Surface Plasmon Resonance Measurement of the Binding of Low-Density Lipoprotein at a Heparin Surface. *Journal of Colloid and Interface Science*, 217(1):111–118, 1999.
- [117] Juliette Sage, Florian Mallèvre, Fabien Barbarin-Costes, Sergey A. Samsonov, Jan-Philip Gehrcke, Maria Teresa Pisabarro, Eric Perrier, Sylvianne Schnebert, André Roget, Thierry Livache, Carine Nizard, Gilles Lalmanach, and Fabien Lecaillon. Binding of chondroitin 4-sulfate to cathepsin S regulates its enzymatic activity. *Biochemistry*, 52(37):6487–6498, 2013.
- [118] Yota Tatara, Shinichiro Suto, and Ken Itoh. Novel roles of glycosaminoglycans in the degradation of type I collagen by cathepsin K. *Glycobiology*, 27(12):1089–1098, 2017.
- [119] Wuxia Zhang, Haibo Mu, Amin Zhang, Guoting Cui, Hua Chen, Jinyou Duan, and Shunchun Wang. A decrease in moisture absorption-retention capacity of N-deacetylation of hyaluronic acid. *Glycoconjugate Journal*, 30(6):577–583, 2013.
- [120] Courtney J. Mycroft-West, Dunhao Su, Isabel Pagani, Timothy R. Rudd, Stefano Elli, Neha S. Gandhi, Scott E. Guimond, Gavin J. Miller, Maria C. Z. Meneghetti, Helena B. Nader, Yong Li, Quentin M. Nunes, Patricia Procter, Nicasio Mancini, Massimo Clementi, Antonella Bisio, Nicholas R. Forsyth, Vito Ferro, Jeremy E. Turnbull, Marco Guerrini, David G. Fernig, Elisa Vicenzi, Edwin A. Yates, Marcelo A. Lima, and Mark A. Skidmore. Heparin Inhibits Cellular Invasion by SARS-CoV-2: Structural Dependence of the Interaction of the Spike S1 Receptor-Binding Domain with Heparin. *Thrombosis and Haemostasis*, 120(12):1700–1715, 2020.
- [121] So Young Kim, Fuming Zhang, David A. Harris, and Robert J. Linhardt. Structural Features of Heparin and Its Interactions With Cellular Prion Protein Measured by Surface Plasmon Resonance. *Frontiers in Molecular Biosciences*, 7:594497, 2020.
- [122] Deling Shi, Peng He, Yuefan Song, Shuihong Cheng, Robert J. Linhardt, Jonathan S. Dordick, Lianli Chi, and Fuming Zhang. Kinetic and Structural Aspects of Glycosaminoglycan-Monkeypox Virus Protein A29 Inter-

- actions Using Surface Plasmon Resonance. *Molecules (Basel, Switzerland)*, 27(18):5898, 2022.
- [123] Peng He, Deling Shi, Yunran Li, Ke Xia, Seon Beom Kim, Rohini Dwivedi, Marwa Farrag, Vitor H. Pomin, Robert J. Linhardt, Jonathan S. Dordick, and Fuming Zhang. SPR Sensor-Based Analysis of the Inhibition of Marine Sulfated Glycans on Interactions between Monkeypox Virus Proteins and Glycosaminoglycans. *Marine Drugs*, 21(5):264, 2023.
- [124] Deling Shi, Peng He, Yuefan Song, Robert J. Linhardt, Jonathan S. Dordick, Lianli Chi, and Fuming Zhang. Interactions of heparin with key glycoproteins of human respiratory syncytial virus. *Frontiers in Molecular Biosciences*, 10:1151174, 2023.
- [125] F. Lustig, J. Hoebeke, G. Ostergren-Lundèn, F. Velge-Roussel, G. Bondjers, U. Olsson, U. Rüetschi, and G. Fager. Alternative splicing determines the binding of platelet-derived growth factor (PDGF-AA) to glycosaminoglycans. *Biochemistry*, 35(37):12077–12085, 1996.
- [126] Tero Pihlajamaa, Hilka Lankinen, Joni Ylöstalo, Leena Valmu, Juha Jäälinoja, Frank Zaucke, Luitgard Spitznagel, Silke Gösling, Anne Puustinen, Matthias Mörgelin, Johan Peränen, Patrik Maurer, Leena Ala-Kokko, and Ilkka Kilpeläinen. Characterization of recombinant amino-terminal NC4 domain of human collagen IX: interaction with glycosaminoglycans and cartilage oligomeric matrix protein. *The Journal of Biological Chemistry*, 279(23):24265–24273, 2004.
- [127] Fuming Zhang, Heather A. Moniz, Benjamin Walcott, Kelley W. Moremen, Lianchun Wang, and Robert J. Linhardt. Probing the impact of GFP tagging on Robo1-heparin interaction. *Glycoconjugate Journal*, 31(4):299–307, 2014.
- [128] Jamie C. Fox, Robert C. Tyler, Francis C. Peterson, Douglas P. Dyer, Fuming Zhang, Robert J. Linhardt, Tracy M. Handel, and Brian F. Volkman. Examination of Glycosaminoglycan Binding Sites on the XCL1 Dimer. *Biochemistry*, 55(8):1214–1225, 2016.
- [129] Fuming Zhang, Jing Zhao, Xinyue Liu, and Robert J. Linhardt. Interactions between Sclerostin and Glycosaminoglycans. *Glycoconjugate Journal*, 37(1):119–128, 2020.

- [130] Luna Dillemans, Karen Yu, Alexandra De Zutter, Sam Noppen, Mieke Gouwy, Nele Berghmans, Lisa Verhallen, Mirre De Bondt, Lotte Vanbraabant, Stef Brusselmans, Erik Martens, Dominique Schols, Patrick Verschueren, Mette M. Rosenkilde, Pedro Elias Marques, Sofie Struyf, and Paul Proost. Natural carboxyterminal truncation of human CXCL10 attenuates glycosaminoglycan binding, CXCR3A signaling and lymphocyte chemotaxis, while retaining angiostatic activity. *Cell communication and signaling: CCS*, 22(1):94, 2024.
- [131] Dominic W. P. Collis, Gokhan Yilmaz, Yichen Yuan, Alessandra Monaco, Guy Ochbaum, Yejiao Shi, Clare O'Malley, Veselina Uzunova, Richard Napier, Ronit Bitton, C. Remzi Becer, and Helena S. Azevedo. Hyaluronan (HA)-inspired glycopolymers as molecular tools for studying HA functions. *RSC chemical biology*, 2(2):568–576, 2021.
- [132] V. B. Lokeshwar, C. Obek, H. T. Pham, D. Wei, M. J. Young, R. C. Duncan, M. S. Soloway, and N. L. Block. Urinary hyaluronic acid and hyaluronidase: markers for bladder cancer detection and evaluation of grade. *The Journal of Urology*, 163(1):348–356, 2000.
- [133] J. M. Williams, R. J. Katz, D. Childs, M. E. Lenz, and E. J. Thonar. Keratan sulfate content in the superficial and deep layers of osteophytic and nonfibrillated human articular cartilage in osteoarthritis. *Calcified Tissue International*, 42(3):162–166, 1988.
- [134] E. M. O'Byrne, H. C. Schroder, C. Stefano, and R. L. Goldberg. Catabolin/interleukin-1 regulation of cartilage and chondrocyte metabolism. *Agents and Actions*, 21(3-4):341–344, 1987.
- [135] Kornelia Kuźnik-Trocha, Katarzyna Winsz-Szczotka, Katarzyna Komosińska-Vassev, Agnieszka Jura-Półtorak, Anna Kotulska-Kucharz, Eugeniusz J. Kucharz, Przemysław Kotyla, and Krystyna Olczyk. Plasma Glycosaminoglycan Profiles in Systemic Sclerosis: Associations with MMP-3, MMP-10, TIMP-1, TIMP-2, and TGF-Beta. *BioMed Research International*, 2020:6416514, 2020.
- [136] Ewa M. Koźma, Kornelia Kuźnik-Trocha, Katarzyna Winsz-Szczotka, Grzegorz Wisowski, Paweł Olczyk, Katarzyna Komosińska-Vassev, Mariusz Kasperczyk, and Krystyna Olczyk. Significant Remodeling Affects the Circulating Glycosaminoglycan Profile in Adult Patients with both Severe and

- Mild Forms of Acute Pancreatitis. *Journal of Clinical Medicine*, 9(5):1308, 2020.
- [137] A. A. Capehart, C. H. Mjaatvedt, S. Hoffman, and E. L. Krug. Dynamic expression of a native chondroitin sulfate epitope reveals microheterogeneity of extracellular matrix organization in the embryonic chick heart. *The Anatomical Record*, 254(2):181–195, 1999.
- [138] Melody Liles, Barbara P. Palka, Anthony Harris, Briedgeen Kerr, Clare Hughes, Robert D. Young, Keith M. Meek, Bruce Caterson, and Andrew J. Quantock. Differential relative sulfation of Keratan sulfate glycosaminoglycan in the chick cornea during embryonic development. *Investigative Ophthalmology & Visual Science*, 51(3):1365–1372, 2010.
- [139] Róbert Agócs, Domonkos Pap, Dániel Sugár, Gábor Tóth, Lilla Turiák, Zoltán Veréb, Lajos Kemény, Tivadar Tulassay, Ádám Vannay, and Attila J. Szabó. Cyclooxygenase-2 Modulates Glycosaminoglycan Production in the Skin During Salt Overload. *Frontiers in Physiology*, 11:561722, 2020.
- [140] Wangping Duan, Yu Zhao, Xiaochun Ren, Ruipeng Zhao, Qi Li, Zhenwei Sun, Wenjie Song, Yanfei Yang, Pengcui Li, and Xiaochun Wei. Combination of chondrocytes and chondrons improves extracellular matrix production to promote the repairs of defective knee cartilage in rabbits. *Journal of Orthopaedic Translation*, 28:47–54, 2021.
- [141] Stefanie Kowarschik, Julian Schöllkopf, Thomas Müller, Songhai Tian, Julian Knerr, Hans Bakker, Stephan Rein, Min Dong, Stefan Weber, Robert Grosse, and Gudula Schmidt. Yersinia pseudotuberculosis cytotoxic necrotizing factor interacts with glycosaminoglycans. *FASEB journal: official publication of the Federation of American Societies for Experimental Biology*, 35(7):e21647, 2021.
- [142] Cédric Laguri, Rabia Sadir, Evelyne Gout, Romain R. Vivès, and Hugues Lortat-Jacob. Preparation and Characterization of Heparan Sulfate-Derived Oligosaccharides to Investigate Protein-GAG Interaction and HS Biosynthesis Enzyme Activity. *Methods in Molecular Biology (Clifton, N.J.)*, 2303:121–137, 2022.
- [143] Ina Coburger, Sven O. Dahms, Dirk Roeser, Karl-Heinz Gührs, Peter Hortschansky, and Manuel E. Than. Analysis of the overall structure of the

- multi-domain amyloid precursor protein (APP). *PloS One*, 8(12):e81926, 2013.
- [144] N Sharma, L Haggstrom, S Sohrabipour, DJ Dwivedi, and PC Liaw. Investigations of the effectiveness of heparin variants as inhibitors of histones. *Journal of thrombosis and haemostasis : JTH*, 20(6), 2022.
- [145] Sean P. Giblin, Sashini Ranawana, Shyreen Hassibi, Holly L. Birchenough, Kyle T. Mincham, Robert J. Snelgrove, Tomoko Tsuchiya, Shiro Kanegasaki, Douglas Dyer, and James E. Pease. CXCL17 binds efficaciously to glycosaminoglycans with the potential to modulate chemokine signaling. *Frontiers in Immunology*, 14:1254697, 2023.
- [146] Christopher M. Johnson. Isothermal Titration Titrations Calorimetry. pages 135–159. 2021.
- [147] R. Tylercros, M. Sobel, D. Marques, D. F. Soler, and R. B. Harris. Heparin-von Willebrand Factor Binding as Assessed by Isothermal Titration Calorimetry and by Affinity Fractionation of Heparins Using Synthetic Peptides. *Archives of Biochemistry and Biophysics*, 306(2):528–533, 1993.
- [148] Amit K. Dutta, Krishna Mohan Sepuru, Jörg Rösger, and Krishna Rajarathnam. Characterizing Thermodynamics of Protein-Glycosaminoglycan Interactions Using Isothermal Titration Calorimetry. *Methods in Molecular Biology (Clifton, N.J.)*, 2303:307–317, 2022.
- [149] R. E. Hileman, R. N. Jennings, and R. J. Linhardt. Thermodynamic analysis of the heparin interaction with a basic cyclic peptide using isothermal titration calorimetry. *Biochemistry*, 37(43):15231–15237, 1998.
- [150] Javier Pérez Quiñones, Oliver Brüggemann, Carlos Peniche Covas, and Dmitri A. Ossipov. Self-assembled hyaluronic acid nanoparticles for controlled release of agrochemicals and diosgenin. *Carbohydrate Polymers*, 173:157–169, 2017.
- [151] Alfonso Maria Ponsiglione, Maria Russo, and Enza Torino. Glycosaminoglycans and Contrast Agents: The Role of Hyaluronic Acid as MRI Contrast Enhancer. *Biomolecules*, 10(12):1612, 2020.
- [152] J. R. Fromm, R. E. Hileman, J. M. Weiler, and R. J. Linhardt. Interaction of fibroblast growth factor-1 and related peptides with heparan sulfate and

- its oligosaccharides. *Archives of Biochemistry and Biophysics*, 346(2):252–262, 1997.
- [153] Jared L. Edwards, Priyanka D. Kadav, Purnima Bandyopadhyay, and Tarun K. Dam. Revealing the Identity of Human Galectin-3 as a Glycosaminoglycan-Binding Protein. *Methods in Molecular Biology (Clifton, N.J.)*, 2442:137–150, 2022.
- [154] Krishna Mohan Sepuru, Balaji Nagarajan, Umesh R. Desai, and Krishna Rajarathnam. Molecular Basis of Chemokine CXCL5-Glycosaminoglycan Interactions. *The Journal of Biological Chemistry*, 291(39):20539–20550, 2016.
- [155] Eva Muñoz and Juan Sabín. The Use of ITC and the Software AFFINImeter for the Quantification of the Anticoagulant Pentasaccharide in Low Molecular Weight Heparin. pages 215–223. 2019.
- [156] Luca M. Lauth, Bruno Voigt, Twinkle Bhatia, Lisa Machner, Jochen Balbach, and Maria Ott. Heparin promotes rapid fibrillation of the basic parathyroid hormone at physiological pH. *FEBS Letters*, 596(22):2928–2939, 2022.
- [157] Iuliia Pilipenko, Viktor Korzhikov-Vlakh, Annika Valtari, Yurii Anufrikov, Stanislav Kalinin, Marika Ruponen, Mikhail Krasavin, Arto Urtti, and Tatiana Tennikova. Mucoadhesive properties of nanogels based on stimuli-sensitive glycosaminoglycan-graft-pNIPAAm copolymers. *International Journal of Biological Macromolecules*, 186:864–872, 2021.
- [158] Weronika Malicka, Rainer Haag, and Matthias Ballauff. Interaction of Heparin with Proteins: Hydration Effects. *The Journal of Physical Chemistry B*, 126(33):6250–6260, 2022.
- [159] Edyta Adrian, Dušana Treľová, Elena Filová, Marta Kumorek, Volodymyr Lobaz, Rafal Poreba, Olga Janoušková, Ognjen Pop-Georgievski, Igor Lacík, and Dana Kubies. Complexation of CXCL12, FGF-2 and VEGF with Heparin Modulates the Protein Release from Alginate Microbeads. *International Journal of Molecular Sciences*, 22(21):11666, 2021.
- [160] Carole Anne De Carufel, Phuong Trang Nguyen, Sabrina Sahnouni, and Steve Bourgault. New insights into the roles of sulfated glycosaminoglycans in islet amyloid polypeptide amyloidogenesis and cytotoxicity. *Biopolymers*, 100(6):645–655, 2013.

- [161] Ishfaq Ahmad Ahanger, Zahoor Ahmad Parray, Khalida Nasreen, Faizan Ahmad, Md Imtaiyaz Hassan, Asimul Islam, and Anurag Sharma. Heparin Accelerates the Protein Aggregation via the Downhill Polymerization Mechanism: Multi-Spectroscopic Studies to Delineate the Implications on Proteinopathies. *ACS omega*, 6(3):2328–2339, 2021.
- [162] Abu Hamza, Abdus Samad, Zahoor Ahmad Parray, Sajda Ara, Anwar Ahmed, Fahad N. Almajhdi, Tajamul Hussain, Asimul Islam, and Shama Parveen. Mutation in the CX3C Motif of G Protein Disrupts Its Interaction with Heparan Sulfate: A Calorimetric, Spectroscopic, and Molecular Docking Study. *International Journal of Molecular Sciences*, 23(4):1950, 2022.
- [163] Fanny Guillaumie, Pascal Furrer, Olivia Felt-Baeyens, Birgit L. Fuhlen-dorff, Søren Nyman, Peter Westh, Robert Gurny, and Khadija Schwach-Abdellaoui. Comparative studies of various hyaluronic acids produced by microbial fermentation for potential topical ophthalmic applications. *Journal of Biomedical Materials Research. Part A*, 92(4):1421–1430, 2010.
- [164] Fatma Krichen, Zohra Ghilissi, Rihab Ben Abdallah, Rim Kallel, Oscar Martinez-Alvarez, M. Carmen Gómez-Guillén, Assaad Sila, Tahia Boudawara, Zouheir Sahnoun, and Ali Bougatef. Glycosaminoglycans from grey triggerfish and smooth hound skins: Rheological, Anti-inflammatory and wound healing properties. *International Journal of Biological Macromolecules*, 118(Pt A):965–975, 2018.
- [165] Rachael Floreani, Cody N. Cranson, and Susan P. James. Synthesis and characterization of a Hyaluronan-polyethylene copolymer for biomedical applications. *Journal of Biomedical Materials Research. Part B, Applied Biomaterials*, 94(2):441–446, 2010.
- [166] Jae-Kyung Kim, Jung-Soo Lee, Hyo-Jung Jung, Jin-Hui Cho, Jee-In Heo, and Yoon-Ho Chang. Preparation and properties of collagen/modified hyaluronic acid hydrogel for biomedical application. *Journal of Nanoscience and Nanotechnology*, 7(11):3852–3856, 2007.
- [167] Piyachat Chuysinuan, Thanyaluck Thanyacharoen, Kitiyaporn Thongchai, Supanna Techasakul, and Sarute Ummartyotin. Preparation of chitosan/hydrolyzed collagen/hyaluronic acid based hydrogel composite with caffeic acid addition. *International Journal of Biological Macromolecules*, 162:1937–1943, 2020.

- [168] Amit Khurana, Anil Kumar Banothu, A. V. Thanusha, Aradhana Nayal, Amit Kumar Dinda, Maneesh Singhal, Kala Kumar Bharani, and Veena Koul. Preclinical efficacy study of a porous biopolymeric scaffold based on gelatin-hyaluronic acid-chondroitin sulfate in a porcine burn injury model: role of critical molecular markers (VEGFA, N-cadherin, COX-2), gamma sterilization efficacy and a comparison of healing potential to IntegraTM. *Biomedical Materials (Bristol, England)*, 16(5), 2021.
- [169] Ulla König, Anja Lode, Petra B. Welzel, Yuichiro Ueda, Sven Knaack, Anja Henß, Anke Hauswald, and Michael Gelinsky. Heparinization of a biomimetic bone matrix: integration of heparin during matrix synthesis versus adsorptive post surface modification. *Journal of Materials Science. Materials in Medicine*, 25(3):607–621, 2014.
- [170] Mark van Beek, Andrea Weeks, Lyndon Jones, and Heather Sheardown. Immobilized hyaluronic acid containing model silicone hydrogels reduce protein adsorption. *Journal of Biomaterials Science. Polymer Edition*, 19(11):1425–1436, 2008.
- [171] Muhammad Suhail, I.-Hui Chiu, Ming-Chia Hung, Quoc Lam Vu, I.-Ling Lin, and Pao-Chu Wu. In Vitro Evaluation of Smart and pH-Sensitive Chondroitin Sulfate/Sodium Polystyrene Sulfonate Hydrogels for Controlled Drug Delivery. *Gels (Basel, Switzerland)*, 8(7):406, 2022.
- [172] Zhipeng Chen, Juan Chen, Li Wu, Weidong Li, Jun Chen, Haibo Cheng, Jinhua Pan, and Baochang Cai. Hyaluronic acid-coated bovine serum albumin nanoparticles loaded with brucine as selective nanovectors for intra-articular injection. *International Journal of Nanomedicine*, 8:3843–3853, 2013.
- [173] Yu-Xin Liu, Da-Yong Zhou, Dong-Dong Ma, Zi-Qiang Liu, Yan-Fei Liu, Liang Song, Xiu-Ping Dong, Dong-Mei Li, Bei-Wei Zhu, Kunihiko Konno, and Fereidoon Shahidi. Effects of endogenous cysteine proteinases on structures of collagen fibres from dermis of sea cucumber (*Stichopus japonicus*). *Food Chemistry*, 232:10–18, 2017.
- [174] J. M. McPherson, S. J. Sawamura, R. A. Condell, W. Rhee, and D. G. Wallace. The effects of heparin on the physicochemical properties of reconstituted collagen. *Collagen and Related Research*, 8(1):65–82, 1988.

-
- [175] G. M. Cherchi, M. Formato, P. Demuro, M. Masserini, I. Varani, and G. DeLuca. Modifications of low density lipoprotein induced by the interaction with human plasma glycosaminoglycan-protein complexes. *Biochimica Et Biophysica Acta*, 1212(3):345–352, 1994.
- [176] M. Bihari-Varga, J. Sztatisz, and S. Gál. Changes in the physical behavior of low density lipoprotein in the presence of glycosaminoglycans and high density lipoprotein. *Atherosclerosis*, 39(1):19–23, 1981.
- [177] John David Jackson. *Classical Electrodynamics*. 1998.
- [178] David J. Griffiths. *Introduction to Electrodynamics*. 2012.
- [179] Sergey A. Samsonov and M. Teresa Pisabarro. Computational analysis of interactions in structurally available protein-glycosaminoglycan complexes. *Glycobiology*, 26(8):850–861, 2016.
- [180] Aurijit Sarkar and Umesh R. Desai. A Simple Method for Discovering Druggable, Specific Glycosaminoglycan-Protein Systems. Elucidation of Key Principles from Heparin/Heparan Sulfate-Binding Proteins. *PloS One*, 10(10):e0141127, 2015.
- [181] M.P. McGee and Liang Liang. Regulation of glycosaminoglycan function by osmotic potentials. Measurement of water transfer during antithrombin activation by heparin. *The Journal of biological chemistry*, 276(52), 2001.
- [182] K. R. Grymonpré, B. A. Staggemeier, P. L. Dubin, and K. W. Mattison. Identification by integrated computer modeling and light scattering studies of an electrostatic serum albumin-hyaluronic acid binding site. *Biomacromolecules*, 2(2):422–429, 2001.
- [183] Lei Wang and Yu Zhao. Quantum chemical electrostatic potential surfaces: An overview and application to biomolecules. *International Journal of Quantum Chemistry*, 118(10):e25538, 2018.
- [184] Takeshi Ishikawa and Yuriko Aoki. Ab initio calculations of electrostatic potentials in biomolecular fragments. *Theoretical Chemistry Accounts*, 137(1):10, 2018.
- [185] Ahmed Badran. A dft investigation of the electrostatic potential distribution in beryllium compounds. *Chemical Physics Letters*, 817:140396, 2023.

- [186] Mohamed E. Elshakre. Density functional theory-based ep maps in halogenated systems. *Journal of Molecular Modeling*, 26:240, 2020.
- [187] Mohammed Majrashi and Abdullah Almutairi. Dft analysis of sulfated carbohydrates for biomedical applications. *Computational and Theoretical Chemistry*, 1221:114954, 2024.
- [188] Jana Khandogin and Charles L. Brooks III. Quantum-based semiempirical potentials for protein simulations. *Journal of Computational Chemistry*, 25(11):1400–1413, 2004.
- [189] David A. Case, Thomas E. Cheatham III, et al. The amber biomolecular simulation package: Advances and applications. *Journal of Computational Chemistry*, 2024.
- [190] Olgun Guvench and Alexander D. MacKerell Jr. Charmm additive all-atom force field for glycosidic linkages. *Journal of Chemical Theory and Computation*, 7(10):3162–3180, 2011.
- [191] Douglas Pimentel and Rodrigo A. Lopes. Validation of theoretical representations of ep surfaces using poisson–boltzmann approaches. *Journal of Chemical Information and Modeling*, 61(3):1039–1047, 2021.
- [192] Francesco Cavani and Rosa Maria C. Di Martino. Molecular modeling in protein–ligand complexes: Advances and pitfalls. *Current Opinion in Chemical Biology*, 66:102086, 2022.
- [193] Nathan A. Baker, David Sept, Simpson Joseph, Michael J. Holst, and J. Andrew McCammon. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences*, 98(18):10037–10041, 2001.
- [194] Steven T. Olson, Benjamin Richard, Gonzalo Izaguirre, Sophia Schedin-Weiss, and Peter G. W. Gettins. Molecular mechanisms of antithrombin-heparin regulation of blood clotting proteinases. A paradigm for understanding proteinase regulation by serpin family protein proteinase inhibitors. *Biochimie*, 92(11):1587–1596, 2010.
- [195] Sophia Schedin-Weiss, Véronique Arocas, Susan C. Bock, Steven T. Olson, and Ingemar Björk. Specificity of the basic side chains of Lys114, Lys125, and Arg129 of antithrombin in heparin binding. *Biochemistry*, 41(41):12369–12376, 2002.

- [196] Annelie Pichert, Denise Schlorke, Sandra Franz, and Juergen Arnhold. Functional aspects of the interaction between interleukin-8 and sulfated glycosaminoglycans. *Biomatter*, 2(3):142–148, 2012.
- [197] Prem Raj B. Joseph, Philip D. Mosier, Umesh R. Desai, and Krishna Rajarathnam. Solution NMR characterization of chemokine CXCL8/IL-8 monomer and dimer binding to glycosaminoglycans: structural plasticity mediates differential binding interactions. *The Biochemical Journal*, 472(1):121–133, 2015.
- [198] G. Cilpa, M. T. Hyvönen, A. Koivuniemi, and M.-L. Riekkola. Atomistic insight into chondroitin-6-sulfate glycosaminoglycan chain through quantum mechanics calculations and molecular dynamics simulation. *Journal of Computational Chemistry*, 31(8):1670–1680, 2010.
- [199] Andrew T. McNutt, Yanjing Li, Rocco Meli, Rishal Aggarwal, and David Ryan Koes. GNINA 1.3: the next increment in molecular docking with deep learning. *Journal of Cheminformatics*, 17(1):28, 2025.
- [200] Gabriele Corso, Hannes Stärk, Bowen Jing, Regina Barzilay, and Tommi Jaakkola. DiffDock: Diffusion Steps, Twists, and Turns for Molecular Docking, 2023.
- [201] Hannes Stärk, Octavian-Eugen Ganea, Lagnajit Pattanaik, Regina Barzilay, and Tommi Jaakkola. EquiBind: Geometric Deep Learning for Drug Binding Structure Prediction, 2022. arXiv:2202.05146 [q-bio].
- [202] Wei Lu, Qifeng Wu, Jixian Zhang, Jiahua Rao, Chengtao Li, and Shuangjia Zheng. TANKBind: Trigonometry-Aware Neural Networks for Drug-Protein Binding Structure Prediction, 2022. Pages: 2022.06.06.495043 Section: New Results.
- [203] Hilbert Yuen In Lam, Jia Sheng Guan, Xing Er Ong, Robbe Pincket, and Yuguang Mu. Protein language models are performant in structure-free virtual screening. *Brief Bioinform*, 25(6):bbae480, 2024.
- [204] Mateusz Zalewski, Björn Wallner, and Sebastian Kmiecik. Protein–Peptide Docking with ESMFold Language Model. *J. Chem. Theory Comput.*, 21(6):2817–2821, 2025.

- [205] Sergey A. Samsonov, Stephan Theisgen, Thomas Riemer, Daniel Huster, and M. Teresa Pisabarro. Glycosaminoglycan monosaccharide blocks analysis by quantum mechanics, molecular dynamics, and nuclear magnetic resonance. *BioMed Research International*, 2014:808071, 2014.
- [206] Sergey A. Samsonov, Martin Zacharias, and Isaure Chauvot de Beauchene. Modeling large protein-glycosaminoglycan complexes using a fragment-based approach. *J Comput Chem*, 40(14):1429–1439, 2019.
- [207] Till Siebenmorgen, Michael Engelhard, and Martin Zacharias. Prediction of protein-protein complexes using replica exchange with repulsive scaling. *J Comput Chem*, 41(15):1436–1447, 2020.
- [208] Till Siebenmorgen and Martin Zacharias. Efficient Refinement and Free Energy Scoring of Predicted Protein-Protein Complexes Using Replica Exchange with Repulsive Scaling. *J Chem Inf Model*, 60(11):5552–5562, 2020.
- [209] Mateusz Marcisz, Martyna Maszota-Zieleniak, and Sergey A. Samsonov. Repulsive Scaling Replica Exchange Molecular Dynamics in Modeling Protein-Glycosaminoglycan Complexes. *Methods Mol Biol*, 2619:153–167, 2023.
- [210] Mateusz Marcisz, Martyna Maszota-Zieleniak, Bertrand Huard, and Sergey A. Samsonov. Advanced Molecular Dynamics Approaches to Model a Tertiary Complex APRIL/TACI with Long Glycosaminoglycans. *Biomolecules*, 11(9):1349, 2021.
- [211] Martyna Maszota-Zieleniak, Mateusz Marcisz, Małgorzata M. Kogut, Till Siebenmorgen, Martin Zacharias, and Sergey A. Samsonov. Evaluation of replica exchange with repulsive scaling approach for docking glycosaminoglycans. *J Comput Chem*, 42(15):1040–1053, 2021.
- [212] Mateusz Marcisz, Margrethe Gaardløs, Krzysztof K. Bojarski, Till Siebenmorgen, Martin Zacharias, and Sergey A. Samsonov. Explicit solvent repulsive scaling replica exchange molecular dynamics (RS-REMD) in molecular modeling of protein-glycosaminoglycan complexes. *J Comput Chem*, 43(24):1633–1640, 2022.
- [213] Neha S. Gandhi and Ricardo L. Mancera. Prediction of heparin binding sites in bone morphogenetic proteins (BMPs). *Biochimica Et Biophysica Acta*, 1824(12):1374–1381, 2012.

-
- [214] Nicolas Sapay, Eric Cabannes, Maurice Petitou, and Anne Imberty. Molecular modeling of the interaction between heparan sulfate and cellular growth factors: bringing pieces together. *Glycobiology*, 21(9):1181–1193, 2011.
- [215] Urszula Uciechowska-Kaczmarzyk, Sándor Babik, Ferenc Zsila, Krzysztof Kamil Bojarski, Tamás Beke-Somfai, and Sergey A. Samsonov. Molecular dynamics-based model of VEGF-A and its heparin interactions. *Journal of Molecular Graphics and Modelling*, 82:157–166, 2018.
- [216] Giovanni Battimelli and Giovanni Ciccotti. Berni Alder and the pioneering times of molecular simulation. *The European Physical Journal H*, 43(3):303–335, 2018.
- [217] Vincent E. Lamberti, Lloyd D. Fosdick, Elizabeth R. Jessup, and Carolyn J. C. Schauble. A Hands-On Introduction to Molecular Dynamics. *Journal of Chemical Education*, 79(5):601, 2002.
- [218] Michel A. Cuendet and Wilfred F. van Gunsteren. On the calculation of velocity-dependent properties in molecular dynamics simulations using the leapfrog integration algorithm. *The Journal of Chemical Physics*, 127(18):184102, 2007.
- [219] Matej Praprotnik and Dusanka Janezic. Molecular dynamics integration meets standard theory of molecular vibrations. *Journal of Chemical Information and Modeling*, 45(6):1571–1579, 2005.
- [220] Jaewoon Jung, Chigusa Kobayashi, and Yuji Sugita. Kinetic energy definition in velocity Verlet integration for accurate pressure evaluation. *The Journal of Chemical Physics*, 148(16):164109, 2018.
- [221] I. P. Omelyan, I. M. Mryglod, and R. Folk. Optimized Verlet-like algorithms for molecular dynamics simulations. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, 65(5 Pt 2):056706, 2002.
- [222] Elizabeth A. Proctor, Feng Ding, and Nikolay V. Dokholyan. Discrete molecular dynamics. *WIREs Computational Molecular Science*, 1(1):80–92, 2011.
- [223] Mark James Abraham, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C. Smith, Berk Hess, and Erik Lindahl. GROMACS: High perfor-

- mance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2:19–25, 2015.
- [224] B.R. Brooks, C.L. Brooks, A.D. MacKerell, L. Nilsson, R.J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caffisch, L. Caves, Q. Cui, A.R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R.W. Pastor, C.B. Post, J.Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D.M. York, and M. Karplus. CHARMM: The Biomolecular Simulation Program. *Journal of computational chemistry*, 30(10):1545–1614, 2009.
- [225] James C. Phillips, David J. Hardy, Julio D. C. Maia, John E. Stone, João V. Ribeiro, Rafael C. Bernardi, Ronak Buch, Giacomo Fiorin, Jérôme Hénin, Wei Jiang, Ryan McGreevy, Marcelo C. R. Melo, Brian K. Radak, Robert D. Skeel, Abhishek Singharoy, Yi Wang, Benoît Roux, Aleksei Aksimentiev, Zaida Luthey-Schulten, Laxmikant V. Kalé, Klaus Schulten, Christophe Chipot, and Emad Tajkhorshid. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *The Journal of Chemical Physics*, 153(4):044130, 2020.
- [226] Aidan P. Thompson, H. Metin Aktulga, Richard Berger, Dan S. Bolinteanu, W. Michael Brown, Paul S. Crozier, Pieter J. in 't Veld, Axel Kohlmeyer, Stan G. Moore, Trung Dac Nguyen, Ray Shan, Mark J. Stevens, Julien Tranchida, Christian Trott, and Steven J. Plimpton. LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Computer Physics Communications*, 271:108171, 2022.
- [227] Louis Lagardère, Luc-Henri Jolly, Filippo Lipparini, Félix Aviat, Benjamin Stamm, Zhifeng F. Jing, Matthew Harger, Hedieh Torabifard, G. Andrés Cisneros, Michael J. Schnieders, Nohad Gresh, Yvon Maday, Pengyu Y. Ren, Jay W. Ponder, and Jean-Philip Piquemal. Tinker-HP: a massively parallel molecular dynamics package for multiscale simulations of large complex systems with advanced point dipole polarizable force fields. *Chemical Science*, 9(4):956–972, 2018.
- [228] Efrem Braun, Justin Gilmer, Heather B. Mayes, David L. Mobley, Jacob I. Monroe, Samarjeet Prasad, and Daniel M. Zuckerman. Best Practices

- for Foundations in Molecular Simulations [Article v1.0]. *Living Journal of Computational Molecular Science*, 1(1):5957, 2019.
- [229] William L. Jorgensen, Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, and Michael L. Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935, 1983.
- [230] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans. Interaction Models for Water in Relation to Protein Hydration. pages 331–342. 1981.
- [231] M. A. González. Force fields and molecular dynamics simulations. *École thématique de la Société Française de la Neutronique*, 12:169–200, 2011.
- [232] Ruslan L. Davidchack, Richard Handel, and M. V. Tretyakov. Langevin thermostat for rigid body dynamics. *The Journal of Chemical Physics*, 130(23):234101, 2009.
- [233] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, 81(8):3684–3690, 1984.
- [234] Martin Lísal, John K. Brennan, and Josep Bonet Avalos. Dissipative particle dynamics at isothermal, isobaric, isoenergetic, and isoenthalpic conditions using Shardlow-like splitting algorithms. *The Journal of Chemical Physics*, 135(20):204105, 2011.
- [235] Jean-Paul Ryckaert, Giovanni Ciccotti, and Herman J. C Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *Journal of Computational Physics*, 23(3):327–341, 1977.
- [236] Berk Hess, Henk Bekker, Herman J. C. Berendsen, and Johannes G. E. M. Fraaije. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18(12):1463–1472, 1997.
- [237] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, 14(1):33–38, 1996.
- [238] Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 1.8. 2015.

- [239] Eric F. Pettersen, Thomas D. Goddard, Conrad C. Huang, Gregory S. Couch, Daniel M. Greenblatt, Elaine C. Meng, and Thomas E. Ferrin. UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13):1605–1612, 2004.
- [240] Serge Pérez, Anita Sarkar, Alain Rivet, Sophie Drouillard, Christelle Breton, and Anne Imberty. Glyco3D: A Suite of Interlinked Databases of 3D Structures of Complex Carbohydrates, Lectins, Antibodies, and Glycosyltransferases. *A Practical Guide to Using Glycomics Databases*, pages 133–161, 2016.
- [241] Arunima Singh, David Montgomery, Xingran Xue, Bethany L. Foley, and Robert J. Woods. GAG Builder: a web-tool for modeling 3D structures of glycosaminoglycans. *Glycobiology*, 29(7):515–518, 2019.
- [242] Sang-Jun Park, Jumin Lee, Yifei Qi, Nathan R Kern, Hui Sun Lee, Sunhwan Jo, InSuk Joung, Keehyung Joo, Jooyoung Lee, and Wonpil Im. CHARMM-GUI Glycan Modeler for modeling and simulation of carbohydrates and glycoconjugates. *Glycobiology*, 29(4):320–331, 2019.
- [243] D. M. van Aalten, R. Bywater, J. B. Findlay, M. Hendlich, R. W. Hooft, and G. Vriend. PRODRG, a program for generating molecular topologies and unique molecular descriptors from coordinates of small molecules. *Journal of Computer-Aided Molecular Design*, 10(3):255–262, 1996.
- [244] Leela S. Dodda, Israel Cabeza de Vaca, Julian Tirado-Rives, and William L. Jorgensen. LigParGen web server: an automatic OPLS-AA parameter generator for organic ligands. *Nucleic Acids Research*, 45(Web Server issue):W331–W336, 2017.
- [245] David A. Case, Hasan Metin Aktulga, Kellon Belfon, David S. Cerutti, G. Andrés Cisneros, Vinícius Wilian D. Cruzeiro, Negin Forouzesh, Timothy J. Giese, Andreas W. Götz, Holger Gohlke, Saeed Izadi, Koushik Kasavajhala, Mehmet C. Kaymak, Edward King, Tom Kurtzman, Tai-Sung Lee, Pengfei Li, Jian Liu, Tyler Luchko, Ray Luo, Madushanka Manathunga, Matias R. Machado, Hai Minh Nguyen, Kurt A. O’Hearn, Alexey V. Onufriev, Feng Pan, Sergio Pantano, Ruxi Qi, Ali Rahnamoun, Ali Risheh, Stephan Schott-Verdugo, Akhil Shajan, Jason Swails, Junmei Wang, Haixin Wei, Xiongwu Wu, Yongxian Wu, Shi Zhang, Shiji Zhao, Qiang Zhu, Thomas E. III Cheatham, Daniel R. Roe, Adrian Roitberg, Carlos Simmerling, Darrin M. York, Maria C. Nagan, and Kenneth M. Jr. Merz.

- AmberTools. *Journal of Chemical Information and Modeling*, 63(20):6183–6191, 2023.
- [246] Karl N. Kirschner, Austin B. Yongye, Sarah M. Tschampel, Jorge González-Outeiriño, Charlisa R. Daniels, B. Lachele Foley, and Robert J. Woods. GLYCAM06: A generalizable biomolecular force field. *Carbohydrates. Journal of Computational Chemistry*, 29(4):622–655, 2008.
- [247] Robert J. Woods, Raymond A. Dwek, Christopher J. Edge, and Bert Fraser-Reid. Molecular Mechanical and Molecular Dynamic Simulations of Glycoproteins and Oligosaccharides. 1. GLYCAM_93 Parameter Development. *The Journal of Physical Chemistry*, 99(11):3832–3846, 1995.
- [248] Karl N. Kirschner and Robert J. Woods. Solvent interactions determine carbohydrate conformation. *Proceedings of the National Academy of Sciences of the United States of America*, 98(19):10541–10545, 2001.
- [249] Anita Plazinska and Wojciech Plazinski. Comparison of Carbohydrate Force Fields in Molecular Dynamics Simulations of Protein–Carbohydrate Complexes. *Journal of Chemical Theory and Computation*, 17(4):2575–2585, 2021.
- [250] Wentao Chen, Sebastian Enck, Joshua L. Price, David L. Powers, Evan T. Powers, Chi-Huey Wong, H. Jane Dyson, and Jeffery W. Kelly. Structural and Energetic Basis of Carbohydrate–Aromatic Packing Interactions in Proteins. *Journal of the American Chemical Society*, 135(26):9877–9884, 2013.
- [251] Martin Karplus. Molecular dynamics of biological macromolecules: A brief history and perspective. *Biopolymers*, 68(3):350–358, 2003.
- [252] Robert Schleif. Modeling and Studying Proteins with Molecular Dynamics. volume 383, pages 28–47. 2004.
- [253] Kaihsu Tai. Conformational sampling for the impatient. *Biophysical Chemistry*, 107(3):213–220, 2004.
- [254] Jan Norberg and Lennart Nilsson. Advances in biomolecular simulations: methodology and recent applications. *Quarterly Reviews of Biophysics*, 36(3):257–306, 2003.

-
- [255] Cecilia Clementi. Coarse-grained models of protein folding: toy models or predictive tools? *Current Opinion in Structural Biology*, 18(1):10–15, 2008.
- [256] Ivet Bahar, Timothy R. Lezon, Lee-Wei Yang, and Eran Eyal. Global dynamics of proteins: bridging between structure and function. *Annual Review of Biophysics*, 39:23–42, 2010.
- [257] Stephen Wells, Scott Menor, Brandon Hespeneide, and M. F. Thorpe. Constrained geometric simulation of diffusive motion in proteins. *Physical Biology*, 2(4):S127–136, 2005.
- [258] Kei Moritsugu and Jeremy C. Smith. Coarse-Grained Biomolecular Simulation with REACH: Realistic Extension Algorithm via Covariance Hessian. *Biophysical Journal*, 93(10):3460–3469, 2007.
- [259] Daniel Seeliger, Jürgen Haas, and Bert L. de Groot. Geometry-based sampling of conformational transitions in proteins. *Structure (London, England: 1993)*, 15(11):1482–1492, 2007.
- [260] Barak Raveh, Angela Enosh, Ora Schueler-Furman, and Dan Halperin. Rapid Sampling of Molecular Motions with Prior Information Constraints. *PLOS Computational Biology*, 5(2):e1000295, 2009.
- [261] Monique M. Tirion. Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. *Physical Review Letters*, 77(9):1905–1908, 1996.
- [262] Yuzo Ueda, Hiroshi Taketomi, and Nobuhiro Gō. Studies on protein folding, unfolding, and fluctuations by computer simulation. II. A. Three-dimensional lattice model of lysozyme. *Biopolymers*, 17(6):1531–1548, 1978.
- [263] A. J. Rader. Coarse-grained models: getting more with less. *Current Opinion in Pharmacology*, 10(6):753–759, 2010.
- [264] Konrad Hinsen, Nathalie Reuter, Jorge Navaza, David L. Stokes, and Jean-Jacques Lacapère. Normal Mode-Based Fitting of Atomic Structure into Electron Density Maps: Application to Sarcoplasmic Reticulum Ca-ATPase. *Biophysical Journal*, 88(2):818–827, 2005.
- [265] Marc Delarue and Philippe Dumas. On the use of low-frequency normal modes to enforce collective movements in refining macromolecular structural models. *Proceedings of the National Academy of Sciences*, 101(18):6957–6962, 2004.

-
- [266] Pablo Chacón, Florence Tama, and Willy Wriggers. Mega-Dalton Biomolecular Motion Captured from Electron Microscopy Reconstructions. *Journal of Molecular Biology*, 326(2):485–492, 2003.
- [267] Sibsanakar Kundu and Robert L. Jernigan. Molecular Mechanism of Domain Swapping in Proteins: An Analysis of Slower Motions. *Biophysical Journal*, 86(6):3846–3854, 2004.
- [268] Nese Kurt, Walter R.P. Scott, Celia A. Schiffer, and Turkan Haliloglu. Co-operative fluctuations of unliganded and substrate-bound HIV-1 protease: A structure-based analysis on a variety of conformations from crystallography and molecular dynamics simulations. *Proteins: Structure, Function, and Bioinformatics*, 51(3):409–422, 2003.
- [269] J. R. Baker and D. G. Pritchard. Action pattern and substrate specificity of the hyaluronan lyase from group B streptococci. *The Biochemical Journal*, 348 Pt 2(Pt 2):465–471, 2000.
- [270] Nobuyasu Koga and Shoji Takada. Roles of native topology and chain-length scaling in protein folding: A simulation study with a Gō-like model. *Journal of Molecular Biology*, 313(1):171–180, 2001.
- [271] Hyunbum Jang, Carol K. Hall, and Yaoqi Zhou. Assembly and Kinetic Folding Pathways of a Tetrameric β -Sheet Complex: Molecular Dynamics Simulations on Simplified Off-Lattice Protein Models. *Biophysical Journal*, 86(1):31–49, 2004.
- [272] Cecilia Clementi, Angel E. Garcíá, and José N. Onuchic. Interplay Among Tertiary Contacts, Secondary Structure Formation and Side-chain Packing in the Protein Folding Mechanism: All-atom Representation Study of Protein L. *Journal of Molecular Biology*, 326(3):933–954, 2003.
- [273] Isaac A. Hubner, Mikael Oliveberg, and Eugene I. Shakhnovich. Simulation, experiment, and evolution: Understanding nucleation in protein S6 folding. *Proceedings of the National Academy of Sciences*, 101(22):8354–8359, 2004.
- [274] Yaakov Levy, Amedeo Caffisch, Jose N. Onuchic, and Peter G. Wolynes. The Folding and Dimerization of HIV-1 Protease: Evidence for a Stable Monomer from Simulations. *Journal of Molecular Biology*, 340(1):67–79, 2004.

-
- [275] Michael Levitt. A simplified representation of protein conformations for rapid simulation of protein folding. *Journal of Molecular Biology*, 104(1):59–107, 1976.
- [276] Arnab Mukherjee and Biman Bagchi. Contact pair dynamics during folding of two small proteins: Chicken villin head piece and the Alzheimer protein β -amyloid. *The Journal of Chemical Physics*, 120(3):1602–1612, 2004.
- [277] Mey Khalili, Adam Liwo, Anna Jagielska, and Harold A. Scheraga. Molecular dynamics with the united-residue model of polypeptide chains. II. Langevin and Berendsen-bath dynamics and tests on model alpha-helical systems. *The Journal of Physical Chemistry. B*, 109(28):13798–13810, 2005.
- [278] Mey Khalili, Jeffrey A. Saunders, Adam Liwo, Stanislaw Oldziej, and Harold A. Scheraga. A united residue force-field for calcium-protein interactions. *Protein Science*, 13(10):2725–2735, 2004.
- [279] Siewert J. Marrink, H. Jelger Risselada, Serge Yefimov, D. Peter Tieleman, and Alex H. de Vries. The MARTINI force field: coarse grained model for biomolecular simulations. *The Journal of Physical Chemistry. B*, 111(27):7812–7824, 2007.
- [280] Anton Arkhipov, Peter L. Freddolino, and Klaus Schulten. Stability and dynamics of virus capsids described by coarse-grained modeling. *Structure (London, England: 1993)*, 14(12):1767–1777, 2006.
- [281] Anton Arkhipov, Ying Yin, and Klaus Schulten. Membrane-Bending Mechanism of Amphiphysin N-BAR Domains. *Biophysical Journal*, 97(10):2727–2735, 2009.
- [282] Gary S. Ayton and Gregory A. Voth. A Hybrid Coarse-graining Approach for Lipid Bilayers at Large Length and Time Scales. *The journal of physical chemistry. B*, 113(13):4413–4424, 2009.
- [283] Andrés Colubri. Prediction of Protein Structure by Simulating Coarse-grained Folding Pathways: A Preliminary Report. *Journal of Biomolecular Structure and Dynamics*, 21(5):625–638, 2004.
- [284] Seung Yup Lee, Yoshimi Fujitsuka, Do Hyun Kim, and Shoji Takada. Roles of physical interactions in determining protein-folding mechanisms: Molecular simulation of protein G and α spectrin SH3. *Proteins: Structure, Function, and Bioinformatics*, 55(1):128–138, 2004.

- [285] Adam D. Schuyler and Gregory S. Chirikjian. Normal mode analysis of proteins: a comparison of rigid cluster modes with C α coarse graining. *Journal of Molecular Graphics and Modelling*, 22(3):183–193, 2004.
- [286] Guohui Li and Qiang Cui. A Coarse-Grained Normal Mode Approach for Macromolecules: An Efficient Implementation and Application to Ca²⁺-ATPase. *Biophysical Journal*, 83(5):2457–2474, 2002.
- [287] Dagmar Flöck and Volkhard Helms. A Brownian Dynamics Study: The Effect of a Membrane Environment on an Electron Transfer System. *Biophysical Journal*, 87(1):65–74, 2004.
- [288] Adrian H. Elcock. Atomic-level observation of macromolecular crowding effects: Escape of a protein from the GroEL cage. *Proceedings of the National Academy of Sciences*, 100(5):2340–2344, 2003.
- [289] Akira R. Kinjo and Shoji Takada. Competition between Protein Folding and Aggregation with Molecular Chaperones in Crowded Solutions: Insight from Mesoscopic Simulations. *Biophysical Journal*, 85(6):3521–3531, 2003.
- [290] Sulin Jiang, Andrei Tovchigrechko, and Ilya A. Vakser. The role of geometric complementarity in secondary structure packing: A systematic docking study. *Protein Science*, 12(8):1646–1651, 2003.
- [291] L. Michel Espinoza-Fonseca. Base docking model of the homomeric $\alpha 7$ nicotinic receptor- β -amyloid1–42 complex. *Biochemical and Biophysical Research Communications*, 320(2):587–591, 2004.
- [292] Teemu Murtola, Emma Falck, Mikko Karttunen, and Ilpo Vattulainen. Coarse-grained model for phospholipid/cholesterol bilayer employing inverse Monte Carlo with thermodynamic constraints. *The Journal of Chemical Physics*, 126(7):075101, 2007.
- [293] Sergei Izvekov and Gregory A. Voth. A multiscale coarse-graining method for biomolecular systems. *The Journal of Physical Chemistry. B*, 109(7):2469–2473, 2005.
- [294] Marcos R. Betancourt and Sheyore J. Omovie. Pairwise energies for polypeptide coarse-grained models derived from atomic force fields. *The Journal of Chemical Physics*, 130(19):195103, 2009.

- [295] Semen O. Yesylevskyy, Lars V. Schäfer, Durba Sengupta, and Siewert J. Marrink. Polarizable Water Model for the Coarse-Grained MARTINI Force Field. *PLoS Computational Biology*, 6(6):e1000810, 2010.
- [296] Luca Monticelli, Senthil K. Kandasamy, Xavier Periole, Ronald G. Larson, D. Peter Tieleman, and Siewert-Jan Marrink. The MARTINI Coarse-Grained Force Field: Extension to Proteins. *Journal of Chemical Theory and Computation*, 4(5):819–834, 2008.
- [297] Adam Liwo, Cezary Czaplewski, Adam K. Sieradzan, Agnieszka G. Lipska, Sergey A. Samsonov, and Rajesh K. Murarka. Theory and Practice of Coarse-Grained Molecular Dynamics of Biologically Important Systems. *Biomolecules*, 11(9):1347, 2021.
- [298] Mark Bathe, Gregory C. Rutledge, Alan J. Grodzinsky, and Bruce Tidor. A Coarse-Grained Molecular Model for Glycosaminoglycans: Application to Chondroitin, Chondroitin Sulfate, and Hyaluronic Acid. *Biophysical Journal*, 88(6):3870–3887, 2005.
- [299] Benedict M. Sattelle, Javad Shakeri, Matthew J. Cliff, and Andrew Almond. Proteoglycans and Their Heterogeneous Glycosaminoglycans at the Atomic Scale. *Biomacromolecules*, 16(3):951–961, 2015.
- [300] Benedict M. Sattelle, Javad Shakeri, and Andrew Almond. Does microsecond sugar ring flexing encode 3D-shape and bioactivity in the heparanome? *Biomacromolecules*, 14(4):1149–1159, 2013.
- [301] Sergey A. Samsonov, Leon Bichmann, and M. Teresa Pisabarro. Coarse-Grained Model of Glycosaminoglycans. *Journal of Chemical Information and Modeling*, 55(1):114–124, 2015.
- [302] Adam Liwo, Cezary Czaplewski, Adam K. Sieradzan, Emilia A. Lubecka, Agnieszka G. Lipska, Lukasz Golon, Agnieszka Karczynska, Pawel Krupa, Magdalena A. Mozolewska, Mariusz Makowski, Artur Gieldon, and Maciej Maciejczyk. Chapter Two - Scale-consistent approach to the derivation of coarse-grained force fields for simulating structure, dynamics, and thermodynamics of biopolymers. volume 170, pages 73–122. 2020.
- [303] Emilia A. Lubecka and Adam Liwo. A general method for the derivation of the functional forms of the effective energy terms in coarse-grained

- energy functions of polymers. II. Backbone-local potentials of coarse-grained O1-4-bonded polyglucose chains. *The Journal of Chemical Physics*, 147(11):115101, 2017.
- [304] Adam Liwo, Maciej Baranowski, Cezary Czaplewski, Ewa Gołaś, Yi He, Dawid Jagieła, Paweł Krupa, Maciej Maciejczyk, Mariusz Makowski, Magdalena A. Mozolewska, Andrei Niadzvedtski, Stanisław Oldziej, Harold A. Scheraga, Adam K. Sieradzan, Rafał Slusarz, Tomasz Wirecki, Yanping Yin, and Bartłomiej Zaborowski. A unified coarse-grained model of biological macromolecules based on mean-field multipole-multipole interactions. *Journal of Molecular Modeling*, 20(8):2306, 2014.
- [305] Yi He, Magdalena A. Mozolewska, Paweł Krupa, Adam K. Sieradzan, Tomasz K. Wirecki, Adam Liwo, Khatuna Kachlishvili, Shalom Rackovsky, Dawid Jagieła, Rafał Slusarz, Cezary R. Czaplewski, Stanisław Oldziej, and Harold A. Scheraga. Lessons from application of the UNRES force field to predictions of structures of CASP10 targets. *Proceedings of the National Academy of Sciences of the United States of America*, 110(37):14936–14941, 2013.
- [306] Rui Zhou, Gia G. Maisuradze, David Sunol, Toni Todorovski, Maria J. Macias, Yi Xiao, Harold A. Scheraga, Cezary Czaplewski, and Adam Liwo. Folding kinetics of WW domains with the united residue force field for bridging microscopic motions and experimental measurements. *Proceedings of the National Academy of Sciences of the United States of America*, 111(51):18243–18248, 2014.
- [307] Ewa Golas, Gia G. Maisuradze, Patrick Senet, Stanisław Oldziej, Cezary Czaplewski, Harold A. Scheraga, and Adam Liwo. Simulation of the opening and closing of Hsp70 chaperones by coarse-grained molecular dynamics. *Journal of Chemical Theory and Computation*, 8(5):1750–1764, 2012.
- [308] Yi He, Maciej Maciejczyk, Stanisław Oldziej, Harold A. Scheraga, and Adam Liwo. Mean-field interactions between nucleic-acid-base dipoles can drive the formation of a double helix. *Physical Review Letters*, 110(9):098101, 2013.
- [309] Aurijit Sarkar, Wenbo Yu, Umesh R Desai, Alexander D MacKerell, and Philip D Mosier. Estimating glycosaminoglycan-protein interaction affinity: water dominates the specific antithrombin-heparin interaction. *Glycobiology*, 26(10):1041–1047, 2016.

- [310] Aishwary T. Shivgan, Jan K. Marzinek, Alexander Krah, Paul Matsudaira, Chandra S. Verma, and Peter J. Bond. Coarse-Grained Model of Glycosaminoglycans for Biomolecular Simulations. *Journal of Chemical Theory and Computation*, 2024.
- [311] Sergey A. Samsonov and M. Teresa Pisabarro. Importance of IdoA and IdoA(2S) ring conformations in computational studies of glycosaminoglycan-protein interactions. *Carbohydr Res*, 381:133–137, 2013.
- [312] Mateusz Marcisz, Bertrand Huard, Agnieszka G. Lipska, and Sergey A. Samsonov. Further analyses of APRIL/APRIL-receptor/glycosaminoglycan interactions by biochemical assays linked to computational studies. *Glycobiology*, 31(7):772–786, 2021.
- [313] Subhomoi Borkotoky, Chetan Kumar Meena, and Ayaluru Murali. Interaction Analysis of T7 RNA Polymerase with Heparin and Its Low Molecular Weight Derivatives - An In Silico Approach. *Bioinform Biol Insights*, 10:155–166, 2016.
- [314] Mateusz Marcisz, Martin Zacharias, and Sergey A. Samsonov. Modeling Protein-Glycosaminoglycan Complexes: Does the Size Matter? *J Chem Inf Model*, 61(9):4475–4485, 2021.
- [315] Mateusz Marcisz, Sebastian Anila, Margrethe Gaardl s, Martin Zacharias, and Sergey A. Samsonov. Studying specificity in protein-glycosaminoglycan recognition with umbrella sampling. *Beilstein J Org Chem*, 19:1933–1946, 2023.
- [316] Sergey A. Samsonov, Sylwia Freza, and Ferenc Zsila. In silico analysis of heparin and chondroitin sulfate binding mechanisms of the antiprotozoal drug berenil and pentamidine. *Carbohydrate Research*, 482:107742, 2019.
- [317] Benedict M. Sattelle and Andrew Almond. Less is more when simulating unsulfated glycosaminoglycan 3D-structure: Comparison of GLYCAM06/TIP3P, PM3-CARB1/TIP3P, and SCC-DFTB-D/TIP3P predictions with experiment. *Journal of Computational Chemistry*, 31(16):2932–2947, 2010.
- [318] Priyanka Samanta, Sushil K. Mishra, Vitor H. Pomin, and Robert J. Doerksen. Docking and Molecular Dynamics Simulations Clarify Binding Sites for Interactions of Novel Marine Sulfated Glycans with SARS-CoV-2 Spike Glycoprotein. *Molecules*, 28(17), 2023.

- [319] Ayako Furukawa, Tsuyoshi Konuma, Saeko Yanaka, and Kenji Sugase. Quantitative analysis of protein-ligand interactions by NMR. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 96:47–57, 2016.
- [320] Charlotte A. Softley, Mark J. Bostock, Grzegorz M. Popowicz, and Michael Sattler. Paramagnetic NMR in drug discovery. *Journal of biomolecular NMR*, 74(6-7):287–309, 2020.
- [321] Christoph Hartmüller, Johannes C. Günther, Antje C. Wolter, Jens Wöhnert, Michael Sattler, and Tobias Madl. RNA structure refinement using NMR solvent accessibility data. *Scientific Reports*, 7(1):5393, 2017.
- [322] Ki-Young Lee, Zhenhao Fang, Masahiro Enomoto, Genevieve Gasmi-Seabrook, Le Zheng, Shohei Koide, Mitsuhiro Ikura, and Christopher B. Marshall. Two Distinct Structures of Membrane-Associated Homodimers of GTP- and GDP-Bound KRAS4B Revealed by Paramagnetic Relaxation Enhancement. *Angewandte Chemie (International Ed. in English)*, 59(27):11037–11045, 2020.
- [323] Christoph Hartmüller, Christoph Göbl, and Tobias Madl. Prediction of Protein Structure Using Surface Accessibility Data. *Angewandte Chemie (International Ed. in English)*, 55(39):11970–11974, 2016.
- [324] Yu Wang, Charles D. Schwieters, and Nico Tjandra. Parameterization of solvent-protein interaction and its use on NMR protein structure determination. *Journal of Magnetic Resonance (San Diego, Calif.: 1997)*, 221:76–84, 2012.
- [325] Christoph Hartmüller, Emil Spreitzer, Christoph Göbl, Fabio Falsone, and Tobias Madl. NMR characterization of solvent accessibility and transient structure in intrinsically disordered proteins. *Journal of biomolecular NMR*, 73(6-7):305–317, 2019.
- [326] S. Arumugam, C. L. Hemme, N. Yoshida, K. Suzuki, H. Nagase, M. Berjanskii, B. Wu, and S. R. Van Doren. TIMP-1 contact sites and perturbations of stromelysin 1 mapped by NMR and a paramagnetic surface probe. *Biochemistry*, 37(27):9650–9657, 1998.
- [327] Tobias Madl, Thomas Güttler, Dirk Görlich, and Michael Sattler. Structural analysis of large protein complexes using solvent paramagnetic relaxation enhancements. *Angewandte Chemie (International Ed. in English)*, 50(17):3993–3997, 2011.

- [328] Jennifer H. Tomlinson, Gary S. Thompson, Arnout P. Kalverda, Anastasia Zhuravleva, and Alex J. O'Neill. A target-protection mechanism of antibiotic resistance at atomic resolution: insights into FusB-type fusidic acid resistance. *Scientific Reports*, 6:19524, 2016.
- [329] Binhuan Yu, Channing C. Pletka, B. Montgomery Pettitt, and Junji Iwahara. De novo determination of near-surface electrostatic potentials by NMR. *Proceedings of the National Academy of Sciences*, 118(25):e2104020118, 2021.
- [330] Binhuan Yu, Channing C. Pletka, and Junji Iwahara. Quantifying and visualizing weak interactions between anions and proteins. *Proceedings of the National Academy of Sciences*, 118(2):e2015879118, 2021.
- [331] Binhuan Yu, Xi Wang, and Junji Iwahara. Measuring Local Electrostatic Potentials Around Nucleic Acids by Paramagnetic NMR Spectroscopy. *The Journal of Physical Chemistry Letters*, 13(42):10025–10029, 2022.
- [332] Elizabeth Jurrus, Dave Engel, Keith Star, Kyle Monson, Juan Brandi, Lisa E. Felberg, David H. Brookes, Leighton Wilson, Jiahui Chen, Karina Liles, Minju Chun, Peter Li, David W. Gohara, Todd Dolinsky, Robert Konecny, David R. Koes, Jens Erik Nielsen, Teresa Head-Gordon, Weihua Geng, Robert Krasny, Guo-Wei Wei, Michael J. Holst, J. Andrew McCammon, and Nathan A. Baker. Improvements to the APBS biomolecular solvation software suite. *Protein Science: A Publication of the Protein Society*, 27(1):112–128, 2018.
- [333] Marco Guerrini, Sara Guglieri, Benito Casu, Giangiacomo Torri, Pierre Mourier, Christian Boudier, and Christian Viskov. Antithrombin-binding octasaccharides and role of extensions of the active pentasaccharide sequence in the specificity and strength of interaction. Evidence for very high affinity induced by an unusual glucuronic acid residue. *The Journal of Biological Chemistry*, 283(39):26662–26675, 2008.
- [334] Vera Hintze, Sergey A. Samsonov, Massimiliano Anselmi, Stephanie Moeller, Jana Becher, Matthias Schnabelrauch, Dieter Scharnweber, and M. Teresa Pisabarro. Sulfated glycosaminoglycans exploit the conformational plasticity of bone morphogenetic protein-2 (BMP-2) and alter the interaction profile with its receptor. *Biomacromolecules*, 15(8):3083–3092, 2014.

- [335] Linda Koehler, Sergey Samsonov, Sandra Rother, Sarah Vogel, Sebastian Köhling, Stephanie Moeller, Matthias Schnabelrauch, Jörg Rademann, Ute Hempel, M. Teresa Pisabarro, Dieter Scharnweber, and Vera Hintze. Sulfated Hyaluronan Derivatives Modulate TGF- β 1:Receptor Complex Formation: Possible Consequences for TGF- β 1 Signaling. *Scientific Reports*, 7(1):1210, 2017.
- [336] Fabian Gude, Jurij Froese, Georg Steffes, and Kay Grobe. The role of glycosaminoglycan modification in Hedgehog regulated tissue morphogenesis. *Biochemical Society Transactions*, 51(3):983–993, 2023.
- [337] Folkert Verkaar, Jody van Offenbeek, Miranda M. C. van der Lee, Lambertus H. C. J. van Lith, Anne O. Watts, Angelique L. W. M. M. Rops, David C. Aguilar, Joshua J. Ziarek, Johan van der Vlag, Tracy M. Handel, Brian F. Volkman, Amanda E. I. Proudfoot, Henry F. Vischer, Guido J. R. Zaman, and Martine J. Smit. Chemokine Cooperativity Is Caused by Competitive Glycosaminoglycan Binding. *The Journal of Immunology*, 192(8):3908–3914, 2014.
- [338] Martha Gschwandtner, Elisabeth Strutzmann, Mauro M. Teixeira, Hans J. Anders, Maria Diedrichs-Möhrling, Tanja Gerlza, Gerhild Wildner, Remo C. Russo, Tiziana Adage, and Andreas J. Kungl. Glycosaminoglycans are important mediators of neutrophilic inflammation in vivo. *Cytokine*, 91:65–73, 2017.
- [339] Pavani Gangavarapu, Lavanya Rajagopalan, Deepthi Kolli, Antonieta Guerrero-Plata, Roberto P. Garofalo, and Krishna Rajarathnam. The monomer-dimer equilibrium and glycosaminoglycan interactions of chemokine CXCL8 regulate tissue-specific neutrophil recruitment. *Journal of Leukocyte Biology*, 91(2):259–265, 2012.
- [340] Sylvie Ricard-Blum, Mickael Beraud, Nicolas Raynal, Richard W. Farnsdale, and Florence Ruggiero. Structural Requirements for Heparin/Heparan Sulfate Binding to Type V Collagen*. *Journal of Biological Chemistry*, 281(35):25195–25204, 2006.
- [341] Caroline Bonnans, Jonathan Chou, and Zena Werb. Remodelling the extracellular matrix in development and disease. *Nature Reviews Molecular Cell Biology*, 15(12):786–801, 2014.

-
- [342] Ji Young Suk, Fuming Zhang, William E. Balch, Robert J. Linhardt, and Jeffery W. Kelly. Heparin Accelerates Gelsolin Amyloidogenesis. *Biochemistry*, 45(7):2234–2242, 2006.
- [343] Masashi Egashira, Hiroka Takase, Izumi Yamamoto, Masafumi Tanaka, and Hiroyuki Saito. Identification of regions responsible for heparin-induced amyloidogenesis of human serum amyloid A using its fragment peptides. *Archives of Biochemistry and Biophysics*, 511(1):101–106, 2011.
- [344] Annemarie Danielsson, Sergey A. Samsonov, Adam Liwo, and Adam K. Sieradzian. Extension of the SUGRES-1P Coarse-Grained Model of Polysaccharides to Heparin. *Journal of Chemical Theory and Computation*, 19(17):6023–6036, 2023.
- [345] Sanaullah Khan, Jayesh Gor, Barbara Mulloy, and Stephen J. Perkins. Semi-rigid solution structures of heparin by constrained X-ray scattering modelling: new insight into heparin-protein complexes. *Journal of Molecular Biology*, 395(3):504–521, 2010.
- [346] Georges Pavlov, Stephanie Finet, Karine Tatarenko, Evgueniya Korneeva, and Christine Ebel. Conformation of heparin studied with macromolecular hydrodynamic methods and X-ray scattering. *European biophysics journal: EBJ*, 32(5):437–449, 2003.
- [347] S. Faham, R. E. Hileman, J. R. Fromm, R. J. Linhardt, and D. C. Rees. Heparin structure and interactions with basic fibroblast growth factor. *Science (New York, N.Y.)*, 271(5252):1116–1120, 1996.
- [348] Ding Xu and Jeffrey D. Esko. Demystifying Heparan Sulfate-Protein Interactions. *Annual review of biochemistry*, 83:129–157, 2014.
- [349] L. D. Thompson, M. W. Pantoliano, and B. A. Springer. Energetic characterization of the basic fibroblast growth factor-heparin interaction: identification of the heparin binding domain. *Biochemistry*, 33(13):3831–3840, 1994.
- [350] Balaji Nagarajan, Nehru Viji Sankaranarayanan, and Umesh R Desai. Rigorous analysis of free solution glycosaminoglycan dynamics using simple, new tools. *Glycobiology*, 30(8):516–527, 2020.

-
- [351] Marta Pagielska and Sergey A. Samsonov. Molecular Dynamics-Based Comparative Analysis of Chondroitin and Dermatan Sulfates. *Biomolecules*, 13(2):247, 2023.
- [352] Sara Guglier, Milos Hricovini, Rahul Raman, Laura Polito, Giangiacomo Torri, Benito Casu, Ram Sasisekharan, and Marco Guerrini. Minimum FGF2 binding structural requirements of heparin and heparan sulfate oligosaccharides as determined by NMR spectroscopy. *Biochemistry*, 47(52):13862–13869, 2008.
- [353] Satoko Ashikari-Hada, Hiroko Habuchi, Noriko Sugaya, Takashi Kobayashi, and Koji Kimata. Specific inhibition of FGF-2 signaling with 2-O-sulfated octasaccharides of heparan sulfate. *Glycobiology*, 19(6):644–654, 2009.
- [354] Eduardo Stancanelli, Stefano Elli, Po-Hung Hsieh, Jian Liu, and Marco Guerrini. Recognition and Conformational Properties of an Alternative Antithrombin Binding Sequence Obtained by Chemoenzymatic Synthesis. *Chembiochem: A European Journal of Chemical Biology*, 2018.
- [355] Pradeep Chopra, Apoorva Joshi, Jiandong Wu, Weigang Lu, Tejabhram Yadavalli, Margreet A. Wolfert, Deepak Shukla, Joseph Zaia, and Geert-Jan Boons. The 3-O-sulfation of heparan sulfate modulates protein binding and lyase degradation. *Proceedings of the National Academy of Sciences of the United States of America*, 118(3):e2012935118, 2021.

Publications included in the PhD Thesis

Publication 1. Detecting Protein-Ligand
Interactions with Nitroxide Based
Paramagnetic Cosolutes.

Detecting Protein-Ligand Interactions with Nitroxide Based Paramagnetic Cosolutes

Anja Penk,^[a] Annemarie Danielsson,^[b] Margrethe Gaardl s,^[b] Cindy Montag,^[a] Andrea Sch ler,^[c] Daniel Huster,^[a] Sergey A. Samsonov,^{*,[b]} and Georg K nze^{*,[c]}

NMR spectroscopy techniques can provide important information about protein-ligand interactions. Here we tested an NMR approach which relies on the measurement of paramagnetic relaxation enhancements (PREs) arising from analogous cationic, anionic or neutral soluble nitroxide molecules, which distribute around the protein-ligand complex depending on near-surface electrostatic potentials. We applied this approach to two protein-ligand systems, interleukin-8 interacting with highly charged glycosaminoglycans and the SH2 domain of Grb2 interacting with less charged phospho-tyrosine tripep-

tides. The electrostatic potential around interleukin-8 and its changes upon binding of glycosaminoglycans could be derived from the PRE data and confirmed by theoretical predictions from Poisson-Boltzmann calculations. The ligand influence on the PREs and NMR-derived electrostatic potentials of Grb2 SH2 was localized to a narrow protein region which allowed the localization of the peptide binding pocket. Our analysis suggests that experiments with nitroxide cosolutes can be useful for investigating protein-ligand electrostatic interactions and mapping ligand binding sites.

Introduction

Elucidating the binding mode of biological macromolecules with their ligands is essential for understanding the biological effects triggered by the ligand binding process as well as for the discovery and design of new bioactive molecules. Nuclear magnetic resonance (NMR) spectroscopy represents a versatile approach to characterize protein-ligand interactions and is sensitive to a wide range of ligand affinities (nanomolar to millimolar).^[1] Several NMR methods are available to obtain structural, thermodynamic, and kinetic information about protein-ligand systems.^[2] For example, the measurement of chemical shift perturbations (CSPs) allows calculation of equilibrium dissociation constants and identification of protein residues that bind the ligand. An advantage of the CSP method is its ability to capture protein conformational changes, which

can occur upon ligand binding. This provides opportunities for identifying and targeting allosteric binding sites.^[3,4] However, care must be taken in the interpretation of CSPs since they can reflect either direct protein-ligand interactions or secondary conformational effects. Other NMR methods, such as saturation transfer difference (STD) NMR spectroscopy,^[5] transferred nuclear Overhauser effect (NOE) NMR spectroscopy,^[6] and the Water-LOGSY method,^[7] rely on detecting the NMR signals of the ligand. STD NMR and Water-LOGSY can be used to determine the ligand moieties that are important for binding^[8] and also to screen mixtures of ligand fragments for their ability to bind to proteins.^[9,10] Transferred NOEs can be utilized to determine the ligand conformation in the bound state,^[11] which can be helpful for modeling the 3D structure of the protein-ligand complex. Moreover, they can be used for ligand screening.^[12]

Paramagnetic NMR methods further expand the application range of NMR in drug research.^[13] An important application of paramagnetic NMR is the collection of structural restraints on the ligand binding pose which can guide the drug design process in medicinal chemistry. This is especially useful in the early stages of drug discovery when crystal structures of weakly binding ligands are difficult to obtain. The paramagnetic restraints such as pseudocontact shifts (PCSs) or paramagnetic relaxation enhancements (PREs) can be integrated via different approaches to guide ligand docking calculations or filter the generated interaction models. PCSs can be used in a similar manner as CSPs to locate the ligand binding site. In order to exploit the paramagnetic effects, different approaches for attaching the paramagnetic center (e.g. a nitroxide radical or lanthanide ion) to the receptor protein or ligand molecule of interest exist.^[14–22] While the PCS approach is well suited for fast-exchanging ligands with micro- to millimolar dissociation constants, PCSs on tightly binding, slow-exchanging ligands

[a] Dr. A. Penk,⁺ C. Montag, Prof. Dr. D. Huster
Institute for Medical Physics and Biophysics
University of Leipzig
H rtelstr. 16/18, D-04107 Leipzig, Germany

[b] A. Danielsson,⁺ Dr. M. Gaardl s, Prof. Dr. S. A. Samsonov
Faculty of Chemistry
University of Gda nsk
Ul. Wita Stwosza 63, 80-308 Gda nsk, Poland
E-mail: sergey.samsonov@ug.edu.pl

[c] A. Sch ler, Dr. G. K nze
Institute for Drug Discovery
University of Leipzig
Br derstr. 34, D-04103 Leipzig, Germany
E-mail: georg.kuenze@uni-leipzig.de

Supporting information for this article is available on the WWW under <https://doi.org/10.1002/chem.202303570>

  2023 The Authors. Chemistry - A European Journal published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution Non-Commercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

could be detected using special reporter groups e.g. tert-butyl^[18] or ¹⁹F nuclei.^[19]

One concern about these covalent tagging approaches is that the addition of such large tags to the ligand molecule could significantly alter its binding mode. Affinity data are also not easily transferrable to the untagged ligand. Thus, tagging-free NMR techniques represent an interesting alternative. Solvent PRE (sPRE) effects can be induced by soluble, paramagnetic probes that are directly added to the protein solution, requiring no chemical modification of the protein or ligand. sPREs offer a rich source of structural information and have been used for a variety of applications in structural biology, e.g., for the mapping of the solvent accessible surface of biomolecules,^[23,24] protein structure prediction,^[25,26] the detection of transient conformational states,^[27] and the localization of the binding interface in protein-protein complexes.^[28–30] Recently, Iwahara and coworkers developed an NMR method that uses paramagnetic 2,2,5,5-tetramethylpyrrolidine-N-oxyl nitroxide (PROXYL) cosolutes to determine the local electrostatic potential (ϕ) around biomolecules.^[31–33] Specifically, two oppositely charged PROXYL molecules are used and their PRE effects are quantified. The PROXYL probes distribute around the macromolecule according to the near-surface electrostatic potential (ϕ_{ENS}), which can be determined with per-residue resolution from the ratio of PRE rates induced by the two PROXYL derivatives. The authors concluded, that the NMR-derived data of ϕ_{ENS} agreed well with predictions from the Poisson-Boltzmann theory and that this method could be used for the *de novo* determination of ϕ_{ENS} .^[31]

Electrostatic interactions play an important role in protein-ligand recognition processes.^[34] They account for a sizeable contribution to the binding affinity and the stabilization of weakly specific transition states that occur along the binding pathway.^[35] Furthermore, electrostatic complementarity at the protein-ligand interface is recognized as an important factor for predicting and optimizing ligand affinity and selectivity.^[36,37] We reasoned that the ionic PROXYL probes could also be useful for the detection of protein-ligand interactions, in particular for ligands bearing charged chemical groups, since those will significantly perturb the protein electrostatic potential in the binding site. Here, we extend the PROXYL method to the investigation of protein-ligand complexes and apply it to two protein-ligand systems, the protein interleukin-8 (IL-8), interacting with glycosaminoglycans (GAGs), and the src-homology 2 (SH2) domain of the growth factor receptor-bound protein 2 (Grb2), interacting with phosphotyrosine peptides.

IL-8 is, depending on its splicing variant, a 8–9 kDa protein that belongs to the CXC chemokine family and plays important roles in the activation and recruitment of neutrophil granulocytes to sites of inflammation.^[38,39] IL-8 strongly interacts with GAGs from the extracellular matrix, such as heparan sulfate or chondroitin sulfate. The interaction with GAGs contributes to the formation of a dynamic chemokine gradient which mediates the attraction and migration of immune cells.^[40] GAGs are highly negatively charged polysaccharides consisting of periodic disaccharide units and carry a large number of sulfate groups. Heparin (HP) is the most highly sulfated member of the

GAG family and an important pharmaceutical anticoagulant. The interaction of IL-8 with HP and other GAGs has been demonstrated by several experimental and computational studies which revealed important GAG binding residues on IL-8 as well as the binding mode for shorter GAG fragments.^[41–47]

Grb2 is a 25 kDa adaptor protein consisting of one SH2 domain and two SH3 domains, arranged in the order SH3-SH2-SH3.^[48] The SH2 domain is a small domain of approximately 100 amino acid residues which binds to specific tyrosine phosphorylation sites on several proteins.^[49] Binding of the Grb2 SH2 domain to phosphotyrosine (pY) motifs on receptors, e.g. EGF receptor,^[50] recruits the Ras exchange factor Sos at the plasma membrane where Sos stimulates nucleotide exchange on Ras, triggering the downstream signaling of Raf and MAP kinases.^[51] The phosphotyrosine (pY) recognition consensus sequence of Grb2 SH2 is pYXN, where X represents any amino acid and N represents asparagine.^[52,53]

In this study, we evaluate the applicability of the PROXYL method for mapping ligand binding sites in IL-8 and Grb2 SH2, and for detecting changes in the protein near-surface electrostatic potential (ϕ_{ENS}) induced by interactions with their respective ligands. Our results show that the PROXYL method can detect ligand-induced changes of ϕ_{ENS} , which are in good agreement with theoretical predictions. A highly charged HP ligand alters the PRE and ϕ_{ENS} data across a large region in IL-8, while pYXN peptides cause more localized changes in a narrow region in Grb2 SH2, which allow a precise localization of the peptide binding site. Our findings suggest that experiments with paramagnetic cosolutes can be a useful addition to the toolbox of NMR methods for studying protein-ligand interactions.

Results and Discussion

Calculation of the electrostatic potential of IL-8

The soluble PROXYL probes distribute around the protein according to its electrostatic potential. We first investigated the strength and distribution of ϕ_{ENS} of IL-8 using computational approaches. Starting from the homodimeric NMR structure of IL-8^[54] we sampled an ensemble of structures of IL-8 using molecular dynamics (MD) simulation and calculated ϕ_{ENS} for these structures using the Adaptive Poisson-Boltzmann Solver (APBS) software.^[55,56] Figure 1A shows mappings of ϕ_{ENS} for different isosurface values. As can be seen in Figure 1A, ϕ_{ENS} is mostly positive over large regions around IL-8 since IL-8 is strongly positively charged at neutral pH.

We then applied molecular docking and MD simulation to model the complex of IL-8 with GAGs and studied the influence of GAG binding on ϕ_{ENS} . We chose two different GAGs, HP and hyaluronan (HA), because of their different charge densities. While HP contains a maximum of four negatively charged groups per disaccharide, HA features only one negatively charged carboxyl group per disaccharide and is the GAG with the lowest net charge. For computational efficiency, both GAGs were modeled as hexasaccharides (dp6, where dp stands for

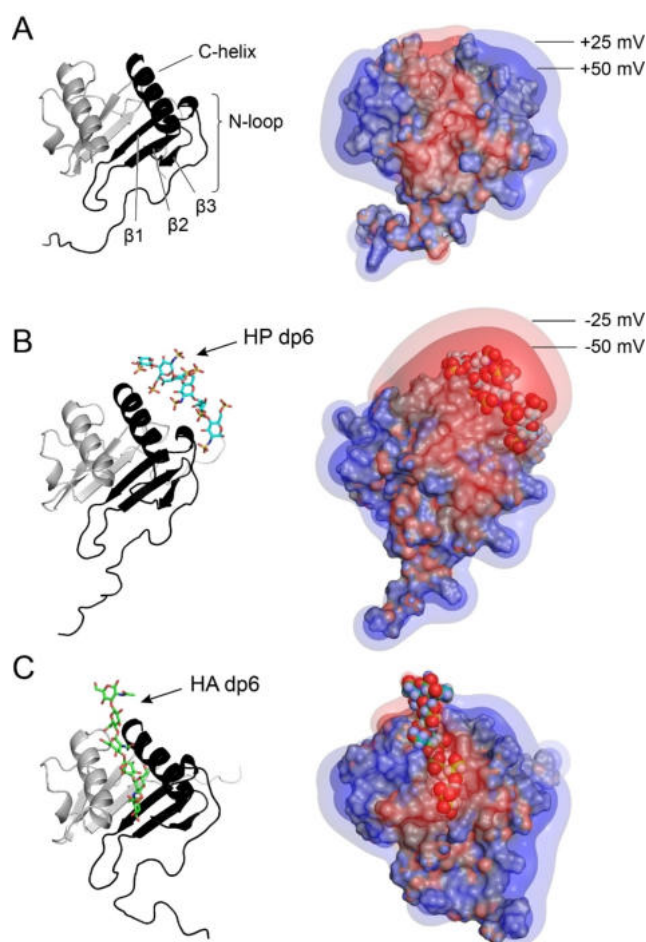


Figure 1. Simulated electrostatic potentials for IL-8 and IL-8-ligand complexes. (A) IL-8 apo. Secondary structure regions are labeled. (B) IL-8/HP dp6 complex. (C) IL-8/HA dp6 complex.

degree of polymerization). The binding pose of HP dp6 with the lowest energy in the MD simulation is displayed in Fig 1B. The HP dp6 molecule adopts an extended conformation located between the C-helix and N-loop of one IL-8 monomer and oriented parallel to the C-terminal C-helix. This binding pose corresponds to the one previously analyzed by us and others, both theoretically and experimentally.^[44,46,47,57,58] The lowest energy binding pose of HA dp6 observed in the MD simulations was located in a central groove formed by the two C-helices of the IL-8 dimer (Figure 1C). This alternative GAG binding pose is not unexpected. It was observed in previous simulation studies,^[44,47,58] however, was less populated than the binding pose between N-loop and C-helix. It was also seen in some MD runs with HP dp6 in this study. As expected from the high negative charge of HP dp6 of $-12e$, it strongly changes ϕ_{ENS} around IL-8 to more negative (Figure 1B). By contrast, HA dp6 has a comparatively small effect and does not significantly change ϕ_{ENS} in the environment of IL-8 (Figure 1C).

Paramagnetic NMR data for IL-8 and IL-8/GAG complexes

Next, the distribution of ϕ_{ENS} around IL-8 and IL-8/GAG complexes was investigated experimentally using paramagnetic NMR. We used ^{15}N -labeled IL-8 and prepared the protein as previously described yielding the homodimeric form of IL-8 (1–77) in the used NMR concentrations.^[44] We used the paramagnetic PROXYL derivatives shown in Figure 2A as cosolutes and measured the ^1H transverse PRE rates (Γ_2) for the $^1\text{H}_\text{N}$ backbone amide groups of IL-8 (example spectrum of IL-8 in Fig S1) using the two time-point approach (see Experimental Section). Specifically, cationic aminomethyl-PROXYL and anionic carboxy-PROXYL were used to measure $\Gamma_{2,+}$ and $\Gamma_{2,-}$ rates, respectively. We observed clear signal intensity changes for many resonances in the ^1H - ^{15}N HSQC spectrum of IL-8 (examples shown in Fig 2B) relative to the diamagnetic reference spectrum. This allowed us to measure $\Gamma_{2,+}$ rates for 55 residues and $\Gamma_{2,-}$ rates for 58 residues in IL-8, respectively (Table S1). As can be seen from the signal slices in Fig 2B and the PRE-vs-residue number plot in Fig 2C, the $\Gamma_{2,+}$ and $\Gamma_{2,-}$ rates vary significantly for several residues of IL-8. Large $\Gamma_{2,+}$ rates above 40 Hz were measured for residues in the N-terminus, the N-loop, and the $\beta 2$ and $\beta 3$ strand. Large $\Gamma_{2,-}$ rates above 40 Hz were observed for residues in the N-loop, the $\beta 3$ strand and the C-helix (Fig 2C). The different patterns of $\Gamma_{2,+}$ versus $\Gamma_{2,-}$ rates are also evident from their mapping onto the 3-dimensional structure of IL-8 shown in Fig S2. This indicates that the two oppositely charged PROXYL molecules distribute with a different probability density in the electric field generated by IL-8.

Next, we prepared samples of IL-8 with HP dp6 or HA dp6, respectively, by stepwise titration with GAG, while recording a ^1H - ^{15}N HSQC spectrum for each titration step (see Fig S3 and S4). As expected from previous NMR studies,^[44] some protein precipitation was observed during the titration of HP, which indicated a strong interaction with GAG. To avoid that the protein was completely precipitated, a concentration not higher than an equimolar ratio of IL-8 and HP had to be used. We monitored the interaction of HP with IL-8 by measuring CSPs from ^1H - ^{15}N HSQC spectra and mapped the CSPs to the IL-8 amino acid sequence (Fig S4A). In line with our previous NMR studies,^[44] the largest CSPs above a threshold of 0.02 ppm were found for residues in the N-loop (K20, H23, K25, K28) and the C-helix (K59, W62, V63, Q64, V66, V67, E68, K69, F70, A74, E75). This set of residues matches the modeled binding pose of HP dp6 that was observed in the MD simulation (Fig 1B), showing that most interactions of HP dp6 are formed with the N-loop and C-helix of IL-8.

Compared to HP dp6, the CSPs induced by HA dp6 were considerably smaller (<0.01 ppm) (Fig S4B), indicating that it binds weakly to IL-8.^[44] However, a similar CSP pattern was observed, with the highest changes found for residues in the $\beta 1$ strand (K28) and the end of the C-helix (A74, E75).

We next monitored the interaction of IL-8 with HP dp6 and HA dp6 using aminomethyl-PROXYL and carboxy-PROXYL probes, respectively, and compared the PRE rates of IL-8 measured in the presence and absence of GAG. The PROXYL-induced signal intensity changes relative to the diamagnetic

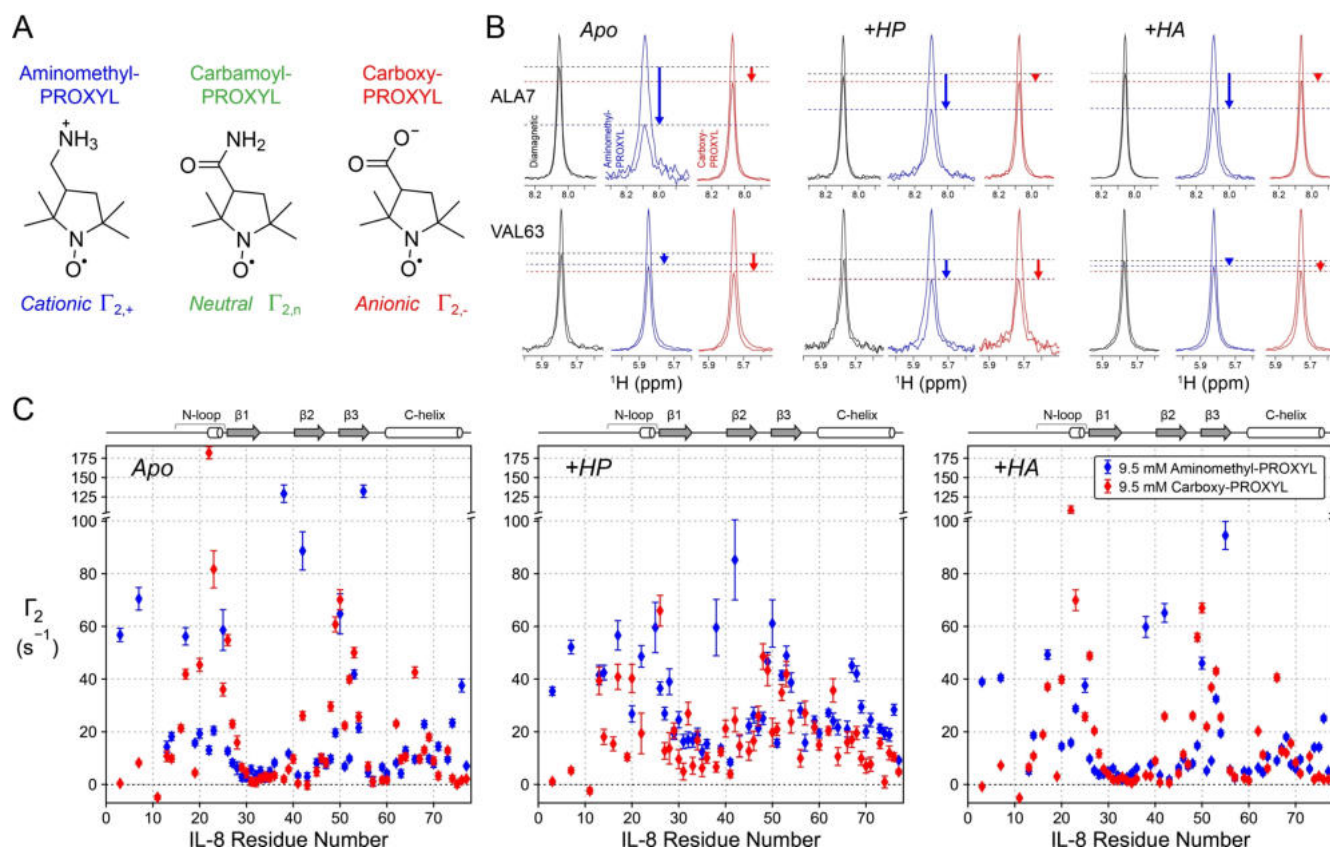


Figure 2. Paramagnetic NMR data used to analyze the electrostatic potentials around IL-8 and IL-8/GAG complexes. (A) Paramagnetic cosolute molecules used in this study. (B) Examples of signal intensity changes used to measure $^1\text{H}_\text{N}$ PRE rates of IL-8 apo protein and IL-8 in the presence of HP dp6 or HA dp6, respectively. The ^1H slices of the $^1\text{H}_\text{N}$ signals of A7 and V63 from the HSQC spectra of IL-8 at 0 ms and 10 ms relaxation measurement delays are shown. The ^1H signal slices in the absence (diamagnetic) or presence of 9.5 mM paramagnetic PROXYL derivative normalized to the intensity at 0 ms are displayed. Signal change differences observed in the presence of aminomethyl-PROXYL or carboxy-PROXYL compared to the diamagnetic case are indicated with blue or red arrows, respectively. (C) The PRE rates $\Gamma_{2,+}$ (blue) and $\Gamma_{2,-}$ (red) measured for the backbone $^1\text{H}_\text{N}$ nuclei in IL-8 in the absence and presence of HP dp6 or HA dp6, respectively, are plotted versus the IL-8 residue number. Error bars represent the Γ_2 rate uncertainties (see Experimental Section). The $^1\text{H}_\text{N}$ PRE rates of IL-8 are summarized in Table S1 in the supporting material.

reference were clearly different for several IL-8 residues upon addition of HP and HA. This is exemplary shown for the NMR signals of A7 and V63 in Fig 2B. Figure 2C displays the $\Gamma_{2,+}$ and $\Gamma_{2,-}$ rates along the IL-8 residue number for the cases without GAG, with HP, and with HA, respectively (see also Table S1). Comparing the GAG-free and HP-bound PRE datasets, the $\Gamma_{2,+}$ rates are clearly increased in the presence of HP for almost all residues in IL-8, as can be seen in Figure 2C and supporting Fig S5A. The $\Gamma_{2,-}$ rates on the other hand are less strongly increased by HP compared to the $\Gamma_{2,+}$ rates (Fig S5A). This result indicates that cationic aminomethyl-PROXYL molecules become more enriched around the IL-8/HP complex and that HP binding to IL-8 renders ϕ_ENS more negative as suggested by the APBS calculations (Figure 1B). Furthermore, the $\Gamma_{2,+}$ rates are most strongly increased for residues in the N-loop, β 1 strand, and C-helix (Fig S5A), which is line with the computationally predicted binding sites for HP dp6 (Figure 1B).

The PRE datasets for IL-8 and IL-8/HA are overall more similar. The $\Gamma_{2,+}$ and $\Gamma_{2,-}$ rates are not increased in the presence of HA, but are even slightly decreased (Fig S5A). This indicates that HA binding has only a small influence on the distribution of PROXYL molecules around IL-8 and consequently on ϕ_ENS .

This can also be seen when comparing the computationally predicted ϕ_ENS isosurfaces for IL-8 apo and IL-8/HA in Figure 1.

Interestingly, the PREs of a few isolated residues are largely decreased in the presence of both HP and HA. The $\Gamma_{2,+}$ rate of H38 and the $\Gamma_{2,-}$ rate of F22 are over 100 Hz in the absence of GAG and are reduced by more than 60 Hz upon addition of HP or HA (Fig S5). Similarly, the $\Gamma_{2,+}$ rates of L3, A7 and C55 are reduced by 20–30 Hz in the presence of HP and HA. A possible explanation could be that those residues in IL-8 undergo direct interactions with the PROXYL molecules, leading to their fast relaxation, and that HP and HA binding blocks those direct IL-8-PROXYL interactions, reducing the strong PRE effects.

In summary, we conclude based on the observed PRE changes that GAGs, in particular HP, profoundly affect the interaction of IL-8 with PROXYL molecules. In accordance with our APBS calculations, this supports the notion that HP renders ϕ_ENS more negative while HA has a comparatively small influence on ϕ_ENS .

NMR-determined near-surface electrostatic potentials for IL-8 and IL-8/GAG complexes

In order to more quantitatively analyze the effect of GAGs on the electrostatic potential of IL-8, we determined ϕ_{ENS} for IL-8 and IL-8/GAG complexes using the measured $\Gamma_{2,+}$ and $\Gamma_{2,-}$ rates (Figure 2C, Table S1) and applying equation 1 (see Experimental Section). We only considered statistically significant PRE rates and $\Gamma_{2,+}/\Gamma_{2,-}$ ratios (see Experimental Section). This procedure yielded ϕ_{ENS} values for 38 residues of IL-8, 41 residues of IL-8/HP dp6, and 41 residues of IL-8/HA dp6, respectively. Fig 3A shows ϕ_{ENS} values (black) for individual residues in IL-8 in the GAG-free case, and in the presence of either HP dp6 or HA dp6, respectively. For comparison with the experimental ϕ_{ENS} values, we calculated theoretical ϕ_{ENS} potentials using the APBS software^[55,56] and structural models of IL-8 or IL-8/GAG complexes generated by docking and subsequent MD simulation. From the APBS-computed potential at each grid point in the exterior of the protein structure, we calculated the theoretical ϕ_{ENS} potential of each $^1\text{H}_\text{N}$ nucleus of IL-8 by applying equations 1 and 2 (see Experimental section). The accessibility of each grid point was assessed by measuring the distance to the closest protein or GAG atom and comparing it to the sum of the van der Waals radius of the biomolecular atom and the accessibility radius of the paramagnetic PROXYL probe of 3.5 Å (Experimental Section). The theoretical ϕ_{ENS} profile plotted in Figure 3A (red) shows the average $\pm$ S.D. of ϕ_{ENS} over a 100 ns MD simulation. Overall, the experimental ϕ_{ENS} data agreed well with the simulated data, in particular for residues in secondary

structure regions. Correlation plots of experimental versus simulated ϕ_{ENS} values for IL-8, IL-8/HP, and IL-8/HA are shown in Fig 3B. The RMSD between experimental and simulated ϕ_{ENS} values in the structured regions was 10.1 mV (IL-8), 10.5 mV (IL-8/HP), and 9.5 mV (IL-8/HA). The correlation coefficients for secondary structure regions were 0.57 (IL-8), 0.40 (IL-8/HP), and 0.49 (IL-8/HA). The largest outliers were observed for residues in loop regions, e.g., for L3 and A7 in the N-terminus, H38 in the β 1- β 2 loop, and N76 in the C-terminus. The discrepancy between experimental and simulated ϕ_{ENS} data for the residues in loop regions is most likely due to increased conformational flexibility of those parts of the IL-8 structure. This is supported by the observations of Yu et al.^[31] who found in their paramagnetic NMR study on ubiquitin the largest outliers in loop regions. The authors attributed the discrepancy to increased conformational heterogeneity, which poses a challenge to structure-based electrostatic potential prediction. We also calculated the standard deviation of the predicted ϕ_{ENS} data using 100 structures equally sampled from the 100 ns MD simulation for each IL-8 or IL-8/GAG system. The standard deviation was highest in the N-terminus, the β 1- β 2 loop, and the β 3 strand (Figure 3A), which are regions of increased flexibility. This supports the notion that considering protein flexibility is critical for the prediction of the electrostatic potential. Improving the conformational sampling of the protein structure, e.g., by extending the time length of the MD simulation, could therefore potentially improve the agreement of the ϕ_{ENS} data with the experiment.

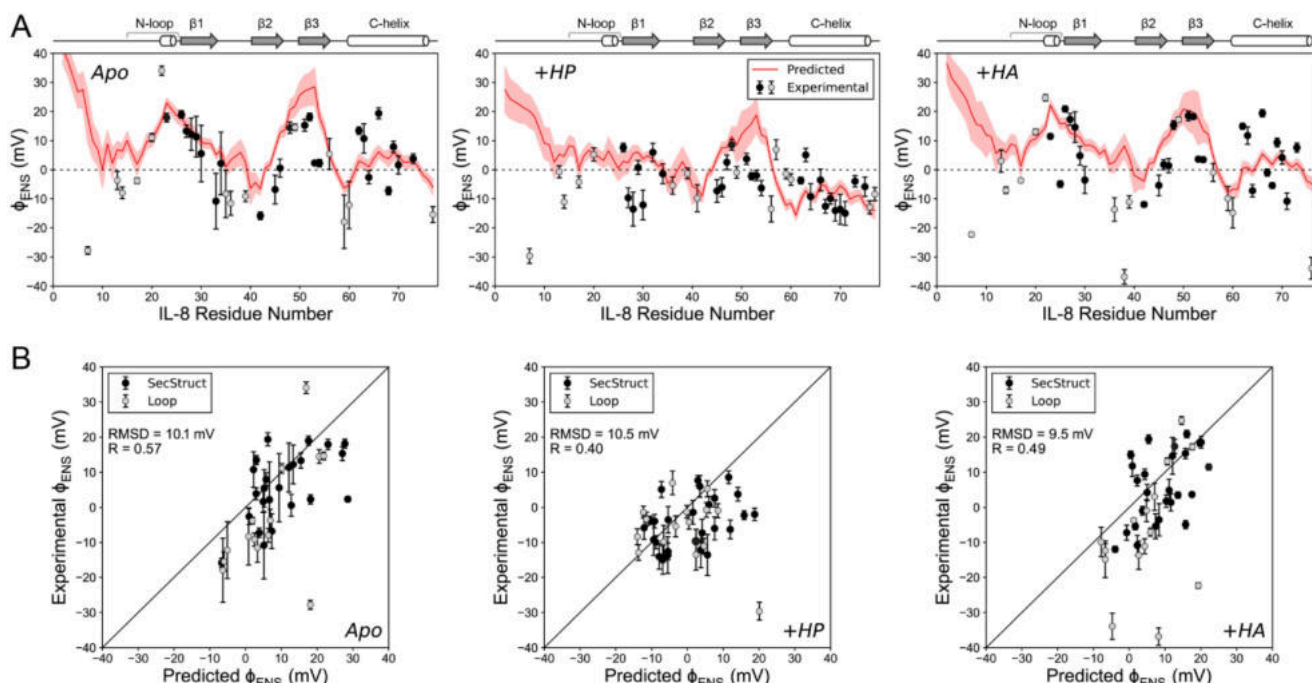


Figure 3. ENS electrostatic potentials (ϕ_{ENS}) determined from the $^1\text{H}_\text{N}$ PRE NMR data of IL-8 and IL-8/GAG complexes. (A) Experimentally determined (black and gray points) and predicted ϕ_{ENS} data (red line) for IL-8 apo (left), IL-8/HP (middle), and IL-8/HA (right) plotted along the residue number for IL-8. The predicted ϕ_{ENS} values represent the mean of the ENS potentials calculated from a 100 ns MD simulation of IL-8 or IL-8/GAG complexes. The red shaded region indicates the standard deviation of the ϕ_{ENS} predictions. Data points for residues located in secondary structure regions are shown in black and those for residues in loop regions are colored gray. (B) Correlation between experimental and predicted ϕ_{ENS} data. Error bars represent the experimental ϕ_{ENS} value uncertainties (see Experimental Section).

Comparing the ϕ_{ENS} data of IL-8 with those of IL-8/HP and IL-8/HA, some distinct differences can be noticed. The presence of HP leads to a reduction of ϕ_{ENS} across large parts of IL-8. The reduction is most significant in the N-loop, $\beta 1$ strand, $\beta 3$ strand, and the C-helix (middle plot in Fig 3A). This can be explained by the highly negatively charged character of HP dp6 ($q = -12e$) and by the fact that HP dp6 binds close to those regions in the IL-8 structure. As seen in Fig 1B, HP dp6 gives rise to an extensive area of negative potential around IL-8. This is in contrast to the IL-8/HA dp6 complex, which is displayed in Figure 1C. The ϕ_{ENS} profile (right plot in Figure 3A) is not significantly different from that of apo IL-8, and the spatial distribution of ϕ_{ENS} is not significantly perturbed. This is consistent with the small negative charge of HA dp6 of $-3e$.

In conclusion, while the interaction of IL-8 with HP dp6 can be clearly detected from the changes of ϕ_{ENS} using paramagnetic NMR, the interaction with HA dp6 cannot be deduced based on experimental ϕ_{ENS} measurements. This could indicate that the PROXYL method is more suitable for the detection of highly charged ligand interactions. In addition, we observed that the HP-induced ϕ_{ENS} changes around IL-8 were not limited to only a handful of amino acid residues, but occurred over multiple regions of the protein. This could be due to the high charge and large size of the HP dp6 molecule. While this effect could be helpful for the detection of biomolecular interactions, it also limits the site-selectivity of the structural information. Furthermore, while HP has the highest negative charge density of any known biological molecule, many small drug molecules are less strongly charged, which could imply possible limitations of the approach. It is an interesting question if drug interactions with proteins can also be studied with paramagnetic PROXYL probes. Thus, to further explore the capabilities and limitations of paramagnetic PROXYL probes for detecting protein-ligand interactions, we decided to test this approach on another group of ligands that were smaller and less strongly charged as HP.

Calculation of the electrostatic potential around Grb2 SH2 and Grb2 SH2/peptide complexes

As an additional test system we used the SH2 domain of Grb2 which recognizes phosphotyrosine peptides containing a pYXN recognition motif (where X can be any residue).^[52,53] Grb2 SH2 was prepared by expression in *E. coli* with a C-terminal His₆ tag for purification purposes (see Experimental Procedures). We tested three different small pYXN tripeptides, which carried in their middle position either a neutral (valine, V), positively charged (lysine, K), or negatively charged amino acid (glutamic acid, E). At physiological pH (7.4), these peptides have a net charge of ca. $-2e$ (pYVN), $-1e$ (pYKN), or $-3e$ (pYEN), respectively. The N- and C-termini were capped with acetyl and amide groups, respectively, to avoid any unspecific interactions due to the free termini. Based on the crystal structure of Grb2 SH2 bound to a longer PSpYVNQN peptide (PDB: 1JYR^[59]), we modeled the complex of Grb2 SH2 with each pYXN tripeptide using docking and MD simulation. Representative structural models from the MD simulation are displayed in Figure 4. Using the APBS software,^[55,56] we calculated ϕ_{ENS} for each Grb2 SH2/peptide complex and apo Grb2 SH2, which is mapped on the protein surface shown in Figure 4.

All pYXN peptides share a common binding mode. The phosphotyrosine side chain binds to a region with high positive ϕ_{ENS} formed by the side chains of R67 ($\alpha 1$ helix), R86 ($\beta 1$ strand), S88 ($\beta 1$ strand), K109 ($\beta 3$ strand), and the backbone HN of E89 ($\beta 1$ - $\beta 2$ loop). The asparagine side chain of pYXN is hydrogen-bonded to the backbone of K109 ($\beta 3$ strand) and capped by the side chain of Trp121 ($\beta 3$ - $\alpha 2$ loop) of Grb2 SH2. The middle residue X binds to a small pocket formed by Q106 ($\beta 3$ strand), F108 ($\beta 3$ strand), S141, and N143 (both in the loop after $\alpha 2$).

As can be seen in Fig 4, Grb2 SH2 exhibits an extended region of positive ϕ_{ENS} which is formed by several cationic residues (R67 ($\alpha 1$ helix), R86 ($\beta 1$ strand), K109 ($\beta 3$ strand)). This

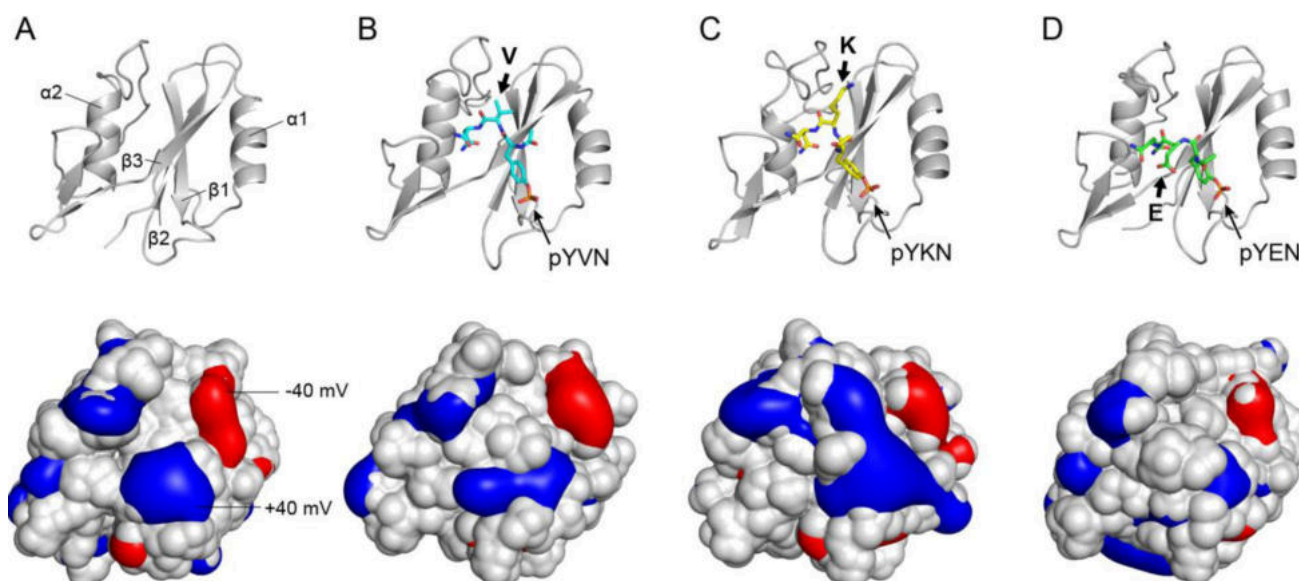


Figure 4. Simulated electrostatic potentials for Grb2 SH2 apo and Grb2 SH2/peptide complexes. (A) Grb2 SH2 apo. Secondary structure regions are labeled. (B) Grb2 SH2 with pYVN ($q = -2e$). (C) Grb2 SH2 with pYKN ($q = -1e$). (D) Grb2 SH2 with pYEN ($q = -3e$).

positive pocket binds the phosphotyrosine group of the pYXN peptides and is one of the structural determinants for the recognition of phospho-tyrosine peptides by Grb2 SH2. Another smaller region of negative ϕ_{ENS} is formed by the side chains of E71, E72 (both $\alpha 1$ helix), and D104 ($\beta 3$ strand), but is not part of the peptide binding pocket. The binding of pYVN and pYEN reduces slightly the size of the positive ϕ_{ENS} region (Figure 4B and D) and the binding of pYKN slightly increases its size, while the negative ϕ_{ENS} region is not significantly changed by any of the tripeptides. To monitor the interaction of pYXN peptides with Grb2 SH2 by NMR and experimentally measure ϕ_{ENS} around the protein as well as the ligand induced changes of ϕ_{ENS} , we next conducted NMR experiments with paramagnetic PROXYL probes.

Paramagnetic NMR data on Grb2 SH2 and Grb2 SH2/peptide complexes

We used ^{15}N -labeled Grb2 SH2 and unlabeled pYXN peptides. The resonance assignment of the ^1H - ^{15}N HSQC spectrum of Grb2 SH2 (see Fig S6) was transferred from previous NMR studies of Grb2 by Yuzawa et al.^[60] (BMRB-ID: 5693) and Sanches et al.^[61] (BMRB-ID: 27781) and confirmed with the help of a ^1H - ^{15}N HSQC-TOCSY experiment. The resonance assignment of the ^1H - ^{15}N HSQC spectrum of peptide-bound Grb2 SH2 was taken from Ogura et al.^[62] (BMRB-ID: 11055). We assured binding of pYXN peptides to Grb2 SH2 by performing titrations of Grb2 SH2 with the individual pYXN peptides and measuring CSPs. Figure S7 shows the series of HSQC spectra recorded during titration with pYVN (Fig S7A), pYKN (Fig S7B), and pYEN (Fig S7C). We observed CSPs (>0.2 ppm) for several residues, a disappearing of peaks at concentrations of 40–60 mol% peptide, and a reappearing of peaks at concentrations of ≥ 100 mol% peptide. This indicates that ligand binding happens in slow exchange on the NMR time scale and that pYXN peptides have a high binding affinity. Figure S8 shows the CSPs induced by pYVN (Fig S8A), pYKN (Fig S8B), and pYEN (Fig S8C) plotted along the Grb2 SH2 residue number. Mapping of the most significant CSPs on the molecular surface of Grb2 SH2 allowed unambiguous identification of the peptide binding region (right side in Fig S8). The CSP-derived binding regions are in perfect agreement with our docking and MD simulation derived models of the individual Grb2 SH2/peptide complexes (Figure 4). The largest CSPs were measured for residues that line the peptide binding pocket and are located in the $\beta 1$ – $\beta 2$ loop, $\beta 2$ strand, $\beta 3$ strand, and the $\beta 3$ – $\alpha 2$ loop.

For measuring $^1\text{H}_\text{N}$ PRE rates, we used 5-fold molar excess of peptide, at which the binding site on Grb2 SH2 is fully saturated, and either cationic aminomethyl-PROXYL ($\Gamma_{2,+}$) or neutral carbamoyl-PROXYL ($\Gamma_{2,n}$) at a concentration of 9.8 mM. We also tested anionic carboxy-PROXYL, but the quality of the Grb2 SH2 NMR spectrum deteriorated significantly at concentrations of carboxy-PROXYL above 3.5 mM. This prevented us from measuring significantly high PREs in the presence of carboxy-PROXYL. However, we observed pronounced signal intensity changes of several NMR signals of Grb2 SH2 in the

presence of 9.8 mM aminomethyl-PROXYL and carbamoyl-PROXYL, respectively. A comparison of the intensity of selected HSQC signals of Grb2 SH2 in the absence (apo) and presence of pYXN peptides is shown in Figure 5A and Fig S9. In the apo state, we observed strong PREs for residues that are part of or near the pYXN binding site, e.g., E89, S90 (both $\beta 1$ – $\beta 2$ loop), Q106, K109 (both $\beta 3$ strand), and W121 ($\beta 3$ – $\alpha 2$ loop). For S90 the HSQC signal is broadened beyond detection by the paramagnetic aminomethyl-PROXYL indicating a very high PRE rate. The small remaining signals of E89 and Q106 challenged the detection limit but allowed to determine a PRE rate (see also Fig S9). Furthermore, the HSQC signals of K109 and W121 are significantly reduced in their intensity by both aminomethyl-PROXYL and carbamoyl-PROXYL. The $\Gamma_{2,+}$ and $\Gamma_{2,n}$ rates calculated from the signal intensity changes are plotted along the Grb2 SH2 amino acid sequence in Figure 5B. The named HSQC signals reappeared upon addition of each pYXN peptide (Fig S9) and a reduction of the PREs of those residues was observed with both tested PROXYLs. This effect was the same for all three peptides – pYVN, pYKN, and pYEN (Figure 5B and S10) – and could indicate that the accessibility of the Grb2 SH2 surface for the PROXYL probes was blocked by the peptide ligands, thus increasing the protein-PROXYL distance and decreasing the PRE for residues within the peptide binding site. In contrast, residues that are outside of the peptide binding site showed no signal intensity modulating effect in the presence of pYXN peptides. This is exemplary shown for G93 in Figure 5A. The PRE rates of residues outside of the peptide binding site were almost the same in the free and peptide-bound state, indicating that the peptide ligand does not affect the accessibility of those residues to the PROXYL probes.

To aid the structural interpretation of the PRE data, we plotted them on the molecular surface of Grb2 SH2 shown in Fig 6 and Fig S11. We compared the strength and distribution pattern of $\Gamma_{2,+}$ and $\Gamma_{2,n}$ rates of the apo state (Figure 6A) with those of the pYVN-bound state (Figure 6B) as well as the other peptide-bound states (Fig S11). In addition, we calculated the PRE differences ($\Delta\Gamma_{2,+}$, $\Delta\Gamma_{2,n}$) between peptide-bound and apo states and analyzed their distribution patterns. Figure 6C displays the PRE differences versus the Grb2 SH2 residue number for the pYVN-bound state and Figure 6D shows protein surface mappings of the $\Delta\Gamma_{2,+}$ and $\Delta\Gamma_{2,n}$ rates. The PRE differences for all peptide-bound states and their protein surface mappings are shown in Fig S12 in the supplement. As can be seen in Fig 6 and Fig S12, PREs change most significantly within a narrow surface region of Grb2 SH2. This area (encircled by a red line in Figure 6) corresponds to the binding site of pYVN. Also for the pYKN- and pYEN-bound states, PREs are most strongly reduced in this surface region of the Grb2 SH2 structure (Fig S11 + S12). Intriguingly, the pattern of PRE changes resembles very closely the pattern of CSPs (Fig S8). These findings support our previous interpretation that the peptide ligand sterically blocks access of the PROXYL probes to the protein surface within a defined region, which corresponds to the ligand binding site. Thus, the PRE data provide precise information about the location of the pYXN peptide binding site. We suppose that this approach may be applicable also to

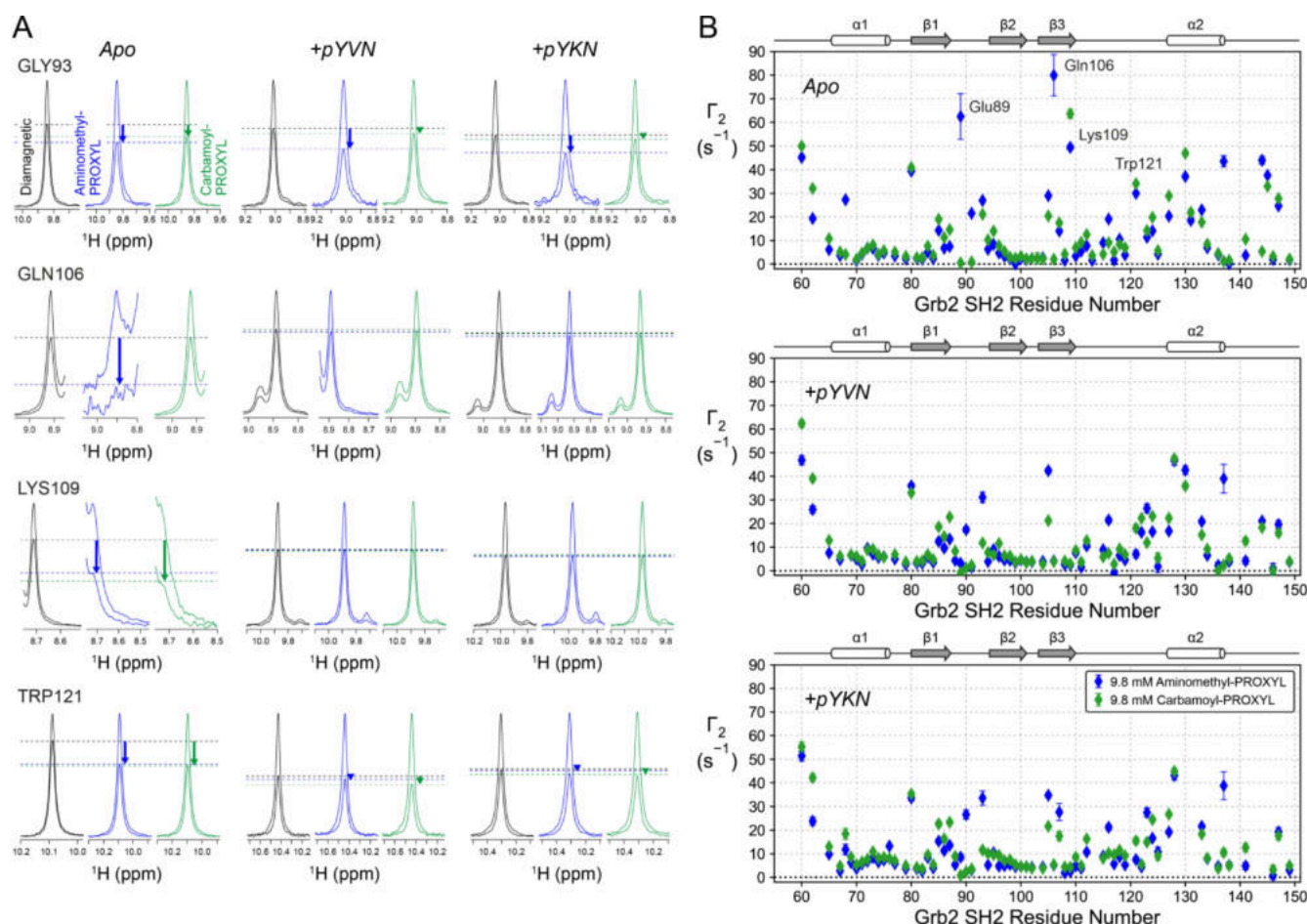


Figure 5. Paramagnetic NMR data used to analyze the electrostatic potentials around Grb2 SH2 and Grb2 SH2/peptide complexes. (A) Examples of signal intensity changes used to measure $^1\text{H}_\text{N}$ PRE rates of Grb2 SH2 apo protein and Grb2 SH2 in the presence of pYVN or pYKN peptide, respectively. The ^1H slices of the ^1H - ^{15}N HSQC signals of G93, Q106, K109, and W121 of Grb2 SH2 at 0 ms and 10 ms relaxation measurement delays are plotted. The ^1H signal slices in the absence (diamagnetic, black) or presence of 9.8 mM aminomethyl-PROXYL (paramagnetic, blue) or carbamoyl-PROXYL (paramagnetic, green) normalized to the intensity at 0 ms are displayed. Signal change differences observed in the presence of PROXYL derivatives relative to the diamagnetic case are indicated with arrows. (B) $^1\text{H}_\text{N}$ PRE rates Γ_{2+} (blue) and Γ_{2n} (green) of Grb2 SH2 in the absence and presence of pYVN or pYKN peptide, respectively, versus the Grb2 SH2 residue number. The signal intensity changes and $^1\text{H}_\text{N}$ PRE rates measured in the presence of pYEN peptide were very similar to those in the presence of pYVN or pYKN, and are shown in Fig S10. Error bars represent the Γ_2 rate uncertainties (see Experimental Section). All $^1\text{H}_\text{N}$ PRE rates of Grb2 SH2 are summarized in Table S2 in the supporting material.

other protein-ligand systems and that PRE data derived from soluble PROXYL probes can be in general helpful for the spatial mapping of ligand binding sites.

NMR-determined electrostatic potentials for Grb2 SH2 and Grb2 SH2/peptide complexes

To analyze how pYXN peptides affect the electrostatic potential of Grb2 SH2, we determined ϕ_ENS from the measured Γ_{2+} and Γ_{2n} rates (Figure 5B, Table S2) by applying equation 1. We obtained ϕ_ENS values for 57 residues of Grb2 SH2 in the apo state, and for 53, 56, and 51 residues of Grb2 SH2 in the pYVN-, pYKN-, and pYEN-bound state, respectively. Figure 7 shows the NMR derived ϕ_ENS data (black) for Grb2 SH2 and Grb2 SH2/peptide complexes. For comparison with the experimental ϕ_ENS values, we also calculated the theoretical ϕ_ENS values using the APBS computed potential grids and the MD-derived models of

Grb2 SH2 or Grb2 SH2/peptide complexes (Figure 4). The theoretical ϕ_ENS data (red) are plotted as average $\pm$ S.D. along the residue number and correlated with the experimental ϕ_ENS values in Figure 7. The agreement between experimental and theoretical data was fairly good, with RMSD values in secondary structure regions of 4.8 mV (Grb2 SH2), 6.2 mV (Grb2 SH2/pYVN), 6.5 mV (Grb2 SH2/pYKN), and 7.2 mV (Grb2 SH2/pYEN). The correlation coefficients were 0.70 (Grb2 SH2), 0.59 (Grb2 SH2/pYVN), 0.58 (Grb2 SH2/pYKN), and 0.66 (Grb2 SH2/pYEN). However, several outliers were observed for residues in loop regions, e.g., G93 (β_1 - β_2 loop), V105 (β_2 - β_3 loop), A115, G116, Y118, W121, V123 (all β_3 - α_2 loop), and S141 (tail after α_2). The discrepancy between experimental and predicted ϕ_ENS data could be explained by increased conformational flexibility for these parts of the Grb2 SH2 structure. In particular, the β_3 - α_2 loop showed increased flexibility in the MD simulation, resulting in standard deviations of ϕ_ENS of up to 10 mV for this loop. The lack of agreement would make it difficult to use ϕ_ENS values

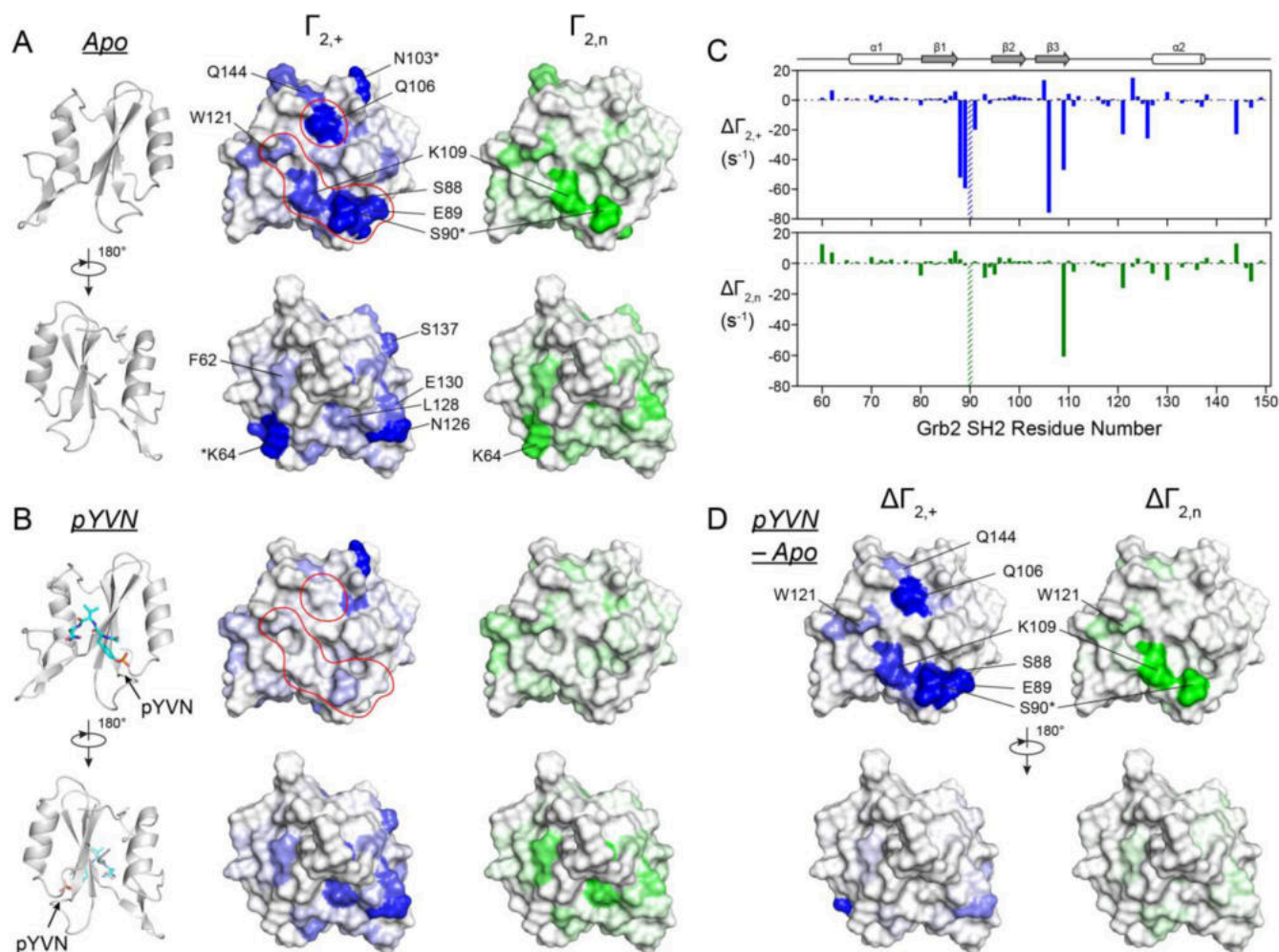


Figure 6. $^1\text{H}_\text{N}$ PRE rates measured for Grb2 SH2 in the presence of either aminomethyl-PROXYL ($\Gamma_{2,+}$) or carbamoyl-PROXYL ($\Gamma_{2,n}$) are plotted on the molecular surface of Grb2 SH2. (A) $\Gamma_{2,+}$ and $\Gamma_{2,n}$ rates for Grb2 SH2 apo protein. Residues with high PRE rates in the apo dataset are labeled. Residues whose signal was wiped out in the presence of PROXYLs are labeled with an asterisk (*). (B) $\Gamma_{2,+}$ and $\Gamma_{2,n}$ rates for Grb2 SH2 bound to pYVN. The region in which strong PRE rates disappear is framed with a red line. (C) Difference of PRE rates between apo and pYVN-bound state ($\Delta\Gamma_{2,+}$, $\Delta\Gamma_{2,n}$) are plotted versus the Grb2 SH2 residue number. The difference PRE rates are defined as: $\Delta\Gamma_{2,+} = \Gamma_{2,+}(\text{pYVN}) - \Gamma_{2,+}(\text{apo})$ and $\Delta\Gamma_{2,n} = \Gamma_{2,n}(\text{pYVN}) - \Gamma_{2,n}(\text{apo})$. The dashed bars at residue 90 indicate that this signal was wiped out by the PROXYL compounds in the apo state but not in the pYVN-bound state, indicating a significant decrease of the $\Gamma_{2,+}$ and $\Gamma_{2,n}$ rates. (D) The absolute values of $\Delta\Gamma_{2,+}$ and $\Delta\Gamma_{2,n}$ are plotted on the surface of Grb2 SH2. Residues with high difference PRE rates are labeled. The strength of PRE rates is indicated with a color gradient. $\Gamma_{2,+}$ or $\Delta\Gamma_{2,+}$: 0 Hz (white) – 60 Hz (blue), $\Gamma_{2,n}$ or $\Delta\Gamma_{2,n}$: 0 Hz (white) – 60 Hz (green). Grb2 SH2 is shown from two different perspectives. Additional surface mappings of Γ_2 and $\Delta\Gamma_2$ rates measured for Grb2 SH2 in the presence of pYKN or pYEN are shown in supporting Figures S11 and S12.

quantitatively as restraints for structure prediction tasks, by evaluating the match between experimental and back-calculated ϕ_ENS data. However, further studies are still needed to assess the predictive power of ϕ_ENS data for modeling protein-ligand interactions and establish best practices how to translate ϕ_ENS data into structural information. One possibility could be to obtain information on the location of the ligand binding site by comparing the ϕ_ENS data collected for the ligand-bound and ligand-free states, similarly to our analysis of the PRE differences. As demonstrated in the previous section, the location of the largest $\Delta\Gamma_2$ rates could be used to precisely narrow down the location of the pYXN binding site in Grb2 SH2 (Figure 6). This kind of information can be very useful as a starting point for modeling of a protein-ligand complex. Looking at the ϕ_ENS curve of Grb2 SH2 and comparing it with those of the Grb2

SH2/peptide complexes (Figure 7), some small but clear differences can be observed. In Figure 8 only the section of the ϕ_ENS curve from K100 to L120 is plotted. It can be seen that the ϕ_ENS curve is shifted to smaller ϕ_ENS values for each pYXN peptide compared to the case without peptide. The downshift of the ϕ_ENS curve is most pronounced for pYVN and pYEN peptides, which carry net charges of -2e and -3e , respectively, and less pronounced for pYKN, which has a more positive charge of -1e . This finding suggests that the pYXN peptides change the ϕ_ENS potential locally in a narrow region of Grb2 SH2. This region corresponds to the β_3 strand and the first loop after β_3 . Both parts of the Grb2 SH2 structure establish direct contacts with the peptide ligand, which could explain why ϕ_ENS is changed in this region.

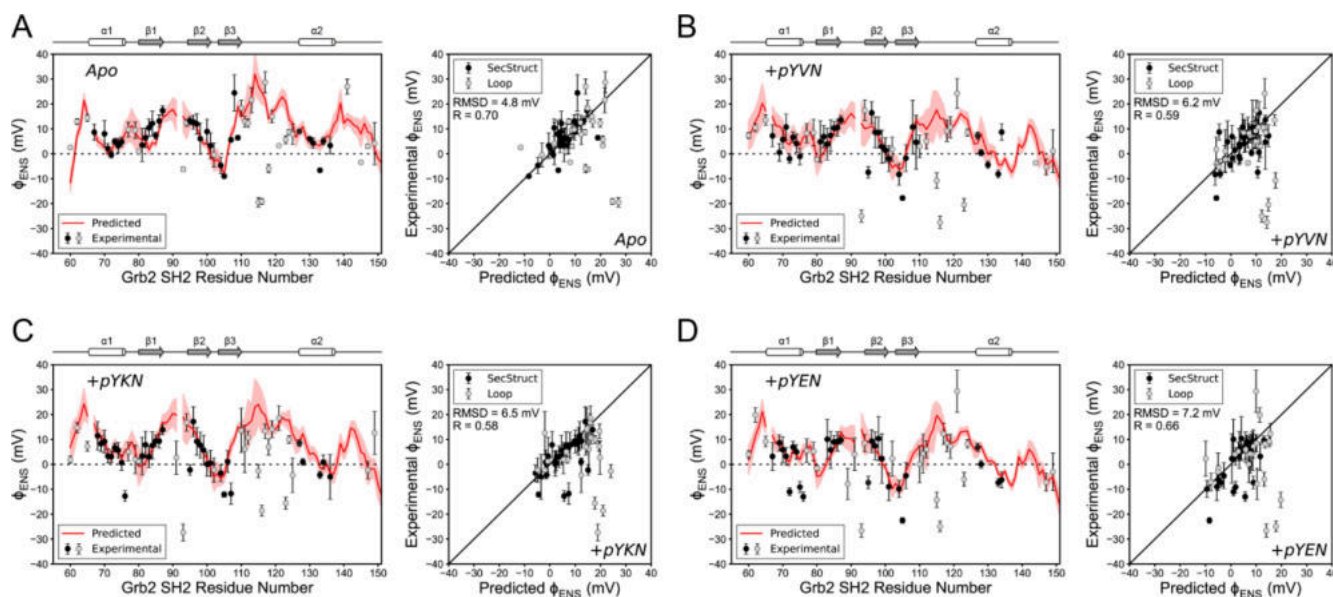


Figure 7. ENS electrostatic potentials determined from the $^1\text{H}_\text{N}$ PRE NMR data of Grb2 SH2 and Grb2 SH2/peptide complexes. Experimentally determined (black and gray points) and predicted ENS potentials (red line) versus residue number plots and experimental versus predicted ϕ_ENS data plots for (A) Grb2 SH2 apo, (B) Grb2 SH2 with pYVN, (C) Grb2 SH2 with pYKN, and (D) Grb2 SH2 with pYEN. The predicted ϕ_ENS values represent the mean of the ENS potentials calculated from a 100 ns MD simulation of Grb2 SH2 or Grb2 SH2/peptide complexes. The red shaded area indicates the standard deviation of the ϕ_ENS predictions. Data points for residues located in secondary structure regions are shown in black and those for residues in loop regions are shown in gray. Error bars represent the experimental ϕ_ENS value uncertainties (see Experimental Section).

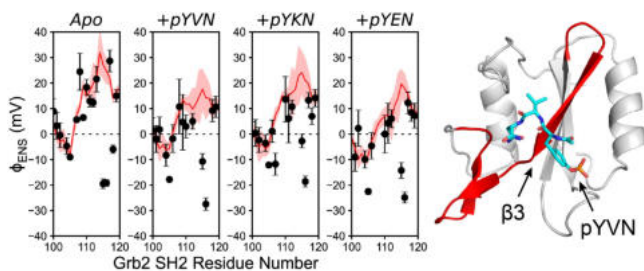


Figure 8. ENS electrostatic potential for residues 100 to 120 in Grb2 SH2 is reduced by the presence of pYXN peptides – pYVN ($q = -2e$), pYKN ($q = -e$), pYEN ($q = -3e$). Left: Experimentally determined (black points) and predicted ENS potentials (red line) versus residue number. Right: Molecular model of Grb2 SH2 with pYVN. The ribbon segment corresponding to residues 100–120 is colored red.

Together, our data show that the ϕ_ENS potential of Grb2 SH2 can be measured using paramagnetic NMR in good agreement with theoretical predictions. The effect of pYXN peptides on ϕ_ENS is localized to a narrow region on the protein. This kind of information could be useful for ligand binding mode discovery.

Conclusions

Paramagnetic NMR methods represent an important NMR tool employed in drug discovery and drug development. They can provide structural restraints for modeling the structure of protein-ligand complexes, even for weakly interacting ligands. However, the covalent labeling of proteins with paramagnetic tags or the generation of metal ion-tagged ligands is complicated and bears the risk of perturbing the protein

structure and/or ligand binding mode. Experiments with paramagnetic cosolutes offer a promising alternative because they require no chemical modification of protein or ligand. Instead, the paramagnetic probes can be directly added to the protein solution. The PROXYL probes used in the present study have small size, which lowers the risk of unspecific binding to the protein. Moreover, the differently charged sidechains of the PROXYL probes influence their distribution around the protein, which offers flexibility in tuning the size of the solvent PRE effect.

Using two test systems, the proteins IL-8 and Grb2 SH2, we have evaluated the utility of the PROXYL method for mapping ligand binding sites in proteins and for measuring the effect of ligand binding on the electrostatic potential of proteins. Our study extends and completes previous investigations, which reported the *de novo* determination of ϕ_ENS for proteins^[31,63] and DNA^[33] as well as studies, which mapped interfaces in protein-protein complexes using soluble, Gd^{3+} -chelating paramagnetic agents.^[29,30]

Our NMR-determined ϕ_ENS data for IL-8 and two IL-8/GAG complexes show that a highly negatively charged HP ligand reduced the ϕ_ENS of IL-8, in agreement with theoretical predictions from APBS calculations, while a less charged HA ligand had only a negligible effect on ϕ_ENS .

The results for Grb2 SH2 and three Grb2 SH2/peptide complexes show that the PROXYL-induced PRE data allow a precise localization of the peptide binding region. The location of the peptide binding region obtained by the PROXYL method matches the region inferred from CSP data and agrees well with the crystallographic binding pose of phosphotyrosine peptides on Grb2 SH2.^[59] Furthermore, the effect on ϕ_ENS was localized to

a narrow region in Grb2 SH2, as expected from the small size of the pYXN tripeptide ligands.

Information on how interacting ligand and drug molecules shape ϕ_{ENS} is important for understanding the protein-ligand interaction mechanism and could give insights into the biological responses that are triggered by the ligand recognition process. Our results show that experiments with soluble, paramagnetic PROXYL probes can be a useful addition to the toolbox of NMR methods for studying protein-ligand interactions. NMR-determined ϕ_{ENS} data could be useful for comparison with computational predictions of protein-ligand electrostatics-driven interactions. In addition, PRE data-derived ligand binding sites can serve as general constraints to guide protein-ligand structure calculations or can be used as filters to validate generated protein-ligand models.

Experimental Section

Proteins and other materials: ^{15}N -labeled IL-8(1-77) (excluding the first 22-amino acid signal peptide from the Uniprot sequence P10145) was expressed in *E. coli* strain ER2566 and purified as described by Pichert et al.^[44] Please note the difference of -5 residues in the nomenclature when comparing to IL-8(1-72). ^{15}N -labeled Grb2 SH2 domain (comprising residues M55–D150 of Grb2 from Uniprot sequence P62993; cloned in pET21b(+)) plasmid) was expressed in *E. coli* strain Rosetta (DE3) at 20°C for 16 hours after induction with 0.4 mM IPTG and was purified by Ni-NTA affinity chromatography and size exclusion chromatography as described in the Supporting Information. The PROXYL derivatives 3-amino-methyl-PROXYL, 3-carboxy-PROXYL, and 3-carbamoyl-PROXYL were purchased from Sigma-Aldrich. The phospho-peptides Ac-pTyr-Val-Asn-NH₂, Ac-pTyr-Glu-Asn-NH₂, and Ac-pTyr-Lys-Asn-NH₂ were synthesized by the Leipzig University peptide synthesis core unit using standard F-moc solid phase synthesis strategy. Peptide identity was confirmed by MALDI-TOF mass spectrometry. Heparin (HP) and hyaluronan (HA) hexasaccharides (dp6, where dp stands for the degree of polymerization) were purchased from Iduron (Manchester, UK).

NMR sample preparation: NMR samples of ^{15}N -labeled IL-8 contained 0.33 mM protein in IL-8 sample buffer (20 mM Na-phosphate (pH 7.0), 40 mM NaCl, 10 mM DMSO, and 5% D₂O). NMR samples of IL-8 with HP contained 0.40 mM protein and 0.40 mM HP dp6 in IL-8 sample buffer, and NMR samples of IL-8 with HA contained 0.29 mM protein and 1.0 mM HA dp6. Samples were prepared by stepwise titration of IL-8 with GAGs and the titration was followed by recording a ^1H - ^{15}N HSQC spectrum for each step. The paramagnetic NMR samples of IL-8 with PROXYLs contained 9.5 mM of either 3-aminomethyl-PROXYL or 3-carboxy-PROXYL. The diamagnetic IL-8 sample was prepared by adding sample buffer containing no PROXYLs to achieve the same concentration of IL-8 or IL-8+GAGs as in the case of the paramagnetic sample. NMR samples of Grb2 SH2 domain contained 0.24 mM protein in Grb2 SH2 sample buffer (20 mM Tris-HCl (pH 7.2), 100 mM NaCl, 10 mM DMSO, and 5% D₂O). NMR samples of Grb2 SH2 domain with phospho-peptides contained 0.24 mM protein and 1.2 mM of either Ac-pYVN-NH₂, Ac-pYEN-NH₂, or Ac-pYKN-NH₂ peptide, respectively. The paramagnetic NMR samples of Grb2 SH2 with or without phospho-peptides contained 9.8 mM 3-aminomethyl-PROXYL or 3-carbamoyl-PROXYL, respectively. The diamagnetic control sample was prepared by adding sample buffer containing no PROXYLs until the same concentration of Grb2 SH2 or Grb2 SH2+peptide as in the case of the paramagnetic sample was reached. The preparation

and quantification of the PROXYL stock solutions is described in the Supporting Information.

NMR experiments: The binding of GAGs to IL-8 and of phospho-peptides to Grb2 SH2 during titrations with ligands were followed by recording ^1H - ^{15}N HSQC spectra.^[64] Spectra were acquired typically with spectral widths of 16 ppm for ^1H and 38 ppm for ^{15}N as well as 128 – 256 data points in the indirect dimension. The NMR resonance assignment of the ^1H - ^{15}N HSQC spectrum of IL-8 was taken from Pichert et al.^[44] The resonance assignment of the ^1H - ^{15}N HSQC spectrum of the apo form of Grb2 SH2 was transferred from Yuzawa et al.^[60] (BMRB-ID: 5693) and Sanches et al.^[61] (BMRB-ID: 27781). NMR signal assignments were confirmed with the help of a $3\text{D } ^1\text{H}$ - ^{15}N HSQC-TOCSY spectrum of apo Grb2 SH2 acquired at a ^1H field strength of 700 MHz and a TOCSY mixing time of 60 ms . The resonance assignment of the ^1H - ^{15}N HSQC spectrum of peptide-bound Grb2 SH2 was transferred from Ogura et al.^[62] (BMRB-ID: 11055). The PRE rates of the ^1H transverse magnetizations (Γ_2) of the protein $^1\text{H}_\text{N}$ nuclei were measured using a ^1H - ^{15}N HSQC experiment with a spin echo sequence implemented during the first INEPT period.^[65] The two time-point approach^[65] with a 10-ms difference was chosen and the two spectra were recorded in an interleaved manner. For each protein and ligand, the PRE experiments were conducted with three samples: one diamagnetic reference sample, containing no PROXYL derivative, and two paramagnetic samples, one containing 3-amino-methyl-PROXYL, and the other one containing either 3-carboxy-PROXYL or 3-carbamoyl-PROXYL. Uncertainties in Γ_2 rates were estimated using an error propagation procedure.^[65] All NMR experiments on IL-8 were performed at 30°C using a Bruker NEO 700 MHz spectrometer equipped with a 5 mm E-free HCN triple resonance probe. All NMR experiments on Grb2 SH2 were performed at 25°C using a Bruker Avance III 600 MHz spectrometer equipped with a 5 mm inverse triple resonance probe with z -gradient. The Bruker software TopspinTM was used for data acquisition and spectrum processing. Analysis of the NMR spectra and quantification of the NMR signal intensities for PRE measurements was carried out using the NMRFAM-SPARKY software.^[66]

Determination of NMR-based near-surface electrostatic potentials: The ENS electrostatic potentials ϕ_{ENS} of individual $^1\text{H}_\text{N}$ nuclei were determined from the Γ_2 rates using equation (1), which was derived by Yu et al.^[31]

$$\phi_{ENS} = \frac{k_B T}{(z_b - z_a)e} \ln \left(\frac{\Gamma_{2,a}}{\Gamma_{2,b}} \right) \quad (1)$$

where k_B is the Boltzmann constant; T is temperature; e is the elementary charge; z is a charge valence of a PROXYL derivative; and suffixes a and b refer to the two PROXYL derivatives. Since a neutral pH was used for the PRE measurements, $z = +1$ for 3-aminomethyl-PROXYL, $z = 0$ for 3-carbamoyl-PROXYL, and $z = -1$ for 3-carboxy-PROXYL. The uncertainties in ϕ_{ENS} were estimated using an error propagation protocol as described in Yu et al.^[31] We used the selection criteria described in Yu et al.^[31] to use only statistically significant Γ_2 rates and $\Gamma_{2,a}/\Gamma_{2,b}$ rates for calculation of ϕ_{ENS} .

Molecular structures: The structure of dimeric IL-8 was obtained from the Protein Data Bank (PDB ID: 1IL8, NMR structure).^[54] Missing N-terminal residues were added in the xLeap module of AMBER20.^[67] The structure of Grb2 SH2 was obtained from the PDB (PDB ID: 1JYR, $1.55\text{ \AA}$).^[59] An eight-residue long histidine tag (LEH₆) was added in xLeap to the protein sequence to match the exact sequence of the protein used in the presented NMR experiments. The GAG ligand structures were extracted from the PDB (PDB ID: 1HPN^[68] and 2BVK^[69] NMR structures), for HP dp6 and HA dp6,

respectively. Three tripeptide ligands Ac-pYXN-NH₂ (X=V, E, or K) containing a phosphotyrosine residue (pY), were constructed based on the peptide ligand from (PDB ID: 1JYR, 1.55 Å).^[59]

Molecular docking: Molecular docking was performed using Autodock3.^[70] The docking simulation of IL-8/HP and IL-8/HA complexes used a grid box of size 60 Å×60 Å×60 Å with a grid step of 0.375 Å, centered on the two α-helices of dimeric IL-8. The docking of each complex was carried out in 100 independent runs using the Lamarckian Genetic Algorithm for 10⁵ generations, an initial population size of 300, and 9995 × 10⁵ energy evaluations. The 50 docking poses with the best Autodock3-score were clustered using the DBSCAN algorithm.^[71] Ten cluster representatives per IL-8/GAG complex with the best Autodock3-score were used for subsequent analyses.

Molecular dynamics simulations: All MD simulations were carried out in AMBER^[67] using the ff14SBonlySC^[72] force field parameters for proteins, and GLYCAM06^[73] force field parameters for GAG ligands. The starting structures of IL-8 and Grb2 SH2 were obtained from the aforementioned PDB files. The cluster representatives previously identified for the IL-8/GAG docking models were used as starting structures of the MD simulations of the IL-8/GAG complexes, resulting in 10 MD systems per complex. The IL-8/HP and IL-8/HA systems were solvated using a truncated octahedron TIP3P periodic box with 15.0 Å distance between the complex and the box boundaries. The starting structures of the Grb2 SH2/peptide complexes were obtained by substituting the ligand in the PDB 1JYR with the phosphotyrosine-containing Ac-pYXN-NH₂ (X=V, E, K) ligands. The ff14SBonlySC^[72] force field parameters in AMBER were used, together with appropriate parameters for the phosphorylated tyrosine residues.^[74] Each structure was placed in a TIP3P water box whose boundaries were at a 15.0 Å distance from the complex atoms.

All analyzed systems were neutralized by adding Na⁺/Cl⁻ counterions prior to the MD simulations. A two-step energy minimization protocol was used, where the first step contained 0.5×10³ steepest descent cycles and 10³ conjugate cycles with harmonic force restraints on solute atoms (100 kcal/mol/Å²), and the second step contained 3×10³ steepest descent cycles and 3×10³ conjugate gradient cycles without restraints. The system was then heated up to 300 K for 10 ps using a Langevin thermostat (collision frequency 1 ps⁻¹), followed by an equilibration step at 300 K in an isothermal isobaric ensemble for 100 ns. The subsequent MD production run was carried out for 100 ns, employing the Particle Mesh Ewald method for the computation of long-range electrostatic interactions and the SHAKE algorithm for constraining all covalent bonds involving hydrogen atoms. The resulting trajectories were visualized in VMD.^[75] The cpptraj module from AMBER was used for the analysis of the obtained trajectories.^[76]

All frames of the MD trajectories were analyzed using Molecular Mechanics-Generalized Born Surface Area (MM-GBSA) in AMBER20 with igb=2.^[77] Out of the ten trajectories obtained per IL-8/GAG complex, a representative MD simulation was chosen based on the lowest binding free energy value, followed by visual inspection and analysis of the obtained trajectories in the context of known binding poses of HP in complex with IL-8.

Poisson-Boltzmann equation-based calculation of electrostatic potentials: For each MD trajectory of the IL-8/GAG and Grb2 SH2/peptide complexes, 100 configurations equally spaced throughout the 100 ns trajectory were extracted and converted to individual PQR files using the xLeap module of AMBER^[67] and the Open Babel chemical toolbox.^[78] The Adaptive Poisson-Boltzmann Solver (APBS) software^[55,56] was used to create the electrostatic potential surface maps. Additional electrostatic potential maps were obtained for the

apo structures of IL-8 and Grb2 SH2. Atomic radii and charges were assigned using the AMBER ff14SBonlySC^[72] force field parameters. Each system was placed in a 144 Å×144 Å×144 Å box with a spacing of 0.5 Å, resulting in 289×289×289 grid points. The monovalent ion concentration was set to 70 mM or 100 mM, respectively, to reflect the actual ionic strength in the IL-8 or Grb2 SH2 sample buffers. The solvent van der Waals radius was set to 1.4 Å and the ionic radius was fixed at 2.0 Å. The dielectric constant of solvent and protein interior was set to 78 and 2, respectively. The theoretical ENS electrostatic potentials φ_{ENS} of individual ¹H_N nuclei were calculated from the electrostatic potential values at all grid points which were outside of the space occupied by protein and ligand using equations (1) and (2):

$$\frac{r_{2,a}}{r_{2,b}} = \frac{\sum_i a_i r_i^{-6} \exp\left(\frac{U_{a,i}}{k_B T}\right)}{\sum_i a_i r_i^{-6} \exp\left(\frac{U_{b,i}}{k_B T}\right)} \quad (2)$$

where i is the index of a grid point, U_i is the electrostatic potential at the grid point i , r_i is the distance between that grid point and a ¹H_N nucleus, and a_i is the accessibility of that grid point (0 for inaccessible grid points, 1 for accessible grid points). The accessibility of a grid point for the paramagnetic PROXYL probe was assessed by comparing the van der Waals radii of protein and ligand atoms to an empirically determined accessibility radius of 3.5 Å for the PROXYL probe. This value was found optimal by Yu et al.^[31] in yielding the best agreement between experimental and predicted φ_{ENS} data. Thus, grid points that were closer to a protein or ligand atom than the sum of the atom van der Waals radius and PROXYL probe accessibility radius were considered to lie in the inaccessible interior of the macromolecule, and otherwise in the accessible exterior. The computation of the φ_{ENS} data from the APBS output grids was conducted using a Python script.

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (Project numbers TRR-386/A2 and TRR-386/B2 to GK) and the National Science Centre of Poland (Narodowe Centrum Nauki, grant number UMO-2018/31/G/ST4/00246 to SS, AD and MG). AP and DH acknowledge DFG funding through project 426256511. We acknowledge further funding from the Open Access Publishing Funds of Leipzig University supported by the German Research Foundation within the program Open Access Publication Funding. We thank the Polish high-performance computing infrastructure PLGrid (HPC Centers: ACK Cyfronet AGH) for providing computer facilities and support within computational grant no. PLG/2020/014275. Open Access funding enabled and organized by Projekt DEAL.

Conflict of Interests

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available in the supplementary material of this article.

Keywords: NMR spectroscopy · protein-ligand interactions · nitroxides · solvent paramagnetic relaxation enhancements · molecular dynamics

- [1] A. Furukawa, T. Konuma, S. Yanaka, K. Sugase, *Prog. Nucl. Magn. Reson. Spectrosc.* **2016**, *96*, 47–57.
- [2] W. Becker, K. C. Bhattacharjee, N. Gubensäk, K. Zangger, *ChemPhysChem* **2018**, *19*, 895–906.
- [3] Y. Mizukoshi, K. Takeuchi, Y. Tokunaga, H. Matsuo, M. Imai, M. Fujisaki, H. Kamoshida, T. Takizawa, H. Hanzawa, I. Shimada, *Sci. Adv.* **2020**, *6*, eabd0480.
- [4] E. Skeens, G. P. Lisi, *Methods* **2023**, *209*, 40–47.
- [5] M. Mayer, B. Meyer, *J. Am. Chem. Soc.* **2001**, *123*, 6108–6117.
- [6] C. B. Post, *Curr. Opin. Struct. Biol.* **2003**, *13*, 581–588.
- [7] C. Dalvit, G. Fogliatto, A. Stewart, M. Veronesi, B. Stockman, *J. Biomol. NMR* **2001**, *21*, 349–359.
- [8] K. J. Yong, T. M. Vaid, P. J. Shilling, F.-J. Wu, L. M. Williams, M. Deluigi, A. Plückthun, R. A. D. Bathgate, P. R. Gooley, D. J. Scott, *ACS Chem. Biol.* **2018**, *13*, 1090–1102.
- [9] S. Barelier, D. Linard, J. Pons, A. Clippe, B. Knoops, J.-M. Lancelin, I. Krimm, *PLoS One* **2010**, *5*, e9744.
- [10] S. Igonet, C. Raingeval, E. Cecon, M. Pučić-Baković, G. Lauc, O. Cala, M. Baranowski, J. Perez, R. Jockers, I. Krimm, A. Jawhari, *Sci. Rep.* **2018**, *8*, 8142.
- [11] M. F. Mesleh, W. A. Shirley, C. E. Heise, N. Ling, R. A. Maki, R. P. Laura, *J. Biol. Chem.* **2007**, *282*, 6338–6346.
- [12] P. M. Nieto, *Front. Mol. Biosci.* **2018**, *5*, 33.
- [13] C. A. Softley, M. J. Bostock, G. M. Popowicz, M. Sattler, *J. Biomol. NMR* **2020**, *74*, 287–309.
- [14] M. John, G. Pintacuda, A. Y. Park, N. E. Dixon, G. Otting, *J. Am. Chem. Soc.* **2006**, *128*, 12910–6.
- [15] T. Zhuang, H.-S. Lee, B. Imperiali, J. H. Prestegard, *Protein Sci.* **2008**, *17*, 1220–1231.
- [16] J. Y. Guan, P. H. Keizers, W. M. Liu, F. Lohr, S. P. Skinner, E. A. Heeneman, H. Schwalbe, M. Ubbink, G. Siegal, *J. Am. Chem. Soc.* **2013**, *135*, 5859–68.
- [17] G. Kuenze, S. Köhling, A. Vogel, J. Rademann, D. Huster, *J. Biol. Chem.* **2016**, *291*, 3100–3113.
- [18] W. N. Chen, C. Nitsche, K. B. Pilla, B. Graham, T. Huber, C. D. Klein, G. Otting, *J. Am. Chem. Soc.* **2016**, *138*, 4539–46.
- [19] K. Zimmermann, D. Joss, T. Müntener, E. S. Nogueira, M. Schäfer, L. Knörr, F. W. Monnard, D. Häussinger, *Chem. Sci.* **2019**, *10*, 5064–5072.
- [20] U. Brath, S. I. Swamy, A. X. Veiga, C.-C. Tung, F. Van Petegem, M. Erdélyi, *J. Am. Chem. Soc.* **2015**, *137*, 11391–11398.
- [21] Á. Canales, A. Mallagaray, M. Á. Berbis, A. Navarro-Vázquez, G. Domínguez, F. J. Cañada, S. André, H.-J. Gabius, J. Pérez-Castells, J. Jiménez-Barbero, *J. Am. Chem. Soc.* **2014**, *136*, 8011–8017.
- [22] S. Köhling, G. Kuenze, K. Lemmnitz, M. Bermudez, G. Wolber, J. Schiller, D. Huster, J. Rademann, *Chemistry* **2016**, *22*, 5563–5574.
- [23] C. Hartlmüller, J. C. Günther, A. C. Wolter, J. Wöhrner, M. Sattler, T. Madl, *Sci. Rep.* **2017**, *7*, 5393.
- [24] K.-Y. Lee, Z. Fang, M. Enomoto, G. Gasmi-Seabrook, L. Zheng, S. Koide, M. Ikura, C. B. Marshall, *Angew. Chem. Int. Ed.* **2020**, *59*, 11037–11045.
- [25] C. Hartlmüller, C. Göbl, T. Madl, *Angew. Chem. Int. Ed.* **2016**, *55*, 11970–11974.
- [26] Y. Wang, C. D. Schwieters, N. Tjandra, *J. Magn. Reson.* **2012**, *221*, 76–84.
- [27] C. Hartlmüller, E. Spreitzer, C. Göbl, F. Falsone, T. Madl, *J. Biomol. NMR* **2019**, *73*, 305–317.
- [28] S. Arumugam, C. L. Hemme, N. Yoshida, K. Suzuki, H. Nagase, M. Berjanskii, B. Wu, S. R. Van Doren, *Biochemistry* **1998**, *37*, 9650–9657.
- [29] T. Madl, T. Güttler, D. Görlich, M. Sattler, *Angew. Chem. Int. Ed.* **2011**, *50*, 3993–3997.
- [30] J. H. Tomlinson, G. S. Thompson, A. P. Kalverda, A. Zhuravleva, A. J. O'Neill, *Sci. Rep.* **2016**, *6*, 19524.
- [31] B. Yu, C. C. Pletka, B. M. Pettitt, J. Iwahara, *Proc. Natl. Acad. Sci. USA* **2021**, *118*, DOI 10.1073/pnas.2104020118.
- [32] B. Yu, C. C. Pletka, J. Iwahara, *Proc. Natl. Acad. Sci. USA* **2021**, *118*, DOI 10.1073/pnas.2015879118.
- [33] B. Yu, X. Wang, J. Iwahara, *J. Phys. Chem. Lett.* **2022**, *13*, 10025–10029.
- [34] P. Kukić, J. E. Nielsen, *Future Med. Chem.* **2010**, *2*, 647–666.
- [35] R. Alsallaq, H.-X. Zhou, *Proteins* **2008**, *71*, 320–335.
- [36] A. Voet, F. Berenger, K. Y. J. Zhang, *PLoS One* **2013**, *8*, e75762.
- [37] M. R. Bauer, M. D. Mackey, *J. Med. Chem.* **2019**, *62*, 3036–3050.
- [38] K. Rajarathnam, B. D. Sykes, C. M. Kay, B. Dewald, T. Geiser, M. Baggiolini, I. Clark-Lewis, *Science* **1994**, *264*, 90–92.
- [39] S. T. Das, L. Rajagopalan, A. Guerrero-Plata, J. Sai, A. Richmond, R. P. Garofalo, K. Rajarathnam, *PLoS One* **2010**, *5*, e11754.
- [40] A. Pichert, D. Schlorke, S. Franz, J. Arnhold, *Biomatter* **2012**, *2*, 142–148.
- [41] D. Spillmann, D. Witt, U. Lindahl, *J. Biol. Chem.* **1998**, *273*, 15487–15493.
- [42] G. S. V. Kuschert, A. J. Hoogewerf, A. E. I. Proudfoot, C. Chung, R. M. Cooke, R. E. Hubbard, T. N. C. Wells, P. N. Sanderson, *Biochemistry* **1998**, *37*, 11193–11201.
- [43] B. Goger, Y. Halden, A. Rek, R. Mösl, D. Pye, J. Gallagher, A. J. Kungl, *Biochemistry* **2002**, *41*, 1640–1646.
- [44] A. Pichert, S. A. Samsonov, S. Theisgen, L. Thomas, L. Baumann, J. Schiller, A. G. Beck-Sickinger, D. Huster, M. T. Pisabarro, *Glycobiology* **2012**, *22*, 134–45.
- [45] K. Nordsieck, A. Pichert, S. A. Samsonov, L. Thomas, C. Berger, M. T. Pisabarro, D. Huster, A. G. Beck-Sickinger, *ChemBioChem* **2012**, *13*, 2558–2566.
- [46] K. Möbius, K. Nordsieck, A. Pichert, S. A. Samsonov, L. Thomas, J. Schiller, S. Kalkhof, M. Teresa Pisabarro, A. G. Beck-Sickinger, D. Huster, *Glycobiology* **2013**, *23*, 1260–1269.
- [47] T. Hofmann, S. A. Samsonov, A. Pichert, K. Lemmnitz, J. Schiller, D. Huster, M. T. Pisabarro, M. von Bergen, S. Kalkhof, *Methods* **2015**, *89*, 45–53.
- [48] S. Maignan, J.-P. Guilloteau, N. Fromage, B. Arnoux, J. Becquart, A. Ducruix, *Science* **1995**, *268*, 291–293.
- [49] P. Chardin, D. Cussac, S. Maignan, A. Ducruix, *FEBS Lett.* **1995**, *369*, 47–51.
- [50] M. Rozakis-Adcock, R. Fernley, J. Wade, T. Pawson, D. Bowtell, *Nature* **1993**, *363*, 83–85.
- [51] J. R. Molina, A. A. Adjei, *J. Thorac. Oncol.* **2006**, *1*, 7–9.
- [52] Z. Songyang, S. E. Shoelson, J. McGlade, P. Olivier, T. Pawson, X. R. Bustelo, M. Barbacid, H. Sabe, H. Hanafusa, T. Yi, R. Ren, D. Baltimore, S. Ratnoffsky, R. A. Feldman, L. C. Cantley, *Mol. Cell. Biol.* **1994**, *14*, 2777–2785.
- [53] H. Gram, R. Schmitz, J. F. Zuber, G. Baumann, *Eur. J. Biochem.* **1997**, *246*, 633–637.
- [54] G. M. Clore, E. Appella, M. Yamada, K. Matsushima, A. M. Gronenborn, *Biochemistry* **1990**, *29*, 1689–1696.
- [55] N. A. Baker, D. Sept, S. Joseph, M. J. Holst, J. A. McCammon, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 10037–10041.
- [56] E. Jurrus, D. Engel, K. Star, K. Monson, J. Brandi, L. E. Felberg, D. H. Brookes, L. Wilson, J. Chen, K. Liles, M. Chun, P. Li, D. W. Gohara, T. Dolinsky, R. Konecny, D. R. Koes, J. E. Nielsen, T. Head-Gordon, W. Geng, R. Krasny, G.-W. Wei, M. J. Holst, J. A. McCammon, N. A. Baker, *Protein Sci.* **2018**, *27*, 112–128.
- [57] P. R. B. Joseph, P. D. Mosier, U. R. Desai, K. Rajarathnam, *Biochem. J.* **2015**, *472*, 121–133.
- [58] K. Nordsieck, L. Baumann, V. Hintze, M. T. Pisabarro, M. Schnabelrauch, A. G. Beck-Sickinger, S. A. Samsonov, *Biopolymers* **2018**, *109*, e23103.
- [59] P. Nioche, W. Q. Liu, I. Broutin, F. Charbonnier, M. T. Latreille, M. Vidal, B. Roques, C. Garbay, A. Ducruix, *J. Mol. Biol.* **2002**, *315*, 1167–77.
- [60] S. Yuzawa, K. Ogura, H. Hatanaka, K. Miura, F. Inagaki, *J. Biomol. NMR* **2003**, *27*, 185–186.
- [61] K. Sanches, Í. P. Caruso, F. C. L. Almeida, F. A. Melo, *Biomol. NMR Assign* **2019**, *13*, 295–298.
- [62] K. Ogura, T. Shiga, M. Yokochi, S. Yuzawa, T. R. Burke Jr., F. Inagaki, *J. Biomol. NMR* **2008**, *42*, 197–207.
- [63] Y. Okuno, C. D. Schwieters, Z. Yang, G. M. Clore, *J. Am. Chem. Soc.* **2022**, *144*, 21371–21388.
- [64] S. Mori, C. Abeygunawardana, M. O. Johnson, P. C. M. Vanzijl, *J. Magn. Reson. Ser. B* **1995**, *108*, 94–98.
- [65] J. Iwahara, C. Tang, G. Marius Clore, *J. Magn. Reson.* **2007**, *184*, 185–195.
- [66] W. Lee, M. Tonelli, J. L. Markley, *Bioinformatics* **2015**, *31*, 1325–1327.
- [67] D. A. Case, K. Belfon, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham III, **2020**, *AMBER 2020*, University of California, San Francisco.
- [68] B. Mulloy, M. J. Forster, C. Jones, D. B. Davies, *Biochem. J.* **1993**, *293*(Pt 3), 849–858.
- [69] A. Almond, P. L. DeAngelis, C. D. Blundell, *J. Mol. Biol.* **2006**, *358*, 1256–1269.
- [70] G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, A. J. Olson, *J. Comput. Chem.* **1998**, *19*, 1639–1662.
- [71] M. Ester, H. P. Kriegel, J. Sander, X. Xiaowei, *Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining (KDD) 1996*, KDD-96, 226–231.

- [72] H. Nguyen, J. Maier, H. Huang, V. Perrone, C. Simmerling, *J. Am. Chem. Soc.* **2014**, *136*, 13959–13962.
- [73] K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. González-Outeiriño, C. R. Daniels, B. L. Foley, R. J. Woods, *J. Comput. Chem.* **2008**, *29*, 622–655.
- [74] T. Steinbrecher, J. Latzer, D. A. Case, *J. Chem. Theory Comput.* **2012**, *8*, 4405–4412.
- [75] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graphics* **1996**, *14*, 33–38.
- [76] D. R. Roe, T. E. Cheatham 3rd, *J. Chem. Theory Comput.* **2013**, *9*, 3084–95.
- [77] A. Onufriev, D. Bashford, D. A. Case, *J. Phys. Chem. B* **2000**, *104*, 3712–3720.
- [78] N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, G. R. Hutchison, *J. Cheminform* **2011**, *3*, 33.

Manuscript received: October 27, 2023
Accepted manuscript online: November 29, 2023
Version of record online: February 13, 2024

Publication 2. Ligand binding of
interleukin-8: a comparison of
glycosaminoglycans and acidic peptides.



Cite this: *Phys. Chem. Chem. Phys.*,
2023, 25, 24930

Ligand binding of interleukin-8: a comparison of glycosaminoglycans and acidic peptides†

Christian Schulze,^{‡a} Annemarie Danielsson,^{‡b} Adam Liwo,^b Daniel Huster,^a Sergey A. Samsonov^{‡*b} and Anja Penk^{‡*a}

Recognition and binding of regulatory proteins to glycosaminoglycans (GAGs) from the extracellular matrix is a process of high biological importance. The interaction between negatively charged sulfate or carboxyl groups of the GAGs and clusters of basic amino acids on the protein is crucial in this binding process and it is believed that electrostatics represent the key factor for this interaction. However, given the rather undirected nature of electrostatics, it is important to achieve a clear understanding of its role in protein–GAG interactions and how specificity and selectivity in these systems can be achieved, when the classical key-lock binding motif is not applicable. Here, we compare protein binding of a highly charged heparin (HP) hexasaccharide with four *de novo* designed decapeptides of varying negative net charge. The charge density of these peptides was comparable to typical GAGs of the extracellular matrix. We used the regulatory protein interleukin-8 (IL-8) because its interactions with GAGs are well described. All four peptide ligands bind to the same epitope of IL-8 but show much weaker binding affinity as revealed in ¹H–¹⁵N HSQC NMR titration experiments. Complementary molecular docking and molecular dynamics simulations revealed further atomistic details of the interaction mode of GAG *versus* peptide ligands. Overall, similar contributions to the binding energy and hydrogen bond formation are determined for HP and the highly charged peptides, suggesting that the entropic loss of the peptides upon binding likely account for the remarkably different affinity of GAG *versus* peptide ligands to IL-8.

Received 28th May 2023,
Accepted 3rd September 2023

DOI: 10.1039/d3cp02457a

rsc.li/pccp

Introduction

Glycosaminoglycans (GAGs) represent a particular class of linear anionic periodic polysaccharides made up of disaccharide repetitive units containing a hexosamine and an uronic acid or galactose in the case of keratan sulfate.^{1,2} These molecules present varying net degrees and patterns of sulfation.³ GAGs are major components of the extracellular matrix and play important roles in numerous cellular processes such as signaling,^{4,5} anticoagulation,^{6,7} angiogenesis,^{8,9} and communication.¹⁰ Their biological role is executed through intermolecular interactions with protein partners such as growth factors, chemokines, proteases and collagen.^{11–13} Disruptions of protein–GAG interactions can cause a variety of pathologies including cancer,^{14–16} Alzheimer's¹⁷ and prion diseases,¹⁸ autoimmune¹⁹ and inflammatory disorders.^{14,20} All this renders GAGs to be very promising

molecules for the design of new biomaterials in regeneration therapies.²¹

Binding of protein to GAGs has been studied extensively and different mechanism to regulate this interaction have been shown to be important.^{13,22–24} The first and most important mechanism is *via* specific interactions, where the classical picture of a highly specific binding motif applies, *e.g. via* hydrogen bonds. In particular, for antithrombin III, a very specific short GAG sequence was discovered to exhibit the strongest binding with important pharmaceutical implications.²⁵ This system represents the best example studied so far of the highly specific protein–GAG interactions. Several hydrogen bonds have been identified in the crystal structure.²⁶ Second, low-affinity binding governed entirely by electrostatics has been reported for many protein targets, including BMP-2,²⁷ CXCL-12,²⁸ TIMP-3,²⁹ or TGF- β 1.³⁰ Here, the most important parameter that affects the binding strength is the net charge per GAG disaccharide unit. Although a clear GAG binding epitope on the protein is also defined in this case, the discrimination between different GAG-ligands and thus the binding strength is solely dependent on the charge per disaccharide unit and not the exact position of the charge. Hence, this mechanism results in a rather low affinity and a rather fuzzy structural ensemble of protein-bound GAGs. Nevertheless, such a non-directed interaction may be of high biological relevance due to

^a Institute for Medical Physics and Biophysics, University of Leipzig, Härtelstr. 16/18, 04107 Leipzig, Germany. E-mail: anja.penk@medizin.uni-leipzig.de

^b Faculty of Chemistry, University of Gdańsk, Fahrenheit Union of Universities, ul. Wita Stwosza 63, 80-308 Gdańsk, Poland. E-mail: sergey.samsonov@ug.edu.pl

† Electronic supplementary information (ESI) available. See DOI: <https://doi.org/10.1039/d3cp02457a>

‡ These authors contributed equally.

the highly ubiquitous abundance of GAGs, *e.g.* in the case of heparansulfate–thrombin interaction.^{22,31} Third — and most challenging to study in detail — GAGs bind to proteins *via* contributions of both aforementioned mechanisms. GAG binding by such a mechanism is characterized by a rather low-affinity and is believed to have the largest contribution from electrostatics and solvation energy. However, some distinct differences in the binding strength of ligands of the same net charge but different sulfation patterns are observed. These result in specificity or discrimination between ligands with the same charge density.^{13,32,33} For cathepsin S, chondroitin sulfates with varying sulfation patterns were predicted to bind to different sites, suggesting the underlying molecular mechanism of its inhibition.³⁴ For CXCL-14, binding poses for dermatan sulfate and chondroitin sulfate of the same net charge are clearly distinguishable both in the experiment and in the computational analysis.³⁵ This is also known as the “sulfation code”, where the sulfation pattern is suggested to be crucial for defining the recognition by protein targets and the involvement in particular biochemical processes.³⁶

Unfortunately, protein–GAG interactions are still insufficiently characterized at the molecular level due to the highly repetitive nature and high variability of GAGs, and relatively few GAG structures are available. Consequently, both experimental and computational approaches experience challenges when dealing with these systems because of GAGs' particular properties such as limited availability of experimental structures, high length, high flexibility, conformational variability, periodicity, pseudosymmetry of the charged groups distribution, multipose binding³⁷ and high variation in the sulfation pattern.^{13,32,33} Considering these challenges, there is a central question in protein–GAG research that remains unanswered: *How is specificity in protein–GAG interaction achieved?* This specificity itself could be understood within several contexts: GAG type (glycosidic linkage and monosaccharide components of a periodic unit), GAG net sulfation (the amount of the GAG sulfate groups defining the charge and, therefore, potential strength of electrostatic interactions established by the molecules), GAG sulfation pattern (particular positions of the sulfate groups in a GAG periodic unit). Despite the central role of GAG binding specificity, this issue is far from being completely understood.

In the current work, we aim to get deeper insights into the particular aspect of the specificity of protein–GAG binding by specifically analyzing the role of electrostatics for the binding affinity and structural properties of the protein–polyelectrolyte complexes. In this generalization, we approach the question by asking: would other molecules with a comparable net charge but of very different chemical nature represent similarly binding behavior with the same protein receptor? Answering this question could yield an improved understanding of the extent to which pure electrostatic interactions in protein–GAG binding are decisive, and could, therefore, affect the molecular mechanisms behind their biological functions. To this end, we chose the small chemokine IL-8, a well-known regulatory protein, for which the interactions with GAGs were extensively studied previously and discussed in terms of its potential specificity.^{37–50} We performed

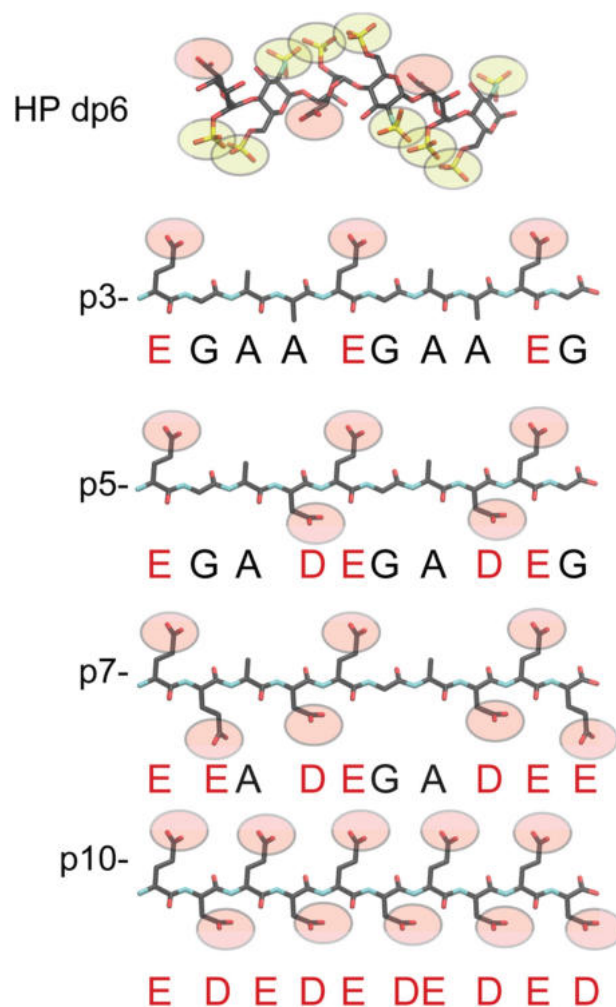


Fig. 1 Chemical structures of HP dp6 and the 4 acidic peptides used in this study. Carboxylate groups are depicted as red and sulfate groups as yellow spheres.

comparative analysis of the interaction of IL-8 with the hexameric heparin (HP) and a series of acidic decapeptides with varying charge density (Fig. 1). We applied a combination of solution NMR and molecular modeling (molecular docking, molecular dynamics, free energy calculations) approaches that were previously demonstrated to be successful in the characterization of the interfaces of ligand binding to small proteins.^{13,51} Although we confirm the strong impact of electrostatics on the binding characteristics of this system, there are also different structural, dynamic and energetic/entropic features that distinguish HP from peptide binding to IL-8 suggesting that electrostatics is not the only driving force responsible for GAG recognition of proteins. Our results contribute to the basic understanding of protein–GAG interactions and represent a fundamental step on the way to deciphering the molecular basis of the molecular specificity in these complex systems.

Materials and methods

Experimental studies

Materials. Heparin hexasaccharide (HP, degree of polymerization of 6; dp6) was purchased from Iduron (Manchester, UK).

The four different peptides, varying in their net charge (p3-: EGAAEGAAEG, p5-: EGADEGADEG, p7-: EEADEGADEE and p10-: EDEDEDEDED, see Fig. 1), were synthesized by the Peptide Synthesis Core Unit (Leipzig University, Germany). The ^{15}N -labeled ammonium salts were purchased from Eurisotop (Saarbrücken, Germany) and all other chemicals from Carl Roth (Karlsruhe, Germany).

Protein production. Human IL-8 (1–77) was expressed, purified, and refolded as previously described,⁴⁵ except the final dialyzing step was against a buffer containing 22.2 mM sodium phosphate and 55.5 mM NaCl at pH 7.0 or only 22.2 mM sodium phosphate at pH 7.0 without any additional NaCl.

Circular dichroism spectroscopy. Peptides were dissolved in the measuring buffer containing 20 mM sodium phosphate and 50 mM NaCl (pH 7.0) to a final concentration of 5 mM. After dissolving the peptides, the pH value of the solution had to be carefully adjusted to pH 7.0 using NaOH. The CD analysis was performed with 70 μM peptide in 20 mM sodium phosphate buffer (pH 7.0) on a Jasco J-1500 spectropolarimeter (Pfungstadt, Germany). All experiments were carried out using a sample size of 350 μl in a 0.1 cm quartz cell at 25 $^{\circ}\text{C}$ and CD spectra were recorded from 190 nm to 260 nm with a scanning rate of 50 nm min^{-1} , 2 nm bandwidth and 0.2 nm data pitch averaged in five scans for each sample. The molar ellipticity (θ) was calculated using the following equation, where m° represents the measured data (mdeg), M the average molecular weight, C the peptide concentration, and L the path length of the cell:

$$[\theta] = m^{\circ} \times M / (10 \times L \times C)$$

NMR spectroscopy. The NMR experiments were performed on Bruker Avance Neo 700 MHz or Avance III 600 MHz spectrometers (Bruker BioSpin GmbH, Rheinstetten, Germany) equipped with 5 mm inverse triple resonance probes with z-gradient. For data acquisition and spectrum processing, the Bruker software TopspinTM and for the analysis of the chemical shift perturbation (CSP) and plotting the software NMRFAM-Sparky was used.⁵²

The peptides were dissolved in 22.2 mM sodium phosphate buffer (0 mM NaCl, pH 7.0) to a concentration between 30 (p3-) and 50 (p10-) mg ml^{-1} and the pH-value was carefully adjusted to 7. Afterwards, an extensive dialysis of roughly 100 μl of the peptide solution against two times 1 l of the buffer without NaCl using SpectrumTM Micro Float-A-LyzerTM with a MWCO of 100–500 Da was performed. The concentration of the resulting peptide solutions ($\sim 300 \mu\text{l}$) was determined using ^1H NMR. For each peptide 10 μl of the solution was mixed with 100 μl valine ($c = 1.7 \text{ mM}$) and filled up to 500 μl including 10% D_2O . A 1D ^1H NOESY with low power water presaturation during the mixing time (10 ms) and a 5 s period during the recycle delay of 40 s were collected, with a 64 scans each. Concentrations of the peptide solutions were determined using the internal valine standard (H_{γ} signal) and the alanine H_{β} signals. For p10- the H_{β} and H_{γ} of glutamic acid and H_{β} of aspartate were integrated and the concentration was determined. Finally, these solutions

were mixed with phosphate buffer (either 0 mM NaCl or 1000 mM NaCl) and D_2O to obtain stem solutions of the peptides with 10 mM concentration (p3-, p5-), 9 mM (p7-) and 7 mM (p10-) in two different buffers. Both buffer systems contained 20 mM sodium phosphate at a pH of 7 including 10% D_2O . One contained no additional NaCl, while the other contained 50 mM NaCl.

The fully ^{15}N -labeled IL-8 (100 μM) sample was measured at a temperature of 30 $^{\circ}\text{C}$ in 20 mM sodium phosphate buffer (pH 7.0) containing 10% D_2O and 4 μM TSP- d_4 with or without additional 50 mM NaCl. The acquired ^1H - ^{15}N HSQC spectra were comparable with published data of the homodimeric form of IL-8,⁵³ and the previously reported assignment was used.⁴⁵ For CSP experiments, increasing amounts of HP hexasaccharide (only in buffer with 50 mM NaCl) or the respective peptides were titrated. Furthermore, pure buffer was titrated in to exclude chemical shift perturbations due to protein dilution. For each titration step, a ^1H - ^{15}N Fast HSQC spectrum,⁵⁴ using a watergate 3919 water suppression⁵⁵ and globally optimized alternating phase rectangular pulse (GARP) for heteronuclear decoupling,⁵⁶ was acquired. For each NMR spectrum, typical pulse lengths of 8.9–10.2 μs for ^1H (depending on the salt concentration of the buffer), 35 μs for ^{15}N 90° pulses and 240 μs for the ^{15}N decoupling were used. In total, 32 scans per increment were acquired with a spectral width of 16 ppm as well as complex 3072 data points in the direct, and 25 ppm as well as real 128 data points in the indirect dimension. To reduce the sample size to 350 μl for p10-, a Shigemi[®] tube matched to D_2O (Shigemi Co., Tokyo, Japan), was used. The resulting weighted chemical shift change for each NH signal of the IL-8 backbone was calculated using the following equation, where $\Delta\delta$ represent the chemical shift perturbation (CSP):

$$\Delta\delta(^1\text{H}, ^{15}\text{N}) = \sqrt{(\delta_{\text{H}})^2 + (\delta_{\text{N}}/5)^2}$$

To determine the apparent k_{D} value for the binding peptide, the calculated CSP was plotted against the ligand concentration and fitted using the following equation,⁵⁷ where $\Delta\delta_{\text{obs}}$ is the observed chemical shift change from the free state, $\Delta\delta_{\text{max}}$ is the maximum chemical shift change on saturation, $[\text{P}]_{\text{t}}$ and $[\text{L}]_{\text{t}}$ are the respective protein/ligand concentrations:

$$\Delta\delta_{\text{obs}} = \Delta\delta_{\text{max}} \frac{([\text{P}]_{\text{t}} + [\text{L}]_{\text{t}} + k_{\text{D}}) - ([\text{P}]_{\text{t}} + [\text{L}]_{\text{t}} + k_{\text{D}})^2 - 4[\text{P}]_{\text{t}}[\text{L}]_{\text{t}}^{1/2}}{2[\text{P}]_{\text{t}}}$$

For the mean k_{D} only the five most affected amino acids (by CSP) were taken into account, as they most likely represent the binding to the ligand instead of effects due to tertiary structure changes. From this the mean k_{D} and its standard deviation was calculated.

For visualization of the CSP on the tertiary protein structure USCF Chimera was used.⁵⁸

Computational studies

Structures. The structure of the dimeric IL-8 protein used for all simulations was obtained from the PDB (PDB ID: 1IL8).⁵⁹

The N-terminal missing residues were built in xLeap module of AMBER16.⁶⁰ The structure of hexameric HP dp6 used for molecular docking simulations was taken from the PDB (PDB: 1HPN).⁶¹ The four anionic peptides, p3-, p5-, p7-, and p10- were built using the xLeap module of AMBER16⁶⁰ based on their amino-acid sequence.

Molecular docking. Docking simulations were performed with Autodock 3 (AD3),⁶² using a grid box with the grid step of 0.375 Å centered on the two α -helices of the IL-8 dimer (the heparin-binding site as described by ref. 39), of the following sizes: 60 × 60 × 60 and 80 × 80 × 60 grid points for the IL-8/HP complex and the IL-8/anionic peptide complexes, respectively. The number of rotatable bonds of the ligands was set to 29 for p3-, 32 for p5-, p7-, p10-, and HP dp6, respectively. IL-8 was kept rigid throughout the docking simulation. Each docking experiment was carried out in 100 independent runs, employing the Lamarckian Genetic Algorithm with an initial population size of 300, set to terminate after 10⁵ generations. The 50 docking solutions with the best AD3-score were clustered using the DBSCAN algorithm.⁶³ The metric used was root mean square atom type distance (RMSatd metric), which is used to calculate distances between atoms in the same manner as RMSD, but uses pairs of atoms of the same type which are spatially close. The neighborhood search radius was determined manually for each complex in order to maximize the amount of identified clusters. For subsequent molecular dynamics (MD) simulations, the cluster representative(s), cluster members with the best AD3-score within the cluster, and docking solutions with the best AD3-score that did not belong to a cluster were used, amounting to a total of 10 solutions per complex. No experimental data was used as a restraint for either docking or MD simulations.

Molecular dynamics

Molecular dynamics of protein-ligand complexes. MD simulations of the complexes obtained from docking experiments were carried out in AMBER16.⁶⁰ The ff14SB force field⁶⁴ parameters were used for the protein and peptide molecules, while GLYCAM06 parameters were used for HP. The simulated IL-8/peptide and IL-8/HP complexes were solvated in a periodic TIP3P octahedron water box with at least 15 Å distance between atoms of the analyzed complex and box boundaries. Charges of the complexes were neutralized by adding Na⁺ and Cl[−] counterions. Prior to the MD analysis, two energy minimization steps, with and without harmonic restraints on the solute atoms, were performed, followed by heating the system up to 300 K for 10 ps, and an equilibration step at 300 K in an isothermal isobaric ensemble (*NTP*) for 100 ps. Subsequently, the 20 ns-long productive MD run was carried out with snapshots of the trajectories being written every 10 ps. The entire protocol used for the MD simulation can be found in detail in ref. 28.

Long MD simulations of unbound ligands. The unbound ligand structures were analyzed in a separate 1 μ s-long MD simulation. The anionic peptides were solvated in a TIP3PBOX with minimum distance of 7 Å between box boundaries and the

solute, and Na⁺ counterions were added to the system. The same protocol for the minimization, heating, and equilibration steps was as described above, followed by the productive MD run. The corresponding simulation of HP dp6 was taken from our previous work.⁶⁵

Free energy calculations. Calculations of the free energy of the IL-8/peptide and IL-8/HP complex trajectories as well as per-residue energy decomposition were performed using the Molecular Mechanics-Generalized Born Surface Area (MM-GBSA) approach in AMBER16 with igb = 2. The analysis was performed for all frames as well as for the last 5 ns of the productive MD run, however no significant difference was observed between the results of those subsets; hence, the energy calculation results for all frames are presented and analyzed. The mean values for total, electrostatic, and van der Waals energy components were compared between the ligands and the statistical significance of the differences was determined using Welch's *t*-test for unpaired samples.⁶⁶ Bonferroni correction for multiple comparisons was applied. The MD simulation length of 20 ns was shown to be sufficient for statistically significant MD-based free energy calculations for another protein-GAG system of a similar size when an ensemble of trajectories starting from multiple docked poses are analyzed.⁶⁷

Comparison of bound and unbound ligand structures

The RMSD (root mean square deviation) of heavy atoms of the same ligand between bound (docked) and unbound states was calculated to make a qualitative estimate of how much the conformation of the ligand changes upon binding to IL-8. Per ligand, the bound structures were taken from all of the 1000 solutions obtained from the docking of the ligand to IL-8 from all AD3 runs. The unbound structures consisted of the 10⁸ conformations obtained from the 1 μ s-long MD simulation.

The number of structures considered similar was calculated for each ligand by counting the amount of bound and unbound structures that were within a certain RMSD cutoff after structural superposition of the structures. The RMSD cutoff value was incremented by a step of 1 Å from 0 Å to 12 Å. The results of this comparison were used to estimate the total energy of the system as a function of the reaction coordinate, which was expressed as the number of structures below the RMSD cutoffs (RMSDi). The resulting Potential of Mean Force (PMF) curves were plotted, and the free energy values were used to obtain an estimate of the dissociation constant k_D using the relationship:

$$\Delta G = -RT \times \ln k_D + C$$

$$= -RT \times \sum_{i=1}^N \left(\ln \left(\frac{\text{Nunder RMSD}_i}{\text{total } N} \right) \times \frac{\text{Nunder RMSD}_i}{\text{total } N} \right) + C,$$

where R is the universal gas constant and T is the temperature of the system, k_D is the binding constant and N is the number of corresponding structures.

The radius of gyration (R_{gyr}) of each ligand was calculated in cpptraj⁶⁸ using default parameters for all atoms and compared between the bound (docked) and unbound states in order to assess the compactness of the structures. Per ligand, the bound

structures were taken from all 1000 AD3 solutions obtained from the docking of the ligand to IL-8. The unbound structures consisted of the 10^8 conformations obtained from the 1 μ s-long MD simulation. The kernel density estimates of R_{gvr} values were calculated and plotted for all ligands in R.⁶⁹

H-Bonds. Hydrogen bonds (H-bonds) established between IL-8 and the ligand molecules were identified from all frames of the 20 ns MD trajectories using cpptraj.⁶⁸ The default parameters were used to determine the presence of a H-bond (bond distance cutoff of 3.0 Å and hydrogen bond angle cutoff of 135°). For each donor-acceptor pair, the fraction of frames in which they formed a H-bond was calculated; the obtained fractions were averaged across the 10 replicates of MD runs per IL-8/ligand complex and subsequently summarized and visualized in Python 3.8.5 using the numpy 1.19.2,⁷⁰ pandas 1.1.3,⁷¹ and matplotlib 3.3.2⁷² libraries as well as in the R package.⁶⁹

Ligand flexibility and secondary structure analysis. The 1 μ s-long MD simulation of the unbound ligands was used to analyze the ligands' flexibility and secondary structure composition using cpptraj.⁶⁸ The secondary structure content of the peptide ligands was determined using the *secstruct* command and subsequently visualized using gnuplot.⁷³ The atomic fluctuations of HP dp6 and the four peptides were obtained using the *rmsd* command, applied to all atoms of the superimposed trajectories. The variation of the rmsd value for the whole trajectory was used as a measure of the ligand flexibility.

Data analysis and visualization. Postprocessing analysis of the trajectories was performed using cpptraj module of AMBER16.⁶⁸ Visualization of the analyzed structures was carried

out in VMD.⁷⁴ The R package⁶⁹ was used for data analysis and statistical analysis (Welch *t*-test for unpaired samples and Bonferroni corrections for multiple testing).

Results

Interaction of IL-8 with charged ligands

The ^1H - ^{15}N HSQC NMR spectrum of IL-8 shows well dispersed, high-resolution signals (red contours in Fig. 2). Titration of the five ligands (HP dp6 and the four negatively charged peptides, see Fig. 1) to the protein caused specific peak shifts in the ^1H - ^{15}N HSQC NMR spectra indicating changes in the chemical environment of the respective amino acid. Fig. 2 shows as example the ^1H - ^{15}N HSQC NMR spectra for the titration of the highest charged peptide ligand, featuring ten negative net charges (p10-). The final titration step used a 20-fold excess of the titrated peptides over IL-8, while for HP only an equimolar GAG/protein ratio was used. Higher HP dp6 concentration leads to a loss of signal intensity in the ^1H - ^{15}N HSQC spectra combined with an increased turbidity of the sample. However, after some incubation time, the solution became clear again and no aggregated protein was observable, while the NMR signal intensity remained the same. All ligands changed the chemical shifts of several residues of IL-8; the affected amino acids were similar, but the magnitude of the chemical shift perturbation (CSP) varied. The CSP increased with increasing net charge of the ligands (Fig. S1, ESI†).

To map out these chemical shift changes for each residue of IL-8, the weighted chemical shift perturbations (CSP)

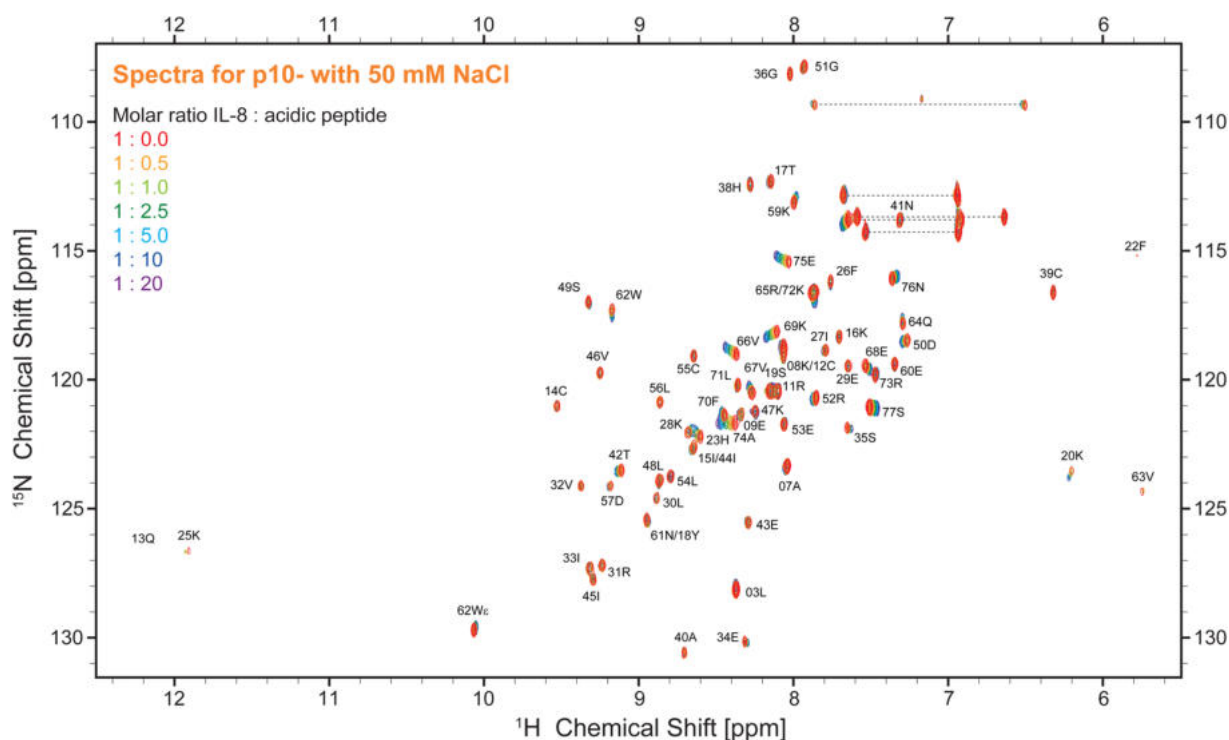


Fig. 2 ^1H - ^{15}N HSQC NMR spectra of ^{15}N -labeled IL-8 (100 μM , 50 mM NaCl, 20 mM NaP, pH 7.0; red) in the presence of the p10- peptide at varying molar ratios of IL-8 to p10- (orange 0.5, light green 1, dark green 2.5 light blue 5, dark blue 10 and purple 20 fold molar excess of p10-).

($\Delta\delta(^1\text{H},^{15}\text{N})$) were calculated as described in the methods section. The threshold for the significance level was set to 0.02 ppm, which was well above 2σ of the CSP for all control spectra. Fig. 3 shows these CSPs of the IL-8 for each peptide as well as for HP dp6 as a function of sequence. By comparing the overall pattern of peak shifts for each ligand, some similarities but also specific differences are observable. The residues with the highest CSP magnitude are concentrated in two regions of IL-8, including (i) residues K20–K28 and (ii) K59–S77. Especially residues V66 and A74 of the C-terminal α -helix show a strong

response in the titration experiments for all ligands. Close to the shorter α -helix of IL-8, the titration with the peptides leads to a similarly pronounced shift of residue K25, H23 and K20, whereby HP dp6 influences predominantly residue H23 and K20. Fig. 3F shows the CSP plotted on the structure of the IL-8 dimer, indicating an overall similar region of IL-8 affected by the binding of p7-, p10- and HP dp6, where the degree of perturbation due to the ligands on IL-8 is ordered as p3- < p5- < p7- < p10- < HP dp6. Please note, that at 20-fold ligand excess, p10-shows similar effects like HP dp6 at equimolar ratio

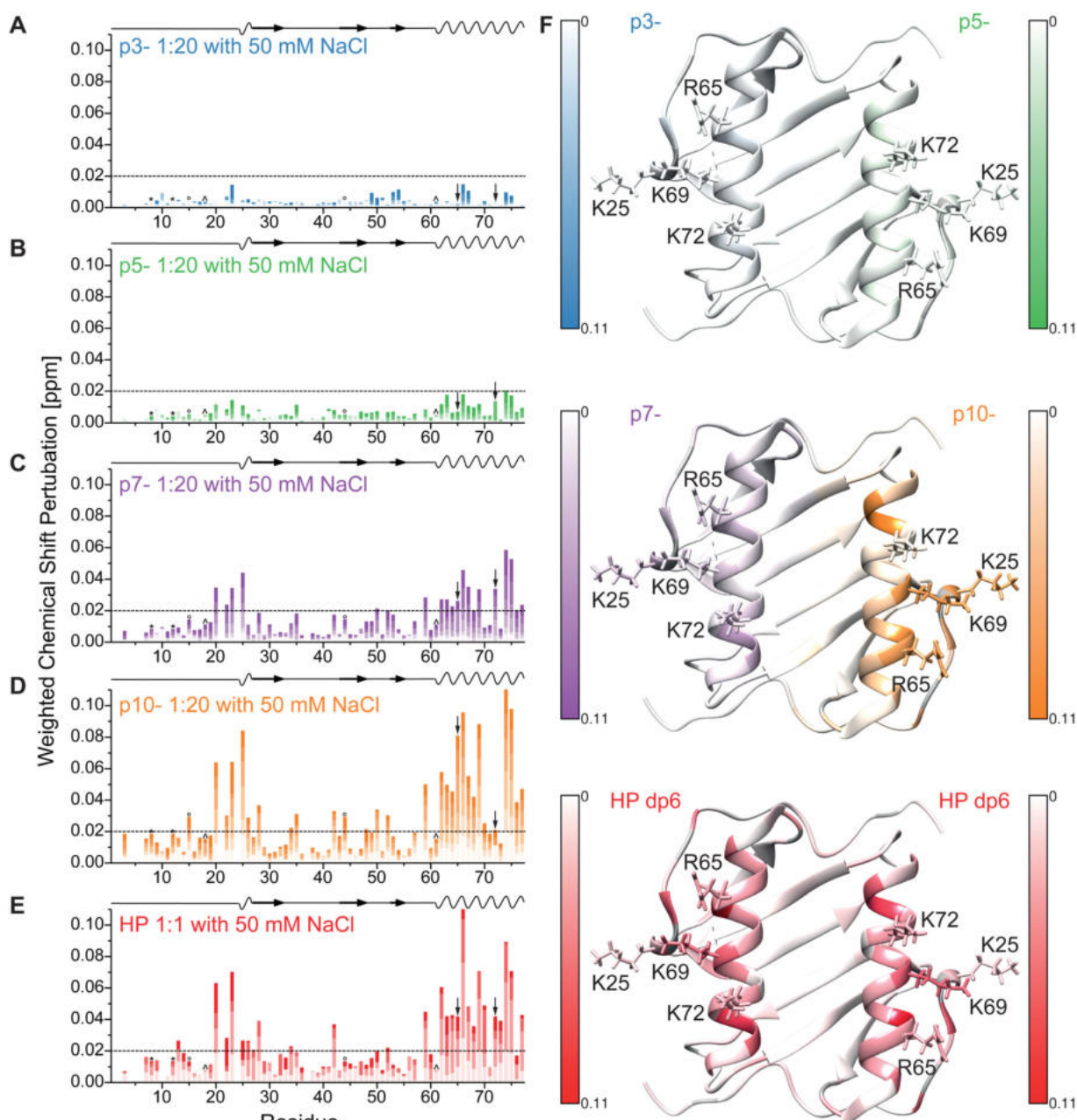


Fig. 3 Weighted chemical shift perturbation of IL-8 in buffer containing 50 mM NaCl upon addition of (A–D) the four peptides (0.5, 1, 2.5, 5, 10 and 20-fold molar excess relative to IL-8 with increasing color) and of (E) HP dp6 (0.2, 0.5, 0.75 and equimolar ratio with IL-8) plotted for each residue of IL-8. The black dashed line denotes the significance threshold of 0.02 ppm. On the top of each graph, the secondary structure of IL-8 is shown, where waves represent α -helical and arrows β -sheet regions. *, ° and ^ denote overlapping peak pairs – K8/C12, 15I/44I and 18Y/61N respectively – which have therefore the same CSP, as they did not separate during the course of titration. Please note, that the arrows indicate the overlapping peak pair R65/72K which separated during the course of titration but could therefore not unambiguously assigned. (F) Shows the highest observed CSP of all ligands (uniform scale from 0 to 0.11 ppm) plotted on the tertiary structure of dimeric IL-8 (pdb: 1IL8). Residues without information about the CSP are colored grey, sidechains of the important residues K25, R65, K69 and K72 are shown and labelled.

between protein and GAG. Nevertheless, for p5-, only A74 shows a CSP above the threshold and for p3- no significant CSP was reached.

Next, we performed the same titration experiments in a NaCl-free buffer because this reduced salt concentration strengthens electrostatic interactions due to reduced screening of the charges. The CSPs for the same titration steps as in Fig. 3 are shown in Fig. 4 under NaCl-free buffer conditions. As expected, the overall magnitude of the CSPs was increased for each peptide ligand in the absence of NaCl (note the different scaling of the y-axis in Fig. 3 and 4). The regions most influenced by ligand binding remain the same (residues K20–K28 and K59–S77) and the overall pattern in the helical region is comparable to the buffer conditions with higher ionic strength. However, especially the CSP pattern for p10- shows some distinct changes for residues K20–K28. Fig. 4E shows the

maximal CSP during titration enlarging this region. This time, the full color represents the maximum CSP of the respective titration to allow for a better distinction of the effects on the influenced amino acid pattern. While for p7- (left) nearly no difference without (top panel) and with 50 mM NaCl (middle panel) is observed, for p10- (right side) no influence on the helical pattern is detected, but in the absence of NaCl a reduced effect on K25 (side chain shown and labeled) is observed compared to 50 mM NaCl. Interestingly, the sequence of residues with maximum CSP in this region changed from K25 > H23 ~ K20 in the presence to K20 > H23 ~ K25, S19 in the absence of NaCl. However, using HP as ligand, an even more reduced influence on K25 and no influence on S19 is observed. Taken together, this shows that the only three charged amino acids in this region (K20, H23, K25) are influenced by the peptide binding, while HP binding clearly perturbs the chemical shift of

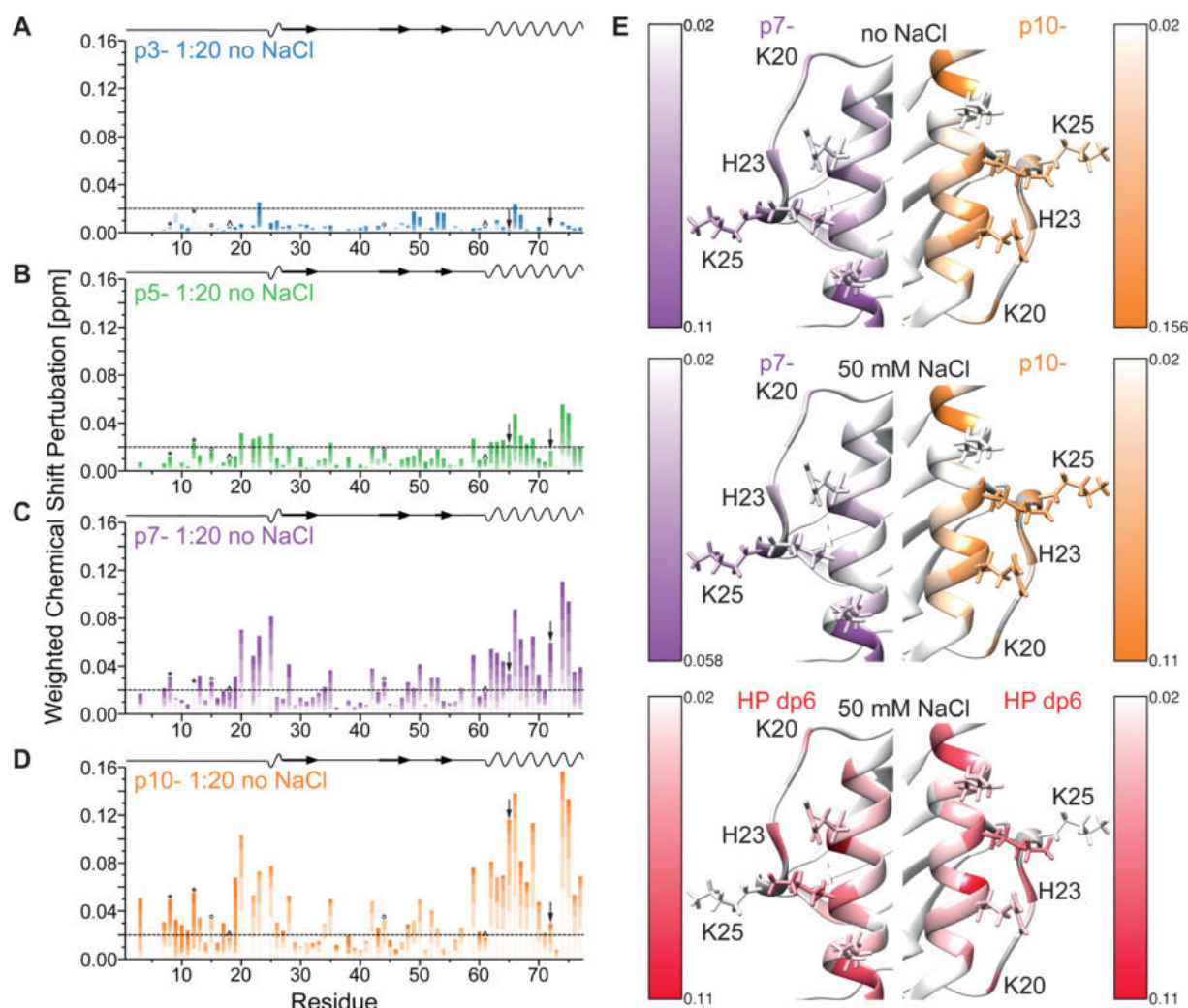


Fig. 4 Weighted chemical shift perturbation of IL-8 in NaCl-free buffer upon addition of (A–D) the four peptides (0.5, 1, 2.5, 5, 10 and 20-fold molar excess relative to IL-8 with increasing color) plotted for each residue of IL-8. The black dashed line denotes the significance threshold of 0.02 ppm. On the top of each graph, the secondary structure of IL-8 is shown, where waves represent α -helical and arrows β -sheet regions. ° and ^ denote overlapping peak pairs 15I/44I and 18Y/61N – which have therefore the same CSP, as they did not separate during the course of titration. * represents the initially slightly close peak pair K8/C12, whose CSPs were evaluated using integration. Please note, that the arrows indicate the overlapping peak pair R65/72K which separated during the course of titration but could therefore not unambiguously assigned. (E) Shows the highest observed CSP in the helical region and from K20–K25 for p7- and p10- in buffer without additional NaCl (top), in buffer with 50 mM additional NaCl (middle) and for HP dp6 in buffer with 50 mM NaCl (scale from 0.02 ppm to the respective maximum CSP) plotted on the tertiary structure of dimeric IL-8 (pdb: 1IL8). Residues without information about the CSP are colored grey, sidechains of K25, R65, K69 and K72 are shown.

K20 and H23 more than K25 (see also Fig SI1 and SI2 lowest panel for enlarged spectra of this region, ESI†). Supporting this binding of the peptides governed mostly by the charged side chains of IL-8, the most highly charged peptide also shows a strong influence on either R65 or K72 (labelled by arrows in Fig. 3 and 4 and side chains shown, see also Fig. SI1 and SI2 upper panel, ESI†), which are adjacent to the K20–K25 region.

Determination of apparent k_D values

We analyzed the binding strength of the negatively charged ligands by plotting the CSP of each titration step of the five most perturbed amino acids against the ligand concentration and calculated the apparent dissociation constant (k_D) for the interaction of IL-8 with each ligand (Fig. 5). Normally, the k_D values determined from the NMR experiments do not represent the intrinsic k_D because the protein concentration necessary for solution NMR measurements is much higher than the intrinsic k_D . However, taking the starting protein concentration of 100 μM and the end titration step with 20 fold excess for the

peptides into account, a determination of the k_D in the range from 20 μM to 2000 μM is reasonable.⁵⁷ Fig. 5A shows exemplarily the plots for residues V66 and A74, which were highly influenced by all ligands, when the CSP threshold was reached (see Fig. 5B). The apparent k_D values ($\pm$ standard deviation) obtained from the fits are (1774 $\pm$ 570) μM for p5-, (1024 $\pm$ 400) μM for p7- and (194 $\pm$ 21) μM for p10- and thus show the expected effect of the net charge in buffer containing 50 mM NaCl. Because the titration experiment with HP dp6 in the same buffer was only possible up to an equimolar ratio, saturation of the CSP was not reached and a k_D could not be determined. However, given the fast increase of the CSP at low ratios, an apparent k_D much smaller than 25 μM is expected. Hence, even the highest charged peptide p10- shows at least an order of magnitude higher k_D at higher salt concentration (50 mM).

The apparent k_D values for ligand binding in the absence of NaCl are larger than 2000 μM for p3- as no saturation was reached, (506 $\pm$ 193) μM for p5-, (142 $\pm$ 22) μM for p7- and (26 $\pm$ 8) μM for p10-. They show again the clear influence on the

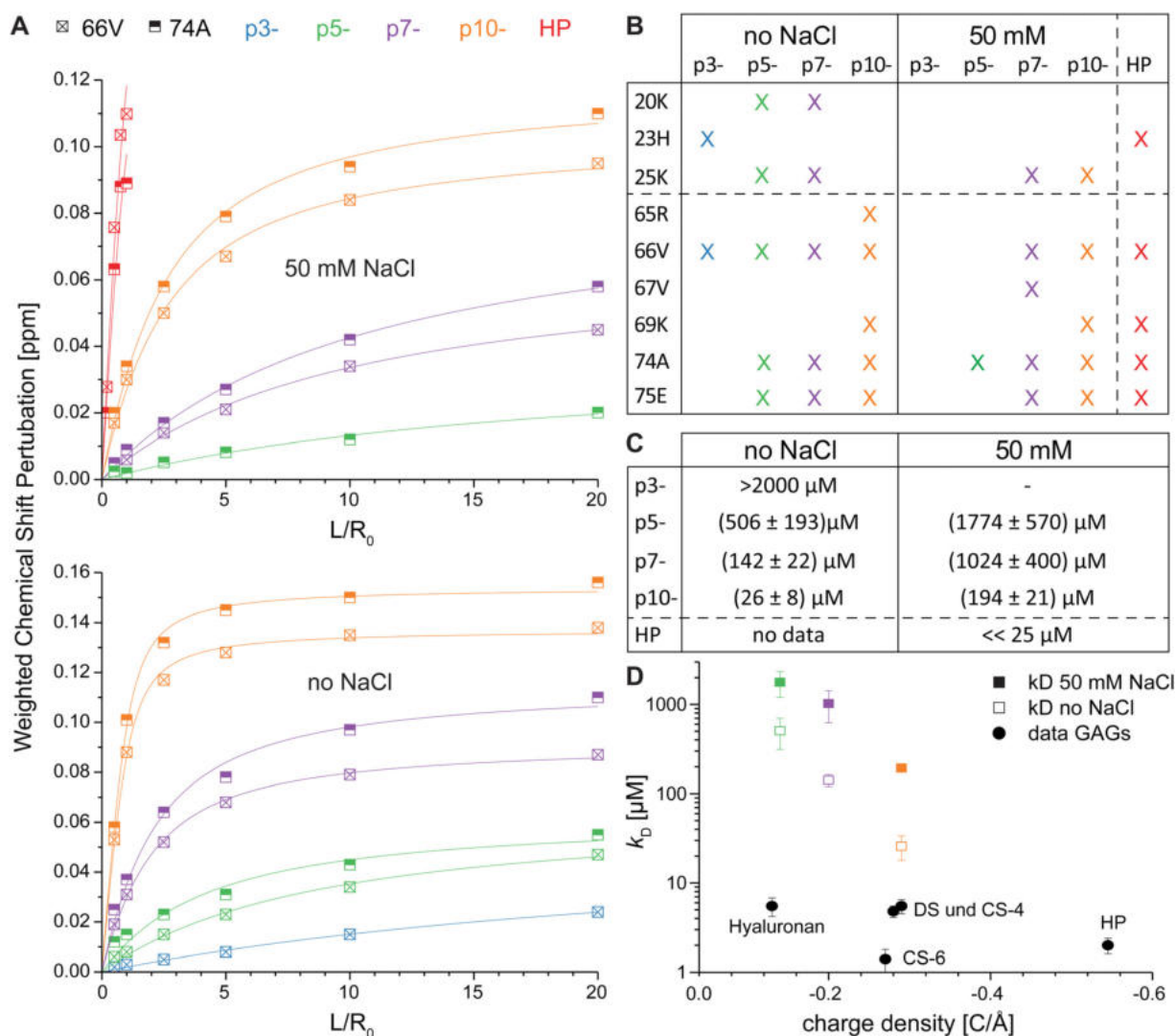


Fig. 5 Plot of the weighted CSP of residue V66 and A74 for the titration experiment with HP dp6 and the four peptide ligands in buffers with different ionic strength (A). Solid lines represent best fits to a 1 : 1 binding model. In (B) the up-to-five amino acids with largest CSP by the binding, from which the k_D were calculated (C). (D) Shows the k_D from this work and data adapted from literature for different hexasaccharides,^{45,46} where the same high ionic strength buffer (filled symbols), but a concentration of 1 μM IL-8 was used.

net charge of the peptide ligand confirming the stronger binding at reduced ionic strength of the buffer.

Ligand docking to dimeric IL-8

In order to predict the structures of the complexes of IL-8 with HP dp6 and the anionic peptides, the molecular docking approach was applied. Docking simulations were performed with the grid box centered on and encompassing only the C-terminal helical regions of IL-8 dimer. Fig. S3 (ESI[†]) shows the results of the docking; while most ligand molecules were placed parallel to the α -helices of IL-8, the solutions appear to be more narrow for HP and p3-, while for p5-, p7- and p10- the docked structures are more spread out along the helices. The results obtained for HP dp6 are in agreement with the previous data we obtained for this system as an alternative binding pose⁴⁵ and by Joseph *et al.*,³⁷ where for longer HP poses, in which HP bound to both C_R-terminal helices, were proposed.

Clustering the 50 top-scoring solutions for each ligand identified two clusters in the case of peptides and only one cluster for HP. For subsequent MD simulations, the cluster representative(s) (the structure(s) closest to the cluster center), high-scoring cluster members, and high-scoring structures not belonging to any cluster were taken, resulting in a total of 10 docking solutions per docking experiment chosen (Fig. 6). Apart from p7-, no clear visual difference in the representative binding poses within the clusters could be determined. For p7- several binding poses stand out from the rest of the structures.

Molecular dynamics

MD simulations of the IL-8/ligand complexes obtained from AD3 docking experiments were performed. To consider the conformational variety of the structural ensemble obtained by molecular docking, 10 different representative initial structures for each ligand were used in the MD simulation. Fig. 7 shows the results of one of the MD simulations for each ligand, while

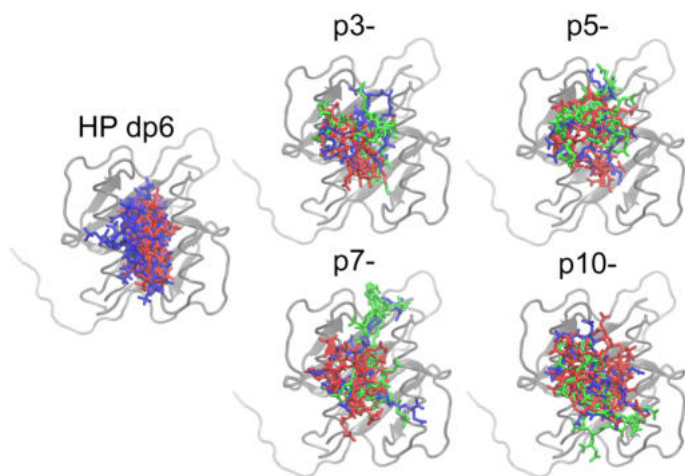


Fig. 6 The docking poses used as input for the MD for IL-8 complexed with HP dp6 and p3-, p5-, p7-, and p10-. The protein is shown in cartoon and the ligands in stick representation. Ligand structures are colored according to cluster membership as identified using DBSCAN clustering, with cluster 1 in red, cluster 2 in light green, and structures not belonging to any cluster in dark blue.

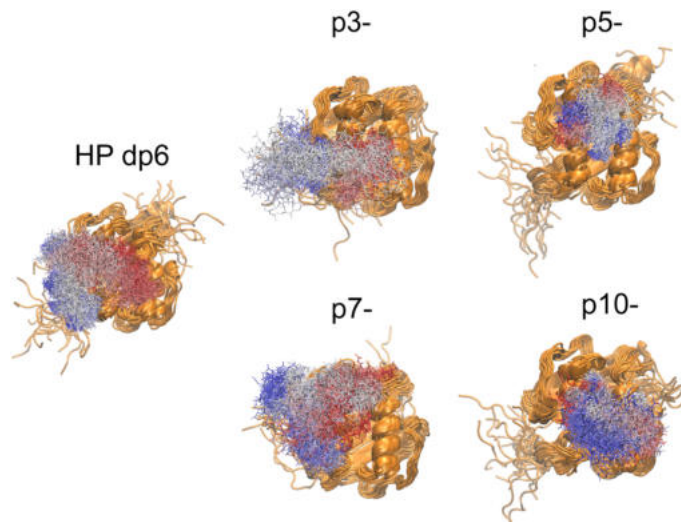


Fig. 7 Representative MD trajectory frames (every 100 ps for ligand, and every 1 ns for the protein) of IL-8 complex with HP dp6 and p3-, p5-, p7-, and p10-. Protein is shown in cartoon and ligand in stick representation (red corresponds to the beginning and blue to the end of the simulation).

Fig. S4 (ESI[†]) shows the poses of the ligand at the start of the MD simulations (red) and at the end of the 20 ns MD simulation (blue). In many cases, the ligand molecules dissociated from the initial binding pose either to move along the α -helices of IL-8 or to bind to the side of IL-8. However, no clear trend in neither binding site nor binding pose preference could be identified for the ligands when compared to each other. Nevertheless, peptide p5- moved the least in all of the 10 simulations performed for the given ligand, suggesting that the initial pose was energetically favorable.

Energetic analysis of the complexes revealed differences in binding strength between the ligands as shown in Fig. 8. A clear difference between HP and the peptides was seen in terms of the total free energy, with IL-8/HP complexes being on average more favorable. The differences were statistically significant between HP and p3- as well as HP and p7-, while between the peptides the only statistically significant difference was between p10-, which had the lowest average total free energy of binding, and p7-. The *in vacuo* electrostatic component of the total energy became more favorable with increasing charge of the ligand molecule, with HP and p10- having similar average electrostatic energies, while IL-8/p3-complexes were the least favorable in terms of electrostatic energy. The differences were statistically significant between HP and p3-, p5-, and p7-, between p10- and p3-, p5-, and p7-, as well as between p3- and p5-, p3- and p7-, and p5- and p7-. The van der Waals interaction energy was the lowest for p5-, and the least favorable for p10-, while HP, p3-, and p7-average energies were comparable. The only differences that remained statistically significant after multiple-testing correction were between p10- and HP dp6, p3-, and p5-. The total electrostatic energy made up of the *in vacuo* and implicit solvation generalized Born component follows the same trend as the electrostatics *in vacuo* for the peptide series. Only for p10- is the total electrostatic impact favorable. Interestingly, it is significantly more favorable

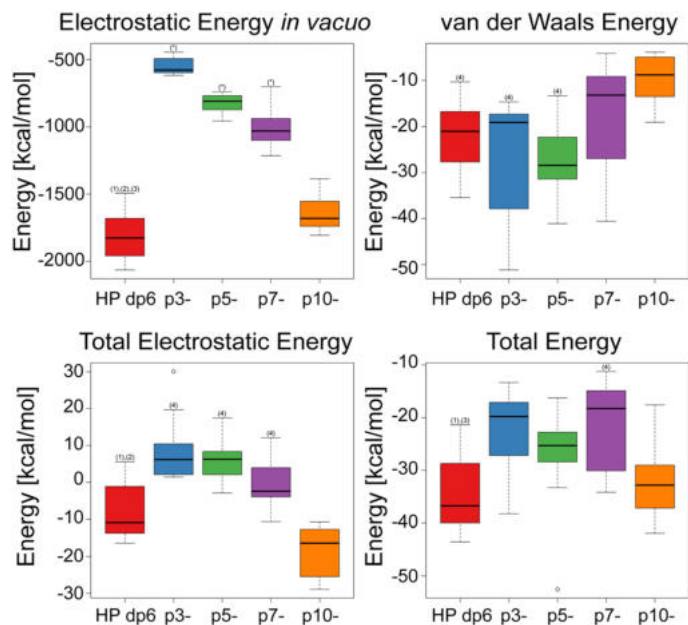


Fig. 8 Boxplots representation of the electrostatic energy *in vacuo* (top left), total electrostatic energy (bottom left), van der Waals energy (top right) and total free energy (bottom right) for the IL-8/ligand complexes determined using MM/GBSA. Statistical significance of differences is shown as (*) for statistical significance when compared to all other groups, (1) statistical significance when compared to p3-, (2) statistical significance when compared to p5-, (3) statistical significance when compared to p7-, (4) statistical significance when compared to p10-.

than for the HP dp6, which has a more negative charge than p10-. This analysis allowed to conclude that HP dp6 and p10- are comparable in terms of electrostatics which justifies the design of the used peptides to meet our original aim which was to analyze if there are substantial differences in binding for representatives of essentially different classes of molecules that are similar in terms of their electrostatic properties.

In order to identify key residues of IL-8 involved in the binding of HP dp6 and the anionic peptides, a per-residue decomposition of the total free binding energy was performed. The residues that contribute most favorably to the binding energy, R65, K69 and K72, are common to all of the studied systems (Table 1). Residue K25 was important for IL-8/HP dp6 and IL-8/p10- complexes, however the magnitude of its influence was greater for the HP-containing complex.

The acidic residues E68 and E75 of IL-8, on average, had a strongly unfavorable impact on binding HP dp6, contributing

on average more than $+1.50 \text{ kcal mol}^{-1}$ to the binding free energy. Only residue E68 had a similar contribution in the IL-8/p10- complex (average energy contribution equal to $+1.55 \text{ kcal mol}^{-1}$), while in case of IL-8 complexed with the other peptides, this influence of E68 was not greater than $+1.50 \text{ kcal mol}^{-1}$. This could be attributed to the higher flexibility of the peptides in comparison to HP dp6, which allowed them to reduce the unfavorable interactions with the negatively charged residues. While all of the HP residues had negative average binding free energy contributions, peptide residue E1 had an unfavorable (positive) average binding free energy contribution above $+1.50 \text{ kcal mol}^{-1}$ for all of the IL-8/peptide complexes.

Comparison of bound and unbound ligand conformations

To explore the changes in conformation and compactness of the ligand structures upon binding to IL-8, the distributions of radii of gyration (R_{gyr}) were compared for each ligand between its bound and unbound state and those differences were compared between ligands. R_{gyr} has been used as a measure for the description of structural specificity of both binding site and the ligand.^{75–77} Fig. 9 shows the differences between the bound and unbound state for each ligand. The distribution of the HP R_{gyr} does not essentially change upon binding, while there is a clear increase of R_{gyr} values for all the peptides. With the increase of the peptide charge, this difference in the R_{gyr} distributions between the bound and unbound states are smaller. This can be explained by the fact that more charged peptides are less likely to be folded into particularly compact structures when unbound due to the intramolecular electrostatic repulsion. This analysis underlines the principal structural differences between the peptides and the HP molecule in IL-8 binding and could potentially be a factor explaining the specificity of HP binding.

The secondary structures of the unbound acidic peptides were analyzed with the DSSP (Dictionary of Secondary Structure of Proteins) approach applied to the corresponding MD trajectories (Fig. S5, ESI†). Neither peptide revealed significant secondary structure elements in the course of the simulation. This was also in agreement with the CD experiments (Fig. S6, ESI†).

The atomic fluctuation analysis of the unbound ligands revealed substantial differences in flexibility between HP dp6 and the peptide ligands. The flexibility of the HP dp6 in terms of the RMSD variance was equal to $0.7 \text{ \AA}$ and lower in comparison to the flexibility values of p3- ($4.1 \text{ \AA}$), p5- ($2.3 \text{ \AA}$), p7- ($1.9 \text{ \AA}$) and p10- ($1.1 \text{ \AA}$). Among the peptides, higher rigidity is observed for more charged peptides as it was also suggested by the analysis of R_{gyr} distributions.

Analysis of hydrogen bonds

Fig. 10 shows the patterns of H-bond formation during MD simulations between IL-8, acting as H-bond donor, and the ligands. The C-terminal helix region (residues 64–73) was an important site of H-bond formation for all of the complexes, especially for residues R65, K69, and K72. A second site of

Table 1 Residues of IL-8 with mean contributions to ΔG below $-1.5 \text{ kcal mol}^{-1}$ for the complexes IL-8/HP and IL-8/peptides, averaged over 10 replicates. Residues with more favorable free energy contributions than $-2.0 \text{ kcal mol}^{-1}$ are shown in bold

	IL-8 residues	
	IL-8 monomeric unit 1	IL-8 monomeric unit 2
HP dp6	R65, K69, K72	K25, R65, K69, K72
p3-	R65, K72	R65
p5-	R65, K69, K72	K72
p7-	R65, K69, K72	R65, K69, K72
p10-	K25, R65, K69, K72, R73	R65, K69, K72

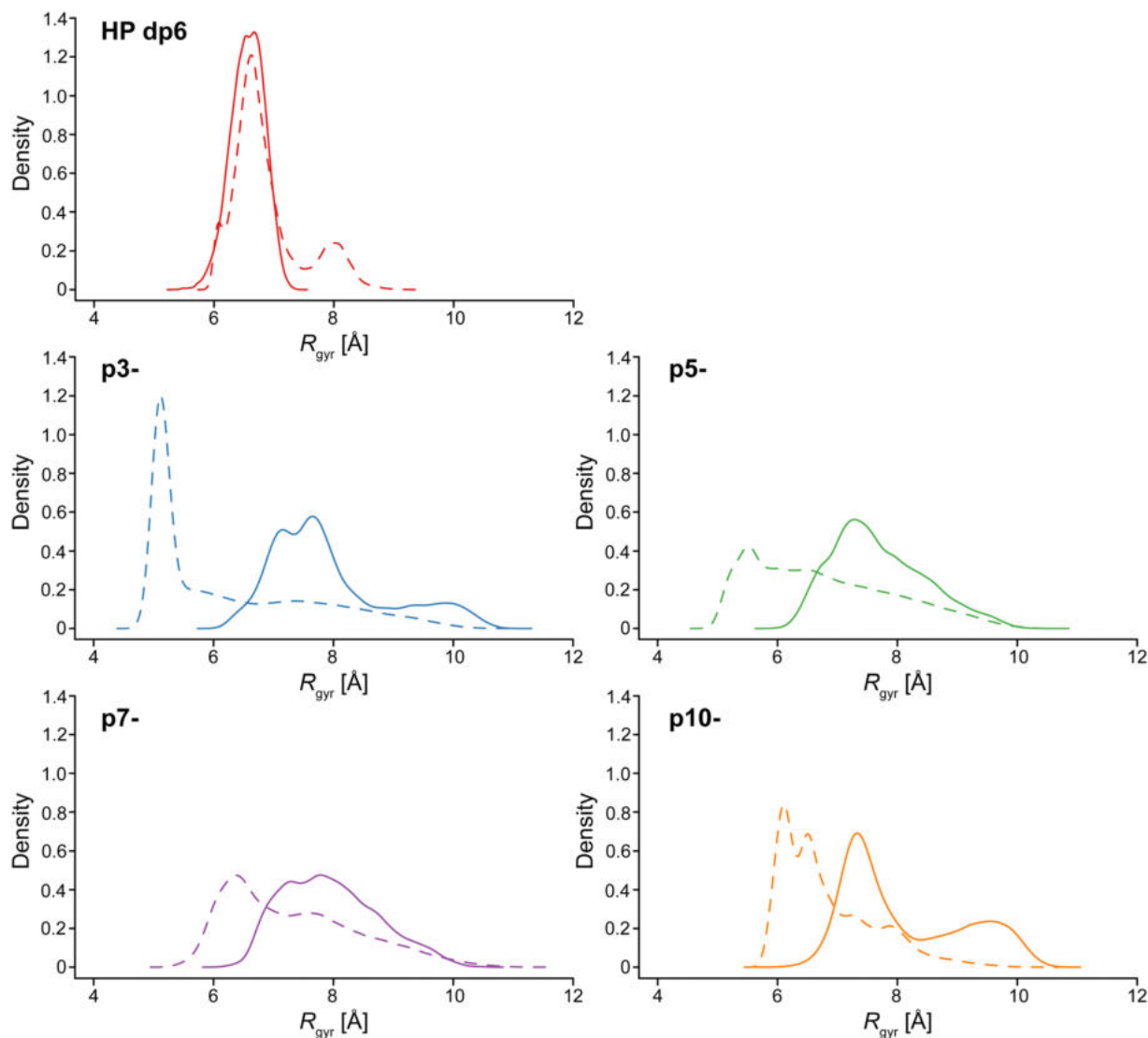


Fig. 9 Comparison of distributions of R_{gyr} values (kernel density estimation) between bound and unbound states in continuous and dashed lines, respectively, for HP dp6, p3-, p5-, p7-, and p10-.

H-bond formation was identified for residues 23–26, located at the side loop of IL-8. Noteworthy is that HP dp6 was more likely to form H-bonds in this site compared to the peptides. In the case of the ligands acting as H-bond donors and IL-8 as H-bond acceptor (Fig. S7, ESI[†]), only residues belonging to the C-terminal helical region of IL-8 were involved in H-bonds with occupancy over 0.1. The average frequency of H-bond formation was similar for all of the examined IL-8/ligand complexes. When ligands act as H-bond donors, the participation of the less charged peptides is higher, while p10- behaves similarly to HP dp6, establishing the fewest H-bonds.

Furthermore, we analyzed the differences in the distribution of the number of simultaneously established H-bonds for these particular residues depending on the bound ligand. When considering only the 14 IL-8 residues that establish most H-bonds as donors (*i.e.*, residues 23–26, 59–61, 64–66, 69, 70, 72, 73) based on Fig. 10, there is a clear difference between the distributions for p5- and especially for p10- in comparison to HP dp6, p3-, p7- (Fig. 11). Peptides p5- and p10- form more H-bonds simultaneously, which could be the explanation for

the observation that these peptides are more stable during the course of the MD simulation in terms of their movements on the protein surface. When analyzing contributions of each individual residue of IL-8 as an H-bond donor, certain patterns of specificity could be revealed (Fig. S8, ESI[†]). For H23 and K25, there is a clear trend of higher H-bond propensity with the increase in peptide charge, with the similarities between HP dp6 and p10-. In contrast, while the trend for the peptides remains the same for K59 (except for p10-), negligible number of H-bonds with this residue are established by HP dp6. Peptide p10- specifically interacts with H23, K25, and R73, while p5- interacts with E60, N61, Q64, K72. V66 and F70 establish specific H-bonds with HP dp6, although the normalized frequencies of these H-bonds are very low. Interestingly, when considering the amino acids of the known binding motif of IL-8 and HP (K25, R65, K69 and K72) especially p7- resembles the H-bond donor pattern overall very well, while the other high propensity amino acids are either in good agreement (H23) or in closest agreement considering the other peptide ligands (N61 and Q64).

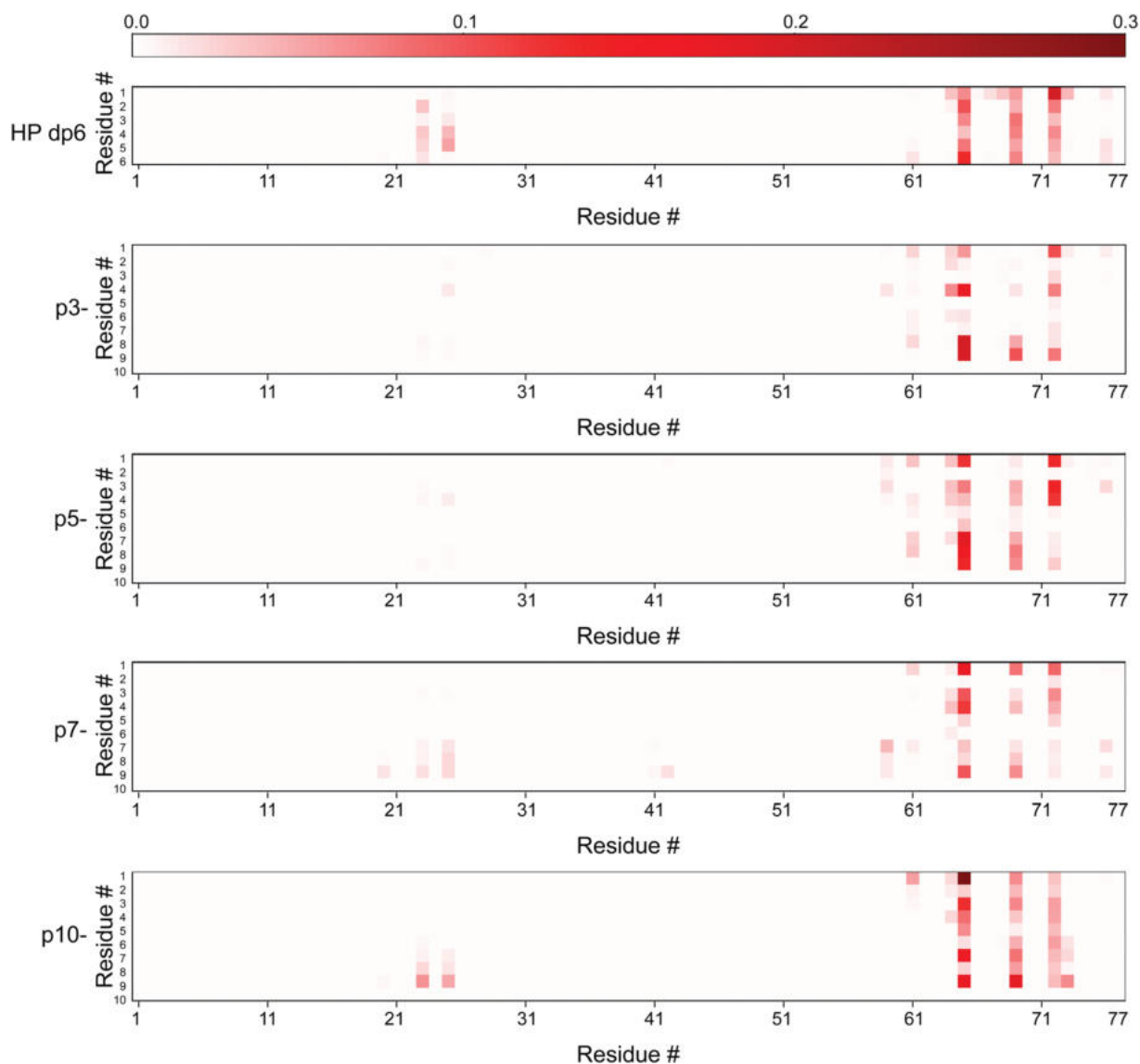


Fig. 10 Heatmap of hydrogen bond interactions between IL-8 (hydrogen atom donor) and HP dp6, p3-, p5-, p7- and p10-, averaged over IL-8 dimer subunits and across replicates of MD simulations. Protein residue numbers are shown on the x-axis, ligand residue numbers are shown on the y-axis. Color intensity corresponds to the frequency of the H-bond, *i.e.* the average fraction of MD trajectory frames in which the bond was formed.

Discussion

Recognition of regulatory proteins by GAGs on the cell surface is a key molecular process critically related to wound healing, cell growth, hemostasis, anticoagulation, tumor progression, inflammation and others.⁷⁸ The molecular characteristics of that interaction is the electrostatic attraction between GAGs, rich in negatively charged sulfate groups, and clusters of positively charged amino acids on the respective protein.⁷⁹ As electrostatic forces represent a rather undirected mode of interaction, the question arises how it can play the key role in such important biological processes. For instance, both pro- and anti-inflammatory cytokines, which induce highly adverse biological effects, expose clusters of basic amino acids on their surface and would be equally attracted by negatively charged GAGs on the cell surface. Although GAGs feature varying charge densities and distributions of negatively charged groups, their interactions strengths with the same protein are not too

different.⁴⁵ Furthermore, also other negatively charged polyelectrolytes (*i.e.* peptides, nucleic acids and nucleotides *etc.*) are found in the extracellular space and could also interact electrostatically with basic proteins. This triggers the question what factors other than electrostatics could provide an additional contribution to the recognition of proteins by GAG molecules.

We approached this question by a combination of experimental and computational methods comparing binding of a classical GAG HP dp6 to IL-8 with a small library of *de novo* designed short acidic decapeptides of varying charge density. Here, the most highly charged peptide is comparable with HP dp6 in terms of its net charge (Fig. 1). While HP is certainly one of the most highly charged GAG with 4 negative charges on average per dp (charge density: -0.52 to -0.57 C Å⁻¹ depending on conformation⁸⁰), also other GAGs with lower charge density such as hyaluronan (1 negative charge per dp, charge density -0.13 C Å⁻¹) or dermatan sulfate (2 negative charges per dp,

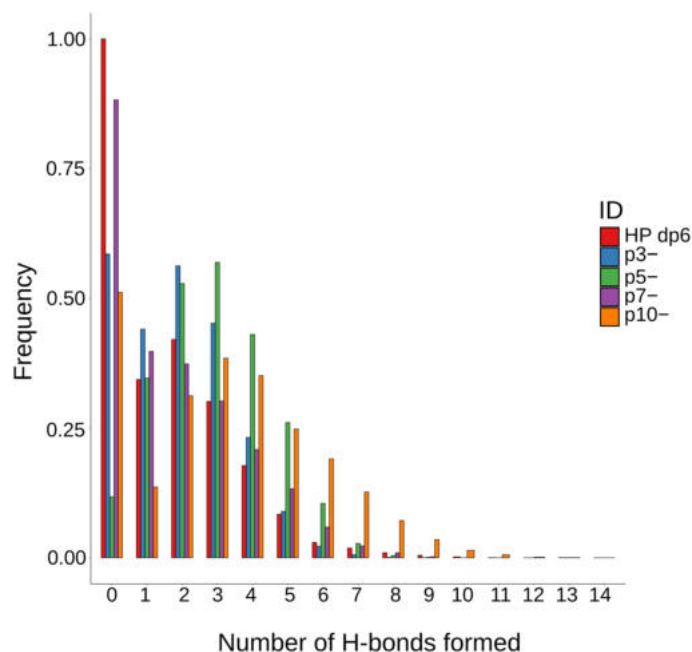


Fig. 11 Frequency of H-bonds established in each frame for 14 IL-8 residues that establish most H-bonds as donors (23–26, 59–61, 64–66, 69, 70, 72, 73). On the x-axis the number of H-bonds simultaneously observed in the same MD frames is provided.

charge density $-0.29 \text{ C } \text{\AA}^{-1}$) and chondroitin sulfate (2 negative charges per dp, charge density -0.28 or $-0.27 \text{ C } \text{\AA}^{-1}$ for chondroitin-4-sulfate and chondroitin-6-sulfate, respectively) are part of the extracellular matrix. The decapeptides varied in their number of charges between -3 and -10 , representing comparable charge densities between non-sulfated hyaluronan and monosulfated chondroitin or dermatan sulfate. The charge densities of the peptides are $-0.09 \text{ C } \text{\AA}^{-1}$ for p3-, $-0.14 \text{ C } \text{\AA}^{-1}$ for p5-, $-0.2 \text{ C } \text{\AA}^{-1}$ for p7-, and $-0.29 \text{ C } \text{\AA}^{-1}$ for p10- and thus well comparable to typical GAGs (Fig. 5D). Furthermore, we included two different negatively charged amino acids, aspartate and glutamate, which differ in the sidechain length similarly to 4-O and 6-O sulfation, where the latter is one CH_2 group further away from the sugar ring. The uncharged amino acids in the peptides had small sidechains (Gly, Ala) to avoid steric effects. For IL-8, it was shown that the binding strength for chondroitin sulfate with the same net charge but an altered position of the sulfate group, differs significantly – e.g. a sulfation in the 6-O position leads to higher affinity (black circles in Fig. 5D).^{45,46}

At the same time, it is worth taking into account that the chemical nature (the absence of sulfate groups in the peptides) and conformational preferences of peptides and GAGs as molecular classes in general are essentially different and,

therefore, their recognition elements are also likely to be distinguishable because of the differences in terms of degrees of freedom, charges distribution and as a consequence solvent-mediated interactions. All these factors could potentially affect the specificity features of their interactions with proteins.

Both GAG and the acidic peptides bind to an epitope defined by similar amino acid residues on IL-8 as shown by NMR titration (Fig. 3 and 4) and MD-based analysis (Table 1). Furthermore, binding is enhanced with increasing net charge from p3- to p10- and highly dependent on the ionic strength of the buffer (Fig. 4D and Table 2). This suggests that non-specific electrostatics may indeed be the key interaction for the formation of GAG/IL-8 complexes. The most involved IL-8 residues in peptide binding are R65, K69, and K72, which are part of the well-known α -helical BxxxBxxBB motif (including R73), where B stands for a basic residue.^{39,41,42} The second known GAG binding motif of IL-8 is localized in the loop connecting the N-terminal regions with the first β -strand and involves residues K20, H23 and K25. Here, all positively charged residues respond to peptide binding, while K25 responds strongest (Table 1 and Fig. 2–4, except for p10- in buffer without NaCl). These two binding motifs combined make the binding parallel to the helix on the side of IL-8 more favorable for HP and the peptides than the binding between the two helices, although both are observed in the MD.

Although similar values for the NMR CSP of IL-8 were induced by HP and the four model peptides, the ligand to protein ratio corresponding to the maximal achieved CSP varied drastically. For HP, the final titration step was achieved at a molar 1:1 ratio between HP and IL-8, higher GAG to IL-8 ratios lead to unspecific aggregation and loss of NMR signal intensity. In contrast, an up to 20-fold peptide excess still allowed reasonable NMR measurements. This experimental result suggests very different affinities of HP and the acidic peptides. Indeed, from plots of CSP vs. ligand concentration, apparent k_D values could be determined for three peptides (except p3-) varying between ~ 0.2 and 1.8 mM in high ionic strength buffer. As no saturation of the CSP was measured in the presence of HP, such an apparent k_D value could not be determined from the NMR measurements. Using Trp fluorescence at much lower protein concentration of $1 \text{ } \mu\text{M}$ in the same buffer, Schlörke *et al.* determined the k_D for GAG binding to IL-8, which varied between $5.5 \text{ } \mu\text{M}$ for HA and $2.0 \text{ } \mu\text{M}$ for HP dp6.^{45,46} These k_D values for GAGs and our NMR values for the acidic decapeptides are not directly comparable because very different protein concentrations were used. It is also known that IL-8 forms dimers with a monomer/dimer dissociation constant of $18 \pm 6 \text{ } \mu\text{M}$.⁸¹ Thus, the NMR measurements are determined for the IL-8 dimer while the Trp fluorescence data

Table 2 Estimate of the free energy of the system based on the comparison of bound and unbound ligand structures and the determined mean dissociation constant k_D for the analyzed complexes

	p3-	p5-	p7-	p10-
Calculated free energy [kcal mol^{-1}] (derived k_D [μM])	1.93 (26.0)	1.79 (20.4)	1.58 (14.3)	1.46 (11.7)

was acquired for monomers. Nevertheless, it is clear from Fig. 5D (closed symbols) that GAGs bind with much higher affinity to IL-8, which is also clearly suggested by the MD-based free energy calculations (Fig. 8). This is remarkable, since the charge density of the peptides and the lowly charged GAGs is comparable. Strikingly, the k_D values vary in the high ionic strength buffer from p5- to p10- by a factor of roughly 10, while in the low ionic strength buffer a factor of 20 is observed. This once again suggests that electrostatics play an important role, as ligand binding is improved in the absence of NaCl.

Since net electrostatics cannot fully explain the very different affinity of the negatively charged GAG or peptides for IL-8, the question arises what other factors contribute to the higher affinity of GAG for the protein. Our computational analysis can provide detailed insight into this question. Free energy calculations (Fig. 8) allowed the separation of electrostatic and van der Waals contributions as well as the impact of solvent that was previously demonstrated to be key in the protein–GAG interfaces.⁸² While the *in vacuo* electrostatic contribution scales approximately proportionally with the ligand charges, the van der Waals contribution is less favorable for the more charged peptides. Total electrostatic contributions are also essentially different for HP dp6 and acidic peptides: in the implicit solvent model, the solvent contribution compensates the electrostatic interactions between HP dp6 and the protein more effectively than between the peptide p10- and the protein. This could be explained by the substantially more hydrated nature of protein–HP interfaces in comparison to protein–peptide interfaces.⁸² Such energetic pattern differences could provide one of the clues for the GAG binding specificity.

An important difference between GAG and acidic peptides is also the nature of the negatively charged groups on either molecule. While sulfated GAGs contain both carboxylate and sulfate groups, the model peptides only feature carboxylate groups. The pK_a value of the Glu sidechain is 4.1 at 25 °C⁸³ and between 2.5 and 0.5 for the sulfate group of mucopolysaccharides.^{84,85} While shifts in these pK_a values can be observed, our results suggest that both groups are relatively strong electrolytes that should be fully charged at neutral pH used in our experiments. However, both functional groups vary in their physical properties as well as their hydrogen bond capacity. The carboxylate group is smaller ($V = 40.6 \text{ \AA}^3$) than the sulfate group ($V = 57.7 \text{ \AA}^3$) as determined by quantum chemical calculations (<https://www.spartan.com>). This results in a lower charge density of the sulfate group ($-11.5 \text{ mC \AA}^{-2}$) compared to the carboxylate group ($-16.6 \text{ mC \AA}^{-2}$). Also, the sulfate group has a higher dipole moment (4.7 D) than the carboxylate group (3.1 D). All these differences influence the charge–charge, charge–dipole, and dipole–dipole electrostatics. However, this hardly explains the difference in binding strength over three orders of magnitude between Hyaluronan dp6 ($5.5 \pm 1.3 \text{ \mu M}$)⁴⁵ and p5- ($\sim 1.7 \pm 0.6 \text{ mM}$), which both only exhibit carboxylate groups and have a comparable charge density.

The conformational changes of the ligands between free and bound form also play a role in the formation of ligand/IL-8 complexes. While HP is considered to be already relatively

elongated in free solution, CD experiments confirm that all peptides are in random coil conformation in solution (Fig. S6, ESI†). While backbone flexibility may help for the ligand to assume the ideal binding pose on the protein surface, the entropic change a coiled peptide ligand undergoes upon binding to a well-defined pose on the protein is rather unfavorable. This could be explained qualitatively by the entropic component of binding. As HP does not alter its radius of gyration or conformation upon protein binding (Fig. 9), it already represents a perfect ligand for IL-8 without the need for further structural adaption and binds with the highest affinity by far. In contrast, a remarkable adaptation of all peptides upon binding to a more elongated structure is observed (Fig. 9); R_g increases by $\sim 2.4 \text{ \AA}$ for p3-, $\sim 1.9 \text{ \AA}$ for p5-, and $\sim 1.4 \text{ \AA}$ for p7- and p10- upon IL-8-binding. This is accompanied with a decrease in entropy due to the loss of motional freedom, which also contributes to the lower binding energy. The suggestion that entropy may be a very important factor is also supported by the fuzzier binding poses assumed by the peptides in comparison to HP (Fig. 6). This looser and more flexible binding mode has more favorable entropy but also results in less optimal binding geometry and thus in overall lower affinity.

The MD results also show that the R_g of the free peptides increases with increasing charge density. This suggests that the entropic loss associated with IL-8-binding is reduced for the more highly charged peptides, in agreement with the higher affinity of these molecules for the protein. As the peptide charges are less screened in the absence of NaCl in the buffer, one would assume that the peptides are even more elongated under these conditions and thus closer to the peptide-bound conformation, which would also result in a smaller entropy loss contributing to the higher affinity of the peptides for IL-8 under these conditions, which amounts to a factor of 3.5, 7.2 and 7.4 for p5-, p7- and p10-, respectively.

The last but very important descriptor of the interaction of IL-8 with these ligands is hydrogen bonding. Based on the MD simulation, we could analyze and compare the hydrogen bond pattern of HP *vs.* the four negatively charged decapeptides (Fig. 10). At first sight, no major differences in hydrogen bond formation are observed between the individual ligands and the C-terminal binding motif of IL-8. However, our analysis reveals that the hydrogen bonds with residues from the more N-terminal binding motif are much less frequent (Fig. 10). There are more H-bonds observed at the same frames of the trajectory for p5- and p10- in comparison to HP dp6, p3- and p7-. But taking only the H-bonds of the binding motive into account, p7- resembles the pattern of HP closest.

Conclusion

Taken together, although long ranged electrostatics is clearly a key player in IL-8/ligand interaction, it cannot fully explain the much higher affinity of GAGs for regulatory proteins compared to other polyelectrolytes, *i.e.*, acidic peptides. Further aspects such as the ligand flexibility and conformational changes upon

binding associated with an entropic contribution, ligand structural adaptability, hydrogen bond capacity, participation of the solvent in the establishment of the interface, and exact localization of negatively charged groups as well as their (chemical) nature have to be taken into account. Physiologically, it is highly important that regulatory proteins do not bind polyacidic ligands — which are found abundantly in extracellular fluids — as strongly as the cell surface GAGs. It appears that a fine balance of physical interactions favors GAG binding over binding of other ligands such as polyions, nucleic acid, or peptide fragments. To describe each of the relevant factors in more detail, investigations with more similar functional groups should be conducted. Nevertheless, all aforementioned factors act synergistically in providing some specificity of certain GAGs to preferentially bind regulatory proteins and direct them to the respective cells. This directed motion, mediated by a well-balanced network of GAGs of varying charge density and distribution, has been referred to as the “electrostatic band-pass”.⁸⁶ While in principle an appealing concept, the current comparison between negatively charged GAGs and peptides suggests that additional factors than electrostatics need to be considered to adequately describe the filtering of regulatory proteins by the extracellular matrix. This contributes a step towards understanding the specificity of protein–GAG biomolecular systems and the sulfation code, which remain highly challenging for analysis at the atomistic level.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This research was funded by the German Research Foundation (DFG) project numbers 426256511 and 58397982 and the National Science Centre of Poland (Narodowe Centrum Nauki, grant number UMO-2018/31/G/ST4/00246). The molecular docking calculations were performed on the ZIH cluster at TU Dresden (grant p_gag). The molecular dynamics simulations were performed on the PROMETHEUS cluster provided by Polish Grid Infrastructure (PL-GRID, grants kchtkat, plggagi, PLG/2020/014274 and PLG/2020/014275) as well as on the local “piasek” cluster. We thank Dr Albert Smith-Penzel for proof reading of the manuscript.

References

- 1 J. D. Esko, K. Kimata and U. Lindahl, in *Essentials of Glycobiology*, ed. A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart and M. E. Etzler, Cold Spring Harbor (NY), 2nd edn, 2009, Proteoglycans and Sulfated Glycosaminoglycans.
- 2 J. Schiller and D. Huster, New methods to study the composition and structure of the extracellular matrix in natural and bioengineered tissues, *Biomatter*, 2012, **2**, 115–131.
- 3 H. Habuchi, O. Habuchi and K. Kimata, Sulfation pattern in glycosaminoglycan: does it have a code?, *Glycoconjugate J.*, 2004, **21**, 47–52.
- 4 J. Gallagher, Fell-Muir Lecture: heparan sulphate and the art of cell regulation: a polymer chain conducts the protein orchestra, *Int. J. Exp. Pathol.*, 2015, **96**, 203–231.
- 5 N. K. Karamanos, Z. Piperigkou, A. D. Theocharis, H. Watanabe, M. Franchi, S. Baud, S. Brézillon, M. Götte, A. Passi, D. Vigetti, S. Ricard-Blum, R. D. Sanderson, T. Neill and R. V. Iozzo, Proteoglycan Chemical Diversity Drives Multifunctional Cell Regulation and Therapeutics, *Chem. Rev.*, 2018, **118**, 9152–9232.
- 6 M. S. G. Pavão, Glycosaminoglycans analogs from marine invertebrates: structure, biological effects, and potential as new therapeutics, *Front. Cell. Infect. Microbiol.*, 2014, **4**, 123.
- 7 K. Karamanou, D. C. R. Espinosa, A. Fortuna-Costa and M. S. G. Pavão, Biological function of unique sulfated glycosaminoglycans in primitive chordates, *Glycoconjugate J.*, 2017, **34**, 277–283.
- 8 P. Chiodelli, A. Bugatti, C. Urbinati and M. Rusnati, Heparin/Heparan sulfate proteoglycans glycomic interactome in angiogenesis: biological implications and therapeutic use, *Molecules*, 2015, **20**, 6342–6388.
- 9 F. M. Spinelli, D. L. Vitale, G. Demarchi, C. Cristina and L. Alaniz, The immunological effect of hyaluronan in tumor angiogenesis, *Clin. Transl. Immunol.*, 2015, **4**, e52.
- 10 S. D. Vallet, O. Clerc and S. Ricard-Blum, Glycosaminoglycan-Protein Interactions: The First Draft of the Glycosaminoglycan Interactome, *J. Histochem. Cytochem.*, 2021, **69**, 93–104.
- 11 Y. Monneau, F. Arenzana-Seisdedos and H. Lortat-Jacob, The sweet spot: how GAGs help chemokines guide migrating cells, *J. Leukoc. Biol.*, 2016, **99**, 935–953.
- 12 A. E. I. Proudfoot, Z. Johnson, P. Bonvin and T. M. Handel, Glycosaminoglycan Interactions with Chemokines Add Complexity to a Complex System, *Pharmaceuticals*, 2017, **10**, 70.
- 13 G. Künze, D. Huster and S. A. Samsonov, Investigation of the structure of regulatory proteins interacting with glycosaminoglycans by combining NMR spectroscopy and molecular modeling – the beginning of a wonderful friendship, *Biol. Chem.*, 2021, **402**, 1337–1355.
- 14 S. Morla, Glycosaminoglycans and Glycosaminoglycan Mimetics in Cancer and Inflammation, *Int. J. Mol. Sci.*, 2019, **20**, 1963.
- 15 D. Vitale, S. Kumar Katakam, B. Greve, B. Jang, E.-S. Oh, L. Alaniz and M. Götte, Proteoglycans and glycosaminoglycans as regulators of cancer stem cell function and therapeutic resistance, *FEBS J.*, 2019, **286**, 2870–2882.
- 16 N. Afratis, C. Gialeli, D. Nikitovic, T. Tsegenidis, E. Karousou, A. D. Theocharis, M. S. Pavão, G. N. Tzanakakis and N. K. Karamanos, Glycosaminoglycans: key players in cancer cell biology and treatment, *FEBS J.*, 2012, **279**, 1177–1197.
- 17 T. Hosono-Fukao, S. Ohtake-Niimi, H. Hoshino, M. Britschgi, H. Akatsu, M. M. Hossain, K. Nishitsuji, T. H. van Kuppevelt, K. Kimata, M. Michikawa, T. Wyss-Coray and K. Uchimura,

- Heparan sulfate subdomains that are degraded by Sulf accumulate in cerebral amyloid β plaques of Alzheimer's disease: evidence from mouse models and patients, *Am. J. Pathol.*, 2012, **180**, 2056–2067.
- 18 N. Iwahashi, M. Ikezaki, T. Nishikawa, N. Namba, T. Ohgita, H. Saito, Y. Ihara, T. Shimanouchi, K. Ino, K. Uchimura and K. Nishitsuji, Sulfated glycosaminoglycans mediate prion-like behavior of p53 aggregates, *Proc. Natl. Acad. Sci. U. S. A.*, 2020, **117**, 33225–33234.
 - 19 C. M. Simonaro, M. D'Angelo, X. He, E. Eliyahu, N. Shtraizent, M. E. Haskins and E. H. Schuchman, Mechanism of glycosaminoglycan-mediated bone and joint disease: implications for the mucopolysaccharidoses and other connective tissue diseases, *Am. J. Pathol.*, 2008, **172**, 112–122.
 - 20 P. Johnson, A. A. Arif, S. S. M. Lee-Sayer and Y. Dong, Hyaluronan and Its Interactions With Immune Cells in the Healthy and Inflamed Lung, *Front. Immunol.*, 2018, **9**, 2787.
 - 21 D. Scharnweber, L. Hübner, S. Rother, U. Hempel, U. Anderegg, S. A. Samsonov, M. T. Pisabarro, L. Hofbauer, M. Schnabelrauch, S. Franz, J. Simon and V. Hintze, Glycosaminoglycan derivatives: promising candidates for the design of functional biomaterials, *J. Mater. Sci.: Mater. Med.*, 2015, **26**, 232.
 - 22 D. Xu and J. D. Esko, Demystifying heparan sulfate-protein interactions, *Annu. Rev. Biochem.*, 2014, **83**, 129–157.
 - 23 D. Shi, A. Sheng and L. Chi, Glycosaminoglycan-Protein Interactions and Their Roles in Human Disease, *Front. Mol. Biosci.*, 2021, **8**, 639666.
 - 24 in *Essentials of Glycobiology*, A. Varki, R. D. Cummings, J. D. Esko, P. Stanley, G. W. Hart, M. Aebi, D. Mohnen, T. Kinoshita, N. H. Packer, J. H. Prestegard, R. L. Schnaar and P. H. Seeberger, (ed.) Cold Spring Harbor (NY), 4th edn, 2022.
 - 25 M. Guerrini, S. Guglieri, B. Casu, G. Torri, P. Mourier, C. Boudier and C. Viskov, Antithrombin-binding octasaccharides and role of extensions of the active pentasaccharide sequence in the specificity and strength of interaction. Evidence for very high affinity induced by an unusual glucuronic acid residue, *J. Biol. Chem.*, 2008, **283**, 26662–26675.
 - 26 L. Jin, J. P. Abrahams, R. Skinner, M. Petitou, R. N. Pike and R. W. Carrell, The anticoagulant activation of antithrombin by heparin, *Proc. Natl. Acad. Sci. U. S. A.*, 1997, **94**, 14683–14688.
 - 27 V. Hintze, S. A. Samsonov, M. Anselmi, S. Moeller, J. Becher, M. Schnabelrauch, D. Scharnweber and M. T. Pisabarro, Sulfated glycosaminoglycans exploit the conformational plasticity of bone morphogenetic protein-2 (BMP-2) and alter the interaction profile with its receptor, *Biomacromolecules*, 2014, **15**, 3083–3092.
 - 28 N. Panitz, S. Theisgen, S. A. Samsonov, J.-P. Gehrcke, L. Baumann, K. Bellmann-Sickert, S. Köhling, M. T. Pisabarro, J. Rademann, D. Huster and A. G. Beck-Sickinger, The structural investigation of glycosaminoglycan binding to CXCL12 displays distinct interaction sites, *Glycobiology*, 2016, **26**, 1209–1221.
 - 29 S. Rother, S. A. Samsonov, T. Hofmann, J. Blaszkiewicz, S. Köhling, S. Moeller, M. Schnabelrauch, J. Rademann, S. Kalkhof, M. von Bergen, M. T. Pisabarro, D. Scharnweber and V. Hintze, Structural and functional insights into the interaction of sulfated glycosaminoglycans with tissue inhibitor of metalloproteinase-3 – A possible regulatory role on extracellular matrix homeostasis, *Acta Biomater.*, 2016, **45**, 143–154.
 - 30 L. Koehler, S. Samsonov, S. Rother, S. Vogel, S. Köhling, S. Moeller, M. Schnabelrauch, J. Rademann, U. Hempel, M. T. Pisabarro, D. Scharnweber and V. Hintze, Sulfated Hyaluronan Derivatives Modulate TGF- β 1:Receptor Complex Formation: Possible Consequences for TGF- β 1 Signaling, *Sci. Rep.*, 2017, **7**, 1210.
 - 31 S. T. Olson, H. R. Halvorson and I. Björk, Quantitative characterization of the thrombin-heparin interaction. Discrimination between specific and nonspecific binding models, *J. Biol. Chem.*, 1991, **266**, 6342–6352.
 - 32 V. H. Pomin and B. Mulloy, Current structural biology of the heparin interactome, *Curr. Opin. Struct. Biol.*, 2015, **34**, 17–25.
 - 33 A. Almond, Multiscale modeling of glycosaminoglycan structure and dynamics: current methods and challenges, *Curr. Opin. Struct. Biol.*, 2018, **50**, 58–64.
 - 34 J. Sage, F. Mallèvre, F. Barbarin-Costes, S. A. Samsonov, J.-P. Gehrcke, M. T. Pisabarro, E. Perrier, S. Schnebert, A. Roget, T. Livache, C. Nizard, G. Lalmanach and F. Lécaille, Binding of chondroitin 4-sulfate to cathepsin S regulates its enzymatic activity, *Biochemistry*, 2013, **52**, 6487–6498.
 - 35 A. Penk, L. Baumann, D. Huster and S. A. Samsonov, NMR and molecular modeling reveal specificity of the interactions between CXCL14 and glycosaminoglycans, *Glycobiology*, 2019, **29**, 715–725.
 - 36 C. I. Gama, S. E. Tully, N. Sotogaku, P. M. Clark, M. Rawat, N. Vaidehi, W. A. Goddard, A. Nishi and L. C. Hsieh-Wilson, Sulfation patterns of glycosaminoglycans encode molecular recognition and activity, *Nat. Chem. Biol.*, 2006, **2**, 467–473.
 - 37 P. R. B. Joseph, P. D. Mosier, U. R. Desai and K. Rajarathnam, Solution NMR characterization of chemokine CXCL8/IL-8 monomer and dimer binding to glycosaminoglycans: structural plasticity mediates differential binding interactions, *Biochem. J.*, 2015, **472**, 121–133.
 - 38 A. D. Cardin and H. J. Weintraub, Molecular modeling of protein-glycosaminoglycan interactions, *Arteriosclerosis*, 1989, **9**, 21–32.
 - 39 G. S. Kuschert, A. J. Hoogewerf, A. E. Proudfoot, C. W. Chung, R. M. Cooke, R. E. Hubbard, T. N. Wells and P. N. Sanderson, Identification of a glycosaminoglycan binding surface on human interleukin-8, *Biochemistry*, 1998, **37**, 11193–11201.
 - 40 D. Spillmann, D. Witt and U. Lindahl, Defining the interleukin-8-binding domain of heparan sulfate, *J. Biol. Chem.*, 1998, **273**, 15487–15493.

- 41 W. Bitomsky and R. C. Wade, Docking of Glycosaminoglycans to Heparin-Binding Proteins: Validation for aFGF, bFGF, and Antithrombin and Application to IL-8, *J. Am. Chem. Soc.*, 1999, **121**, 3004–3013.
- 42 H. Lortat-Jacob, A. Grosdidier and A. Imberty, Structural diversity of heparan sulfate binding domains in chemokines, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 1229–1234.
- 43 E. Krieger, E. Geretti, B. Brandner, B. Goger, T. N. Wells and A. J. Kungl, A structural and dynamic model for the interaction of interleukin-8 and glycosaminoglycans: support from isothermal fluorescence titrations, *Proteins*, 2004, **54**, 768–775.
- 44 N. S. Gandhi and R. L. Mancera, Molecular dynamics simulations of CXCL-8 and its interactions with a receptor peptide, heparin fragments, and sulfated linked cyclitols, *J. Chem. Inf. Model.*, 2011, **51**, 335–358.
- 45 A. Pichert, S. A. Samsonov, S. Theisgen, L. Thomas, L. Baumann, J. Schiller, A. G. Beck-Sickinger, D. Huster and M. T. Pisabarro, Characterization of the interaction of interleukin-8 with hyaluronan, chondroitin sulfate, dermatan sulfate and their sulfated derivatives by spectroscopy and molecular modeling, *Glycobiology*, 2012, **22**, 134–145.
- 46 D. Schlorke, L. Thomas, S. A. Samsonov, D. Huster, J. Arnhold and A. Pichert, The influence of glycosaminoglycans on IL-8-mediated functions of neutrophils, *Carbohydr. Res.*, 2012, **356**, 196–203.
- 47 K. Möbius, K. Nordsieck, A. Pichert, S. A. Samsonov, L. Thomas, J. Schiller, S. Kalkhof, M. Teresa Pisabarro, A. G. Beck-Sickinger and D. Huster, Investigation of lysine side chain interactions of interleukin-8 with heparin and other glycosaminoglycans studied by a methylation-NMR approach, *Glycobiology*, 2013, **23**, 1260–1269.
- 48 K. Nordsieck, A. Pichert, S. A. Samsonov, L. Thomas, C. Berger, M. T. Pisabarro, D. Huster and A. G. Beck-Sickinger, Residue 75 of interleukin-8 is crucial for its interactions with glycosaminoglycans, *ChemBioChem*, 2012, **13**, 2558–2566.
- 49 T. Hofmann, S. A. Samsonov, A. Pichert, K. Lemmnitzer, J. Schiller, D. Huster, M. T. Pisabarro, M. von Bergen and S. Kalkhof, Structural analysis of the interleukin-8/glycosaminoglycan interactions by amide hydrogen/deuterium exchange mass spectrometry, *Methods*, 2015, **89**, 45–53.
- 50 P. R. B. Joseph, K. V. Sawant, J. Iwahara, R. P. Garofalo, U. R. Desai and K. Rajarathnam, Lysines and Arginines play non-redundant roles in mediating chemokine-glycosaminoglycan interactions, *Sci. Rep.*, 2018, **8**, 12289.
- 51 G. Paiardi, M. Milanese, R. C. Wade, P. D'Ursi and M. Rusnati, A Bittersweet Computational Journey among Glycosaminoglycans, *Biomolecules*, 2021, **11**, 739.
- 52 W. Lee, M. Tonelli and J. L. Markley, NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy, *Bioinformatics*, 2015, **31**, 1325–1327.
- 53 S. Berkamp, S. H. Park, A. A. de Angelis, F. M. Marassi and S. J. Opella, Structure of monomeric Interleukin-8 and its interactions with the N-terminal Binding Site-I of CXCR1 by solution NMR spectroscopy, *J. Biomol. NMR*, 2017, **69**, 111–121.
- 54 S. Mori, C. Abeygunawardana, M. O. Johnson and P. C. van Zijl, Improved sensitivity of HSQC spectra of exchanging protons at short interscan delays using a new fast HSQC (FHSQC) detection scheme that avoids water saturation, *J. Magn. Reson., Ser. B*, 1995, **108**, 94–98.
- 55 V. Sklenar, M. Piotto, R. Leppik and V. Saudek, Gradient-Tailored Water Suppression for ^1H - ^{15}N HSQC Experiments Optimized to Retain Full Sensitivity, *J. Magn. Reson., Ser. A*, 1993, **102**, 241–245.
- 56 A. Shaka, P. Barker and R. Freeman, Computer-optimized decoupling scheme for wideband applications and low-level operation, *J. Magn. Reson.*, 1985, **64**, 547–552.
- 57 M. P. Williamson, Using chemical shift perturbation to characterise ligand binding, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2013, **73**, 1–16.
- 58 E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, UCSF Chimera—a visualization system for exploratory research and analysis, *J. Comput. Chem.*, 2004, **25**, 1605–1612.
- 59 G. M. Clore, E. Appella, M. Yamada, K. Matsushima and A. M. Gronenborn, Three-dimensional structure of interleukin 8 in solution, *Biochemistry*, 1990, **29**, 1689–1696.
- 60 D. A. Case, R. M. B. D. S. Cerutti, T. E. Cheatham, III, T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T. S. Lee, S. LeGrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K. M. Merz, G. Monard, H. Nguyen, H. T. Nguyen, I. Omelyan, A. Onufriev, D. R. Roe, A. Roitberg, C. Sagui, C. L. Simmerling, W. M. Botello-Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, L. Xiao and P. A. Kollman, *AMBER 2016*, University of California, San Francisco, 2016.
- 61 B. Mulloy, M. J. Forster, C. Jones and D. B. Davies, N.m.r. and molecular-modelling studies of the solution conformation of heparin, *Biochem. J.*, 1993, **293**, 849–858.
- 62 G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew and A. J. Olson, Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, *J. Comput. Chem.*, 1998, **19**, 1639–1662.
- 63 M. Ester, H.-P. Kriegel, J. Sander and X. Xu, *Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining (KDD)*, AAAI Press, Portland, Oregon, 1996, pp. 226–231.
- 64 J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB, *J. Chem. Theory Comput.*, 2015, **11**, 3696–3713.
- 65 K. K. Bojarski, A. K. Sieradzian and S. A. Samsonov, Molecular dynamics insights into protein-glycosaminoglycan systems from microsecond-scale simulations, *Biopolymers*, 2019, **110**, e23252.
- 66 B. L. Welch, The generalisation of student's problems when several different population variances are involved, *Biometrika*, 1947, **34**, 28–35.
- 67 M. Marcisz, B. Huard, A. G. Lipska and S. A. Samsonov, Further analyses of APRIL/APRIL-receptor/glycosaminoglycan

- interactions by biochemical assays linked to computational studies, *Glycobiology*, 2021, **31**, 772–786.
- 68 D. R. Roe and T. E. Cheatham, PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data, *J. Chem. Theory Comput.*, 2013, **9**, 3084–3095.
 - 69 R Core Team, R: A language and environment for statistical computing, www.R-project.org/, (accessed 19 August 2022).
 - 70 C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. Del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke and T. E. Oliphant, Array programming with NumPy, *Nature*, 2020, **585**, 357–362.
 - 71 W. McKinney, in *Proceedings of the 9th Python in Science Conference SciPy 2010*, ed. S. van der Walt and J. Millman, 2010, pp. 56–61.
 - 72 J. D. Hunter, Matplotlib: A 2D Graphics Environment, *Comput. Sci. Eng.*, 2007, **9**, 90–95.
 - 73 T. Williams and C. Kelley, Gnuplot 4.5: An interactive plotting program, <http://gnuplot.inf>, (accessed 19 August 2022).
 - 74 W. Humphrey, A. Dalke and K. Schulten, VMD: visual molecular dynamics, *J. Mol. Graphics*, 1996, **14**(33–38), 27–28.
 - 75 K. W. Walker and R. A. Bradshaw, Yeast methionine aminopeptidase I. Alteration of substrate specificity by site-directed mutagenesis, *J. Biol. Chem.*, 1999, **274**, 13403–13409.
 - 76 L. He, S. André, H.-C. Siebert, H. Helmholz, B. Niemeyer and H.-J. Gabius, Detection of Ligand- and Solvent-Induced Shape Alterations of Cell-Growth-Regulatory Human Lectin Galectin-1 in Solution by Small Angle Neutron and X-Ray Scattering, *Biophys. J.*, 2003, **85**, 511–524.
 - 77 S. Baranova, F. V. Tuzikov, O. D. Zakharova, N. A. Tuzikova, C. Calmels, S. Litvak, L. Tarrago-Litvak, V. Parissi and G. A. Nevinsky, Small-angle X-ray characterization of the nucleoprotein complexes resulting from DNA-induced oligomerization of HIV-1 integrase, *Nucleic Acids Res.*, 2007, **35**, 975–987.
 - 78 N. S. Gandhi and R. L. Mancera, The structure of glycosaminoglycans and their interactions with proteins, *Chem. Biol. Drug Des.*, 2008, **72**, 455–482.
 - 79 A. Imberty, H. Lortat-Jacob and S. Pérez, Structural view of glycosaminoglycan-protein interactions, *Carbohydr. Res.*, 2007, **342**, 430–439.
 - 80 S. A. Samsonov, L. Bichmann and M. T. Pisabarro, Coarse-grained model of glycosaminoglycans, *J. Chem. Inf. Model.*, 2015, **55**, 114–124.
 - 81 S. D. Burrows, M. L. Doyle, K. P. Murphy, S. G. Franklin, J. R. White, I. Brooks, D. E. McNulty, M. O. Scott, J. R. Knutson and D. Porter, Determination of the monomer-dimer equilibrium of interleukin-8 reveals it is a monomer at physiological concentrations, *Biochemistry*, 1994, **33**, 12741–12745.
 - 82 S. A. Samsonov, J. Teyra and M. T. Pisabarro, Docking glycosaminoglycans to proteins: analysis of solvent inclusion, *J. Comput.-Aided Mol. Des.*, 2011, **25**, 477–489.
 - 83 D. L. Nelson and M. M. Cox, *Lehninger Biochemie*, Springer, Berlin, Heidelberg, New York, Barcelona, Hongkong, London, Mailand, Paris, Tokio, 3rd edn, 2001.
 - 84 K. E. Kuettner and A. Lindenbaum, Analysis of mucopolysaccharides in partially aqueous media, *Biochim. Biophys. Acta*, 1965, **101**, 223–225.
 - 85 M. Remko, R. Broer and P. T. van Duijnen, How acidic are monomeric structural units of heparin?, *Chem. Phys. Lett.*, 2013, **590**, 187–191.
 - 86 O. Lieleg, R. M. Baumgärtel and A. R. Bausch, Selective filtering of particles by the extracellular matrix: an electrostatic bandpass, *Biophys. J.*, 2009, **97**, 1569–1577.

**Publication 3. Extension of the
SUGRES-1P Coarse-Grained Model of
Polysaccharides to Heparin.**

Extension of the SUGRES-1P Coarse-Grained Model of Polysaccharides to Heparin

Annemarie Danielsson, Sergey A. Samsonov, Adam Liwo, and Adam K. Sieradzan*

Cite This: *J. Chem. Theory Comput.* 2023, 19, 6023–6036

Read Online

ACCESS |



Metrics & More

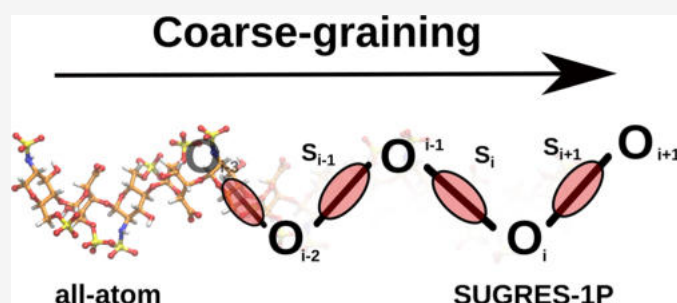


Article Recommendations



Supporting Information

ABSTRACT: Heparin is an unbranched periodic polysaccharide composed of negatively charged monomers and involved in key biological processes, including anticoagulation, angiogenesis, and inflammation. Its structure and dynamics have been studied extensively using experimental as well as theoretical approaches. The conventional approach of computational chemistry applied to the analysis of biomolecules is all-atom molecular dynamics, which captures the interactions of individual atoms by solving Newton's equation of motion. An alternative is molecular dynamics simulations using coarse-grained models of biomacromolecules, which offer a reduction of the representation and consequently enable us to extend the time and size scale of simulations by orders of magnitude. In this work, we extend the UNified COarse-grained RaiNed (UNICORN) model of biological macromolecules developed in our laboratory to heparin. We carried out extensive tests to estimate the optimal weights of energy terms of the effective energy function as well as the optimal Debye–Hückel screening factor for electrostatic interactions. We applied the model to study unbound heparin molecules of polymerization degree ranging from 6 to 68 residues. We compare the obtained coarse-grained heparin conformations with models obtained from X-ray diffraction studies of heparin. The SUGRES-1P force field was able to accurately predict the general shape and global characteristics of heparin molecules.



1. INTRODUCTION

Glycosaminoglycan (GAG) heparin (HP) is a linear, anionic polysaccharide composed of repeating disaccharide units of 2-O-sulfated iduronic acid (IdoA2S) and 6-O-sulfated and N-sulfated glucosamine (GlcNS6S).¹ With a -4 net charge per disaccharide unit, HP is the most negatively charged polysaccharide among all GAGs. Synthesized in mast cells of connective tissues as part of proteoglycans, after its synthesis HP is stored in secretory granules of mast cells² and released into the extracellular matrix, where it is involved in a range of important biological processes, such as angiogenesis,^{3,4} anticoagulation,⁵ cell proliferation,⁶ cell adhesion,^{7,8} and cell migration.^{9–11} HP fulfills its role via interactions with protein partners. The disruption of protein–HP interactions can lead to the development of diseases and pathologies, including tumor growth and metastasis,^{3,12–14} neurodegenerative,^{15–18} prion diseases,^{19–21} and autoimmune disorders.^{22,23} The participation of HP in essential biological processes and pathways renders it a promising and interesting target in medicine.²⁴

The structure and dynamics of HP as well as its interactions with proteins have been investigated by various experimental techniques, including nuclear magnetic resonance (NMR),^{25–28} X-ray diffraction,^{29–32} surface plasmon resonance,^{33–35} mass spectrometry,^{36,37} and capillary electrophoresis.^{38–41} However, experimental techniques face a range of challenges in the study of HP and other GAGs, mainly stemming from the periodicity,⁴²

considerable length, and high molecular weight of native GAGs,^{43–45} diversity of sequences and sulfation patterns,⁴⁴ conformational flexibility,⁴⁶ and tendency to cause oligomerization and precipitation of proteins.⁴⁵

To fully elucidate the structure and dynamics of biomolecules, computational approaches are often used alongside experimental techniques.^{47,48} A standard computational method is all-atom molecular dynamics (MD) simulations, in which Newtonian laws of motion are applied to determine the movement of each individual atom of the simulated system over time.^{47,48} The trajectories resulting from MD simulations, therefore, show the dynamic behavior of biomolecules and their complexes. MD simulations of HP molecules of different lengths were used to study their conformation in solution^{49,50} and in complexes with proteins.^{51–54}

While MD simulations are invaluable in the elucidation of the three-dimensional structures of GAGs and the atomistically detailed mechanisms of binding with proteins, they might also

Received: May 17, 2023

Published: August 16, 2023



fail to produce satisfactory results. Native GAG chains can achieve sizes of up to 100 kDa (degree of polymerization up to approximately 30,000,000)⁵⁵ and are, therefore, too large to be simulated using classical all-atom MD approaches. The addition of protein receptors, solvents, and ions increases the system size even when modeling only short HP fragments, which reduces the timescales that can be covered by all-atom MD techniques. The length of HP molecules together with their high flexibility result in a large conformational space that may be sampled insufficiently. Additionally, the binding of GAGs by proteins is predominantly electrostatic in nature and GAG-binding sites are most often patches of positively charged amino acid residues on the protein surface, as opposed to deep binding pockets.^{56,57} This, as well as multipose binding, a phenomenon in which multiple binding poses of comparable binding energy coexist, causes difficulties in the determination and prediction of the exact binding poses of GAGs in protein-GAG systems.^{58,59}

An alternative to classical all-atom MD methods that allows the coverage of larger spatial and temporal scales is coarse-grained (CG) modeling. CG approaches are based on a reduction of the representation of a system studied by introducing the so-called pseudoatoms that represent groups of atoms. When compared to all-atom MD representations, coarse graining of the system representation significantly reduces the number of degrees of freedom. As a result, the computational resources required to simulate the system are reduced. At the same time, the reduction in resolution is inherently accompanied by a reduction in detail and accuracy, so care has to be taken in the construction of the CG model in order to keep key features of the system and retain the essential characteristics of the biomolecules.⁶⁰

In spite of the significance and ubiquity of GAGs in the biological context, only a handful of CG models have been developed for this class of molecules, possibly due to the complexity of their sequence and structure.⁶¹ The first CG model of GAGs was proposed by Bathe et al. in 2005,⁶² designed for the modeling of chondroitin, chondroitin sulfate, and hyaluronic acid. The sugar residue was represented by a total of five CG beads: two carbon atoms and an oxygen atom to model the glycosidic linkage, an interaction site corresponding to the center of mass, used to model steric interactions, and an interaction site corresponding to the center of charge, modeling the electrostatic interactions. The model was able to correctly reproduce the conformation of the GAGs studied as well as their titration characteristics.⁶²

Sattelle et al.^{63,64} designed a CG model of heparan sulfate by reducing the representation of the sugar residue to two interaction centers: one representing the sugar ring and the second one representing the glycosidic linkage oxygen atom. The model was successfully applied to the study of heparan sulfate of large size⁶⁴ and heparan sulfate-containing proteoglycans.⁶³

A more detailed model, comprising 28 different pseudoatoms for the simulation of 17 different GAG types was proposed by Samsonov et al.⁶⁵ The pseudoatoms corresponded to different functional groups of the GAGs, including oxygens of the glycosidic linkages, sulfate groups, the carboxylate and *N*-acetyl groups, centers of mass of the pyranose rings and the CH₂OH group of the C₆ atom of *N*-acetyl-glucosamine and *N*-acetyl-galactosamine. The number of CG pseudoatoms per repeating unit ranged from 2 to 5 depending on the type and sulfation pattern of the sugar residue (*N*-acetyl-galactosamine/*N*-acetyl-glucosamine and glucuronic acid/iduronic acid) and its location

in the GAG chain (internal/terminal). Geometrical parameters for virtual bonds, virtual bond angles, and virtual bond torsional angles were obtained from all-atom MD simulations of different GAGs, while parameters of nonbonded interactions were obtained using steered MD. Pseudoatoms representing carboxyl and sulfate groups were assigned charges equal to -1 , while all other pseudoatom charges were set to 0. The CG model achieved good results in the modeling of global and local characteristics of GAG chains of different lengths.⁶⁵

Another CG model of GAGs is based on the SUGRES-1P model of carbohydrates, which is a part of the UNified COarse gRaiNed (UNICORN) model for biomacromolecules.^{66,67} The UNICORN model relies on a strictly physics-based approach, reducing the representation to only one or two CG sites per repeating unit depending on the type of macromolecule, and the effective energy function of the modeled system.⁶⁸ The transferability of the model to different systems is ensured by the decomposition of the potential of mean force of a system into a sum of contributions from its parts, corresponding to interactions within the CG sites, pairs of CG sites, as well as groups of CG sites. The UNICORN model has been successful in the prediction of protein structures,⁶⁹ folding kinetics,⁷⁰ conformational changes,⁷¹ and RNA and DNA structure and dynamics.⁷²

In this study, we present the implementation and calibration of the SUGRES-1P CG model.^{67,73} We based our implementation on the theoretical background and initial parametrization of the sugar rings with each other published by our group.^{67,73} Furthermore, we have fine tune the weights and parameters of the force field and applied the SUGRES-1P force field to the simulation of free HP of degree of polymerization (dp) ranging from 6 to 68. The conformations obtained from CG simulations were compared to experimentally determined HP structures.^{74,75} The results show a good agreement with the experimental data in terms of the general shape and conformation of the CG HP molecules as well as their global characteristics: end-to-end distance (EED) and radius of gyration (R_g).

2. METHODOLOGY

2.1. SUGRES-1P CG Force Field. The SUGRES-1P model^{67,73} employed in this work is a physics-based model of polysaccharide chains which is a part of the CG UNICORN model of biomolecules, alongside the UNited RESidue (UNRES), and united Nucleic Acid RESidue (NARES-2P) models for polypeptides and nucleic acids, respectively.⁶⁸ In the SUGRES-1P model, the polysaccharide chain representation is reduced to a single interaction site per sugar residue located halfway between glycosidic linkage oxygen atoms, which serve as anchor points of the polysaccharide chain (Figure 1). The geometry of the polysaccharide chain in the SUGRES-1P model is defined by virtual bonds (corresponding to O1 $\rightarrow$ O4 glycosidic linkages), the virtual bond angles θ_i , and virtual bond dihedral angles γ_i , as shown in Figure 1.

Each interaction center of the polysaccharide consists of two sections, termed the “head” and “tail”. This is brought by the necessity of accounting for the anisotropy of the CG HP residues, as the center of the charge is off the geometrical center of the residue. Therefore, the location of the head corresponds to the center of the charge, while that of the tail corresponds to the uncharged part of the residue. As a result, the energy of interactions between HP residues is calculated as a sum of the head–head, tail–tail, and head–tail interaction energies. The

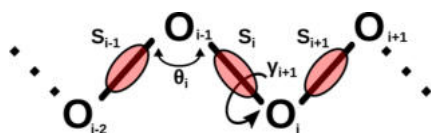


Figure 1. Illustration of the SUGRES-1P model. The interaction sites are united sugar rings, represented by transparent red ellipsoids, located half-way between glycosidic oxygen atoms (white spheres) which are not interaction sites but serve to define the geometry of the polysaccharide molecule. The virtual bonds connecting the oxygen atoms are shown as thick black lines. The geometry of the polysaccharide chain is defined by the virtual bond angles θ_i and torsional angles γ_i .

SUGRES-1P physics-based effective energy function of a polysaccharide chain is expressed by eq 1

$$U = w_{SS}^{vdW} \sum_i \sum_{i < j} U_{SS_j}^{vdW} + w_{SSf_2}^{el}(T) \sum_i \sum_{j < i} U_{SS_j}^{el} + w_{bond} \sum_i U_{bond}(d_i) + w_b \sum_i U_b(\theta_i) + w_{tor} \sum_i f_2(T) U_{tor}(\gamma_i, \theta_{i-1}, \theta_i) \quad (1)$$

The term $U_{bond}(d_i)$ accounts for the energetics of the virtual-bond-deformation, where d_i corresponds to the length of the i -th virtual bond, $U_b(\theta_i)$ is the virtual bond angle deformation term for angle θ_i , and $U_{tor}(\gamma_i, \theta_{i-1}, \theta_i)$ is the virtual-bond-torsional energy term for virtual bond dihedral angle γ_i , where angles (θ_{i-1} and θ_i) are the adjacent virtual bond angles.

The term $U_{SS_j}^{vdW}$ is represented by a sum of energy terms of the interaction of the polar and charged parts of the residue, excluding the Coulombic charge–charge interactions, as expressed by eq 2

$$U_{SS_j}^{vdW} = E_{GBerne} + E_{pol}^{GB} + E_{pol} + \Delta F_{cav} + \Delta F_{cav}^{iso} + E_{LJ} \quad (2)$$

The interaction between the uncharged parts of the interaction sites is modeled using the Gay–Berne potential,⁷⁶ expressed by

$$E_{GBerne} = 4\epsilon_{ij} \left(\left(\frac{\sigma_{ij}^0}{r_{ij} - \sigma_{ij} + \sigma_{ij}^0} \right)^{12} - \left(\frac{\sigma_{ij}^0}{r_{ij} - \sigma_{ij} + \sigma_{ij}^0} \right)^6 \right) \quad (3)$$

where r_{ij} is the distance between the uncharged tails of the interacting particles, σ_{ij} is the distance corresponding to the zero value of E_{GBerne} for arbitrary orientation of the particles (σ_{ij}^0 is the distance corresponding to the zero value of E_{GBerne} for the side-to-side approach of the particles), and ϵ_{ij} is the van der Waals well depth.

The contribution to the energy arising from polarization of the solvent, E_{pol}^{GB} , by charged parts of the interaction sites is computed using the generalized Born model

$$E_{pol}^{GB} = 332q_i q_j \left(\frac{1}{\epsilon_{in}} - \frac{1}{\epsilon_{out}} \right) \frac{1}{f_{GB}(r_{ij}')} \quad (4)$$

where q_i , q_j are the charges of the interacting particles, ϵ_{in} is the effective dielectric constant of the “inside” of the interacting particles, ϵ_{out} is the effective dielectric constant of the solvent, r_{ij}' is the distance between the charged heads of the interacting

particles, and $f_{GB}(R)$ is the Generalized Born function expressed by

$$f_{GB}(R) = \sqrt{R^2 + a_i a_j \exp\left(-\frac{R^2}{4a_i a_j}\right)} \quad (5)$$

where a_i and a_j are Born radii and R corresponds to the distance between the given sites of the interacting particles (e.g., the distance between the charged heads, r_{ij}' , in the case of E_{pol}^{GB}).

E_{pol} is the polarization energy corresponding to the interactions between the charged and uncharged parts of the interaction sites of two sugar residues

$$E_{pol} = 332 \left(\frac{1}{\epsilon_{in}} - \frac{\exp(-\kappa_D f_{GB}(r_{ij}'))}{\epsilon_{out}} \right) \left(\frac{\alpha_1}{f_{GB}(r_{ji}'')} + \frac{\alpha_2}{f_{GB}(r_{ij}'')} \right) \quad (6)$$

where r_{ij}' is defined under eq 3, r_{ji}'' is the distance between the uncharged tail of particle i and the charged head of particle j ; r_{ij}'' is the distance between the head of particle i and the tail of particle j ; ϵ_{in} and ϵ_{out} are dielectric constants within the particles and in bulk, respectively; κ_D is the length of Debye screening due to the presence of counterions; and α_1 and α_2 are solvation parameters of the tails of the particles.

The cavity term of the isotropic charged heads is expressed by ΔF_{cav}^{iso}

$$\Delta F_{cav}^{iso} = \frac{\alpha_{ij}^{(1)iso} (x)^{(1/2)} + \alpha_{ij}^{(2)iso} x - \alpha_{ij}^{(3)iso}}{1 + \alpha_{ij}^{(4)iso} (x)^{12}} \quad (7)$$

with

$$x = \frac{r_{ij}'}{\sqrt{(\sigma_i^{iso})^2 + (\sigma_j^{iso})^2}} \quad (8)$$

In the above equations, r_{ij}' is the distance between two charged parts of the interaction site of particles i and j , and σ_i^{iso} and σ_j^{iso} are equivalent to the minimum distance between the center of charge of particle i or j , respectively.

The cavity term of the uncharged tails is calculated by

$$\Delta F_{cav} = \frac{\alpha_{ij}^{(1)} ((x\lambda)^{(1/2)}) + \alpha_{ij}^{(2)} x\lambda - \alpha_{ij}^{(3)}}{1 + \alpha_{ij}^{(4)} (x\lambda)^{12}} \quad (9)$$

where

$$x = \frac{r_{ij}}{\sqrt{\sigma_i^2 + \sigma_j^2}} \quad (10)$$

and

$$\lambda = \left(1 - \frac{\chi_{ij}^{(1)} \omega_{ij}^{(1)2} + \chi_{ij}^{(2)} \omega_{ij}^{(2)2} - 2\chi_{ij}^{(1)} \chi_{ij}^{(2)} \omega_{ij}^{(1)} \omega_{ij}^{(2)} \omega_{ij}^{(12)}}{1 - \chi_{ij}^{(1)} \chi_{ij}^{(2)} \omega_{ij}^{(12)2}} \right) \quad (11)$$

where $\chi_{ij}^{(1)}$ and $\chi_{ij}^{(2)}$ are anisotropies related to ΔF_{cav} , r_{ij} is the distance between the tails of the interacting particles, and σ_i and σ_j are calculated with the minimum distance between the centers of the interacting particles.

The isotropic Lennard-Jones potential is used to model the van der Waals interaction energy between two polar heads

$$E_{\text{LJ}} = 4\epsilon'_{ij} \left[\left(\frac{\sigma'_{ij}}{r'_{ij}} \right)^{12} - \left(\frac{\sigma'_{ij}}{r'_{ij}} \right)^6 \right] \quad (12)$$

where r'_{ij} is the distance between the polar heads, σ'_{ij} is the distance corresponding to the zero value of E_{LJ} , and ϵ'_{ij} is the van der Waals well depth.

The electrostatic interaction energy between the charged parts of the interaction sites is expressed by

$$E_{\text{el}} = 332 \frac{q_i q_j}{\epsilon_{\text{in}} r'_{ij}} \quad (13)$$

The energy terms are multiplied by the weights w as well as, in the case of $U_{\text{S,S}_j}^{\text{el}}$, temperature factors $f_n(T)$ that reflect the temperature dependence of the appropriate effective-energy terms, defined by

$$f_n(T) = \frac{\ln[\exp(1) + \exp(-1)]}{\ln\{\exp[(T/T_0)^{n-1}] + \exp[-(T/T_0)^{n-1}]\}} \quad (14)$$

where $T_0 = 300$ K.

For a more detailed description of the CG SUGRES-1P force field, the reader is referred to refs 67 and 73.

2.2. HP Structures. The NMR structure of HP from PDB: 1HPN²⁵ was used as a template for the construction of HP fragments used in CG simulations. HP molecules of length from 6 to 68 residues, which are referred to as dp6, dp8, dp10, dp12, dp14, dp16, dp18, dp24, dp30, dp32, dp36, dp48, and dp68, respectively, were constructed using the xLeap module of AMBER⁷⁷ by either elongating or shortening the template structure. The choice of HP molecule lengths was dictated by the availability of experimental data used as a reference to compare with the results obtained using the CG MD simulation.⁷⁴

As a reference, measures of EED and radius of gyration (R_g) from analytical centrifugation and synchrotron X-ray scattering experiments for HP dp6, dp12, dp18, dp24, dp30, and dp36⁷⁴ and HP dp32, dp48, and dp68⁷⁵ were used. The conformations of HP oligomers obtained from CG simulations were compared with the structures of HP dp18, dp24, dp30, and dp36 obtained by⁷⁴ using constrained scattering modeling based on the structure of HP in ref 25.

2.3. Estimation of Energy-Term Weights and the Debye–Hückel Screening Factor. Due to the high negative charge of HP,¹ the inclusion of an appropriate amount of counterions in simulations is necessary. In the SUGRES-1P model, this is possible by the adjustment of the Debye–Hückel screening factor κ^{-1} , which describes the electrostatic screening distance of charges in an electrolyte^{78,79} and is linked to the ionic strength of a solution, as shown in eq 15, for the case of monovalent ions

$$\kappa^{-1} = \sqrt{\frac{\epsilon_r \epsilon_0 k_B T}{2N_A e^2 I}} \quad (15)$$

where κ^{-1} is the Debye–Hückel screening length, ϵ_r is the dielectric constant of water, ϵ_0 is the vacuum dielectric permittivity, k_B is the Boltzmann constant, T is the absolute temperature, N_A is the Avogadro number, e is the elementary charge, and I is the ionic strength of the electrolyte.⁷⁹ The Debye–Hückel length κ^{-1} is part of the $U_{\text{S,S}_j}^{\text{vdW}}$ term in eq 6 in the SUGRES effective energy equation. The expression for $U_{\text{S,S}_j}^{\text{vdW}}$ has

been adapted from previous studies of the effective interactions of like-charged side chains in the UNRES force field.⁸⁰ While a direct comparison between values of κ^{-1} as used in SUGRES-1P and actual ionic strength of solutions cannot be made due to the implicit nature of the solvent in the UNICORN CG model, the value of κ can be expected to increase with the square of ionic strength, as can be deduced from eq 15.

The appropriate weights of the energy terms from eq 1, as well as the optimal value of the Debye–Hückel screening factor, were determined empirically using HP of three different lengths (HP dp12, HP dp24, and HP dp68). For each effective-energy term, the corresponding weight was increased from 1 to 10 with increments of 1. At the same time, for each examined set of energy term weights, the value of the κ parameter was increased from 0.0 to 1.0 with increments of 0.1, as detailed in Table S1, emulating the increase of salt concentration.

Canonical Langevin dynamics simulations of each HP molecule were conducted with the SUGRES-1P force field for all combinations of effective energy-term weights and κ values without any restraints imposed on the size of the molecule, starting from the extended conformation. The time step size stems from the energy and distance units conversion from kcal/mol and Å, respectively (molecular time unit). Previous publications have investigated the influence of the step length used in the MD algorithm of the UNRES force field,^{81,82} especially in the context of energy drift in the CG simulations. The time step equal to 4.89 fs was recommended as a safe value for the stability of the MD algorithm.⁸² To obtain approximately 1 μ s real time per simulation, the simulations in our work were therefore carried out for 2,000,000 steps per simulation. The RMSD of the HP molecules in reference to the starting structures has been inspected to ensure the convergence of the MD simulations. The EED and radii of gyration (R_g) of the obtained HP conformations were compared to experimental data.^{74,75}

All CG simulations have been conducted at $T = 300$ K. Due to the dependence of the energy function employed by the SUGRES-1P force field on the temperature, a change of the simulation temperature would likely affect the obtained results and the choice of optimal parameters. Nevertheless, all simulations were conducted at $T = 300$ K as it represents a temperature close to the physiological temperature and corresponds to the experimental conditions employed most commonly in experiments, including the ones we used as a reference for our data.⁷⁴

2.4. CG Simulations of Short and Long Free HP Molecules for Selected Weights of Effective Energy Terms and Selected Values of κ . Canonical Langevin CG dynamics were conducted for HP dp6 to dp68 using three combinations of κ and weights of the effective energy terms:

1. κ_2 and electrostatic interaction energy weight (w_{eel}) equal 7,
2. κ_7 and electrostatic interaction energy weight (w_{eel}) equal 7,
3. κ_7 and virtual bond-stretching energy weight (w_{bond}) equal 4.

The trajectories comprised 2,000,000 steps with a 4.89 fs step length. The EED and radii of gyration (R_g) of the obtained HP conformations were compared to experimental data.^{74,75} The overall conformations of CG HP dp18, dp24, dp30, and dp38 were compared to the conformation obtained by ref 74 using

HP dp12

End-to-end distance

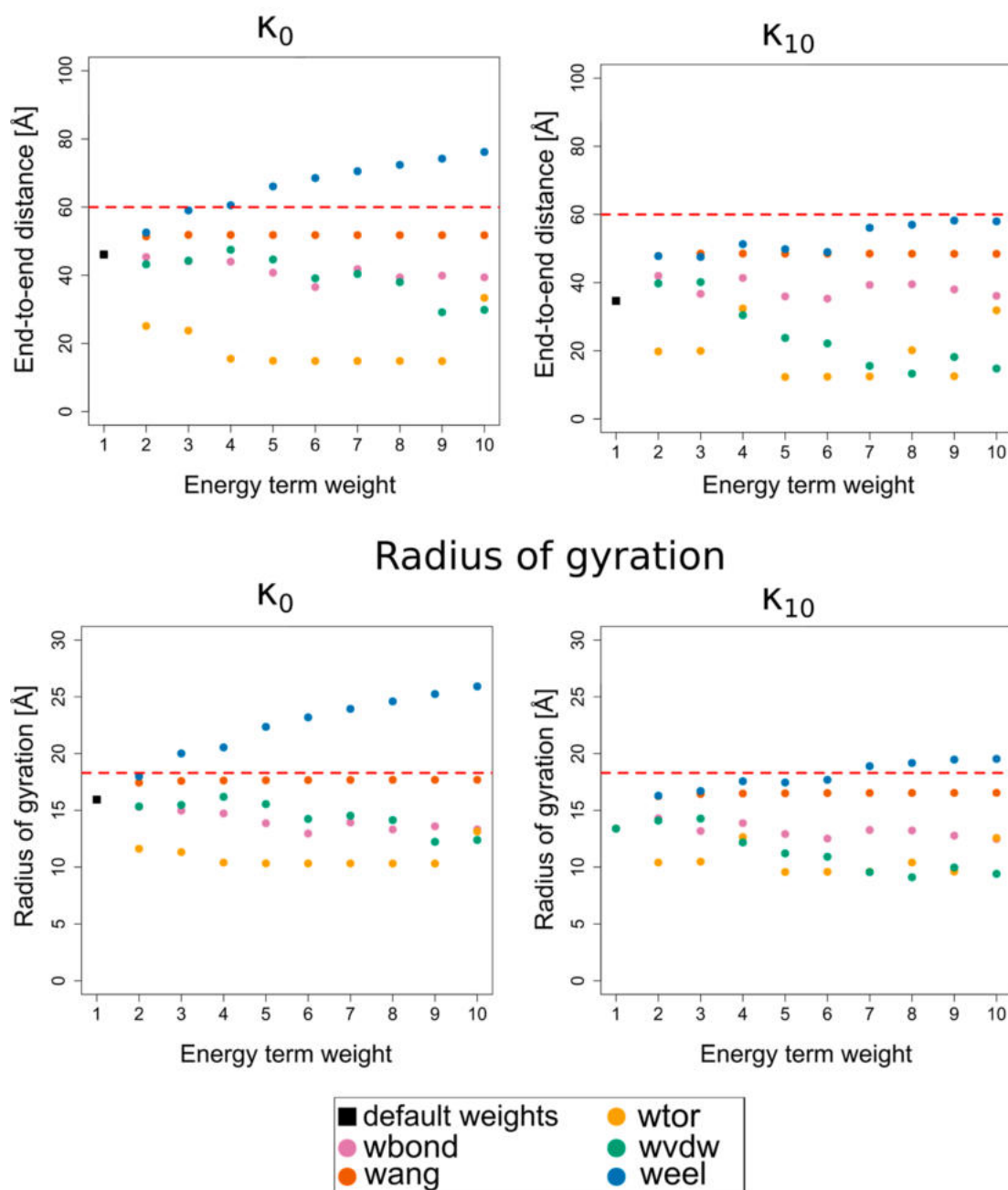


Figure 2. Dependence of the EED and the radius of gyration (R_g) on the energy term weight and ion concentration expressed by the κ parameter for HP dp12. The black square corresponds to all weights set to 1 (default) while colored circles (color legends shown below the graph) correspond to varying a given weight from 1 to 10 with a step of 1. See eq 1 and the text below this equation for the energy-term and energy-term-weight symbols. The experimental values of EED and R_g from refs 74 and 75 are shown as dashed red lines.

constrained scattering modeling based on the NMR structure of HP.²⁵

2.5. Visualization and Analysis. The CG conformations of HP were visualized using VMD⁸³ and analyzed using the cpptraj module of AMBER.⁷⁷ Comparisons of EED and R_g values as well as their visualization were performed using R.⁸⁴ Clustering of frames of the CG trajectory has been performed using the DBSCAN⁸⁵ algorithm using the cpptraj module of AMBER⁷⁷ with a minimum cluster size of 2 and distance cutoff of 4 Å.

3. RESULTS AND DISCUSSION

3.1. Estimation of Energy Term Weights and the Debye–Hückel Screening Factor. An initial CG simulation of free HP using default effective-energy term weights and parameters determined by refs 67 and 73 did not yield satisfactory results. The HP molecules formed tight coils, which are not likely to correspond to in vivo conformations.

The optimal weights of the effective energy terms from eq 1 as well as the optimal value of the κ parameter accounting for

HP dp24

End-to-end distance

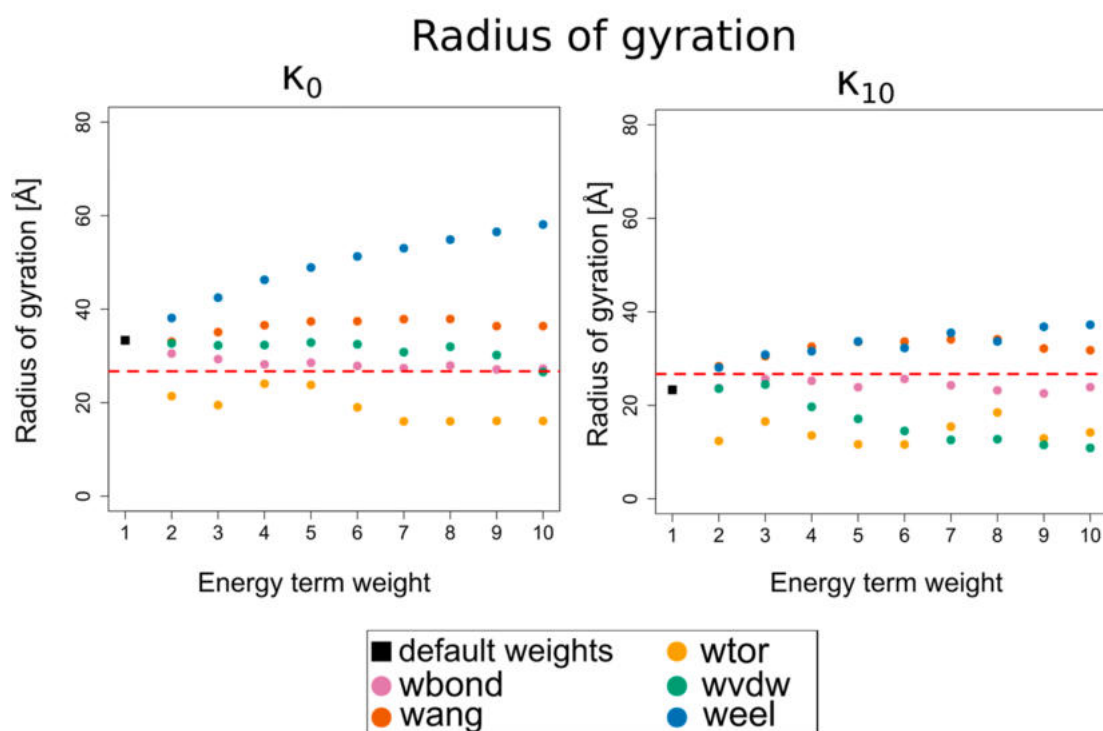
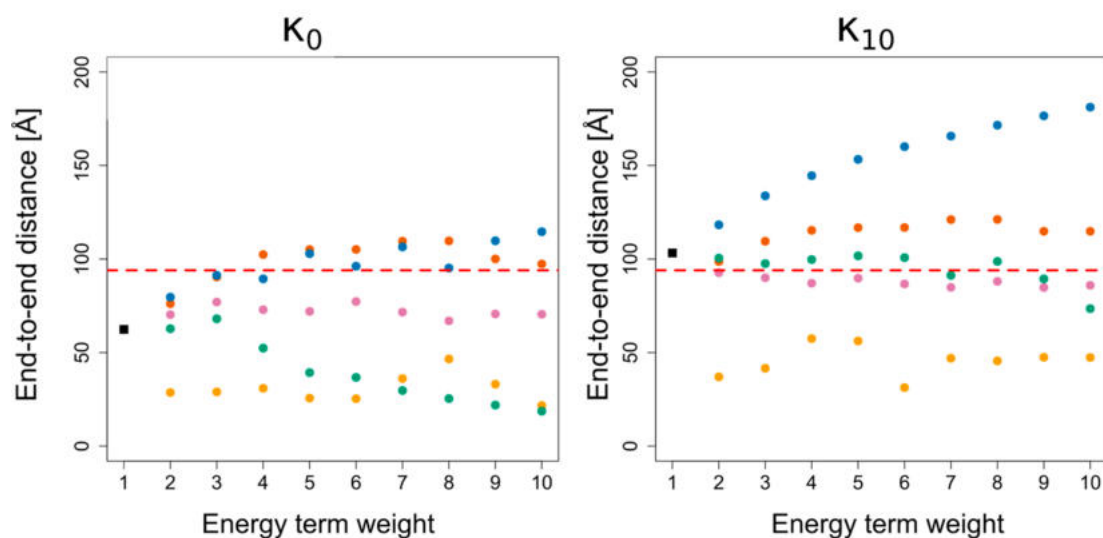


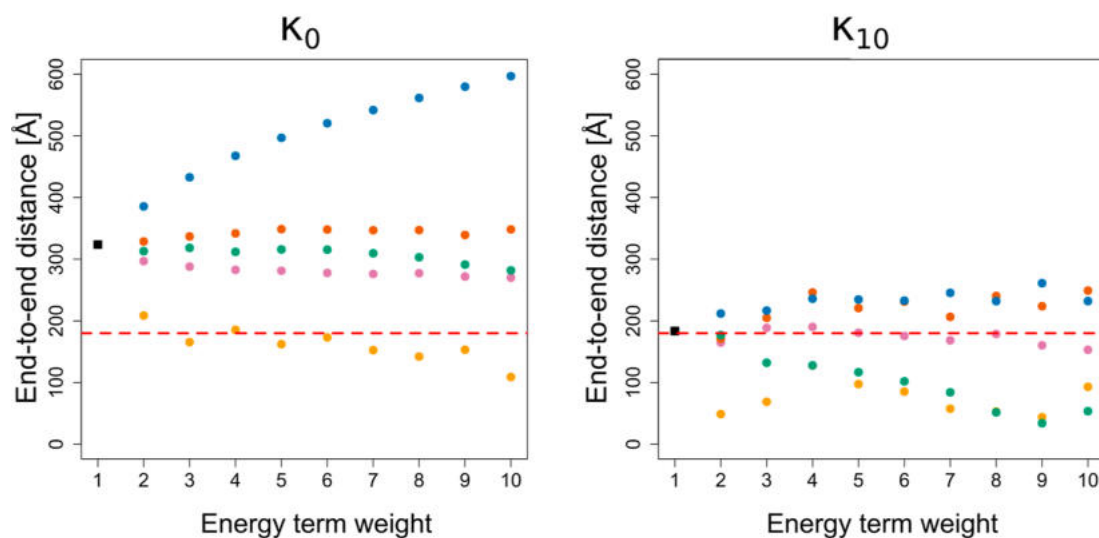
Figure 3. Dependence of the EED and the radius of gyration (R_g) on the energy term weight and ion concentration expressed by the κ parameter for HP dp24. The black square corresponds to all weights set to 1 (default) while colored circles (color legends shown below the graph) correspond to varying a given weight from 1 to 10 with a step of 1. See eq 1 and the text below this equation for the energy-term and energy-term-weight symbols. The experimental values of EED and R_g from refs 74 and 75 are shown as dashed red lines.

counterion screening have been determined empirically by conducting CG simulations of HP dp12, dp24, and dp68 for all combinations of energy term weights and κ as described in the Methodology section. In short, each energy-term weight was increased from 1 to 10 with a step of 1 while keeping all other energy-term weights at a default value of 1. For every considered value of the weights, the value of κ was increased from κ_0 , corresponding to a total lack of counterions in the simulated system, to κ_{10} , as detailed in Table S1. The average EED and R_g

values across the CG simulation trajectory were compared to experimental values from refs 74 and 75 in order to determine which combination of weights and κ resulted in conformations with characteristics most similar to the experimental HP conformations. Figures 2–4 summarize the findings for HP dp12, dp24, and dp68 for κ_0 and κ_{10} , which correspond to the two extremes of the tested values, i.e., a complete lack of ions and high salinity of the simulated system.

HP dp68

End-to-end distance



Radius of gyration

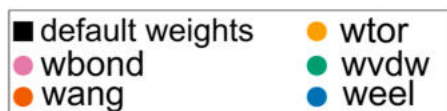
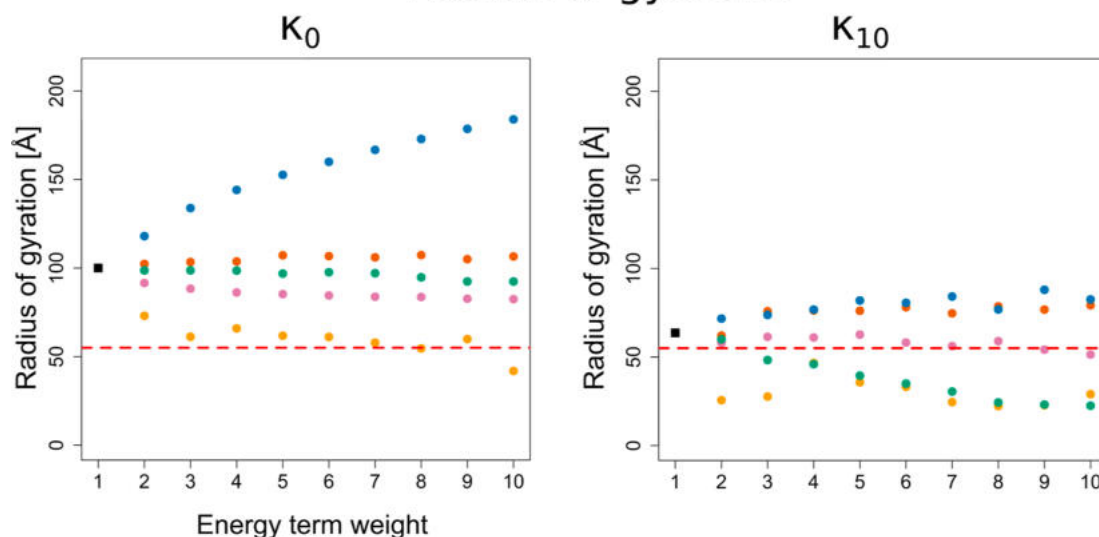


Figure 4. Dependence of the EED and the radius of gyration (R_g) on the energy term weight and ion concentration expressed by the κ parameter for HP dp68. The black square corresponds to all weights set to 1 (default) while colored circles (color legends shown below the graph) correspond to varying a given weight from 1 to 10 with a step of 1. See eq 1 and the text below this equation for the energy-term and energy-term-weight symbols. The experimental values of EED and R_g from refs 74 and 75 are shown as dashed red lines.

The increase of the κ parameter, corresponding to an increase in the ion concentration of the implicit solvent, led to a decrease in the EED and R_g of the three examined HP molecules. The same effect, i.e., the adoption of more coiled conformations, was achieved by increasing the weight of the torsional energy term. On the other hand, more elongated molecules were obtained by increasing the weight of the electrostatic energy term. Such conformations were also characterized by a larger radius of gyration.

Depending on the chain length of the analyzed HP molecules, the structures obtained from CG simulations were in good agreement with experimental data at different ion concentrations.^{74,75} Interestingly, the experimental EED of the short HP dp12 lied within the range of achievable values for simulations conducted in the absence of ions (i.e., for κ_0). The percentage of error (PE) of EED ranged from 75.3% below to 73.0% above the experimental value of 60 Å. In contrast, the EED of CG HP dp12 simulated in κ_{10} achieved values close to the experimental EED only in the case of very high weights of the

electrostatic energy term, with the closest value lying 2.9% below the experimental EED. While less pronounced than in the case of EED, the same can also be observed for the R_g of HP dp12. The experimental EED and R_g of HP dp24 were within the ranges of EED and R_g values of the simulated HP molecules for all the considered κ_i values. In general, the obtained HP dp68 conformations were characterized by an EED and R_g of the chain closer to the experimentally determined EED and R_g in higher ion concentrations compared to the shorter HP molecules. The long HP dp68 chains have a higher propensity for forming coiled conformations owing to their length; hence, the EED of the simulated chain is probably controlled less by the repulsive electrostatic forces and more by the increased flexibility of the virtual bonds and virtual bond angles caused by an increased weight of the bond stretching and torsional angle deformation energies.

The need for the adjustment of the energy-term weights is evident, as the use of default weights resulted in either too compact or too extended molecules compared to the experimentally determined conformations. At the same time, the different preferences for ion concentration and energy term weights for short and long HP chains highlight the challenge of finding one universal set of parameters and weights for all lengths of HP. The choice of optimal energy-term weights and κ_i employed in this study relied on an empirical approach that resulted in a set of parameters enabling the simulated HP molecules to adopt conformations similar to the ones observed in experiments (e.g., in ref 74). While this approach proved successful in this study, due to the interdependence of the parameters of the force field, more advanced optimization techniques may be applied in future developments of the SUGRES-1P force field. Recent research showed the successful application of a range of sophisticated optimization approaches, including a variety of machine-learning algorithms.^{86–94} While such complex parametrization techniques are without a doubt an interesting approach in possible future developments of the SUGRES-1P force field, for the aim of proof of concept of this study, we have restricted ourselves to the empirical approach. For this purpose, we have determined 10 combinations of energy term weights and κ_i for each of the three tested HP lengths that resulted in the best agreement with experimental EED and R_g values from refs 74 and 75 (detailed in Tables S2–S4).

Simulations at lower ion concentrations (κ_i ranging from κ_0 to κ_5) together with an increased weight of the electrostatic energy term (w_{eel} from 3 to 10) resulted in only small discrepancies between the experimental EED and the EED of CG HP dp12 molecules (Table S2). The percentage error of the EED of CG molecules ranged from 1.6% below to 1.9% above the experimental value, with the closest EED lying just 0.2% above the experimental EED in the case of κ_2 and w_{eel} equal 7. The R_g of the conformation obtained in simulations with these settings were, however, higher than the corresponding experimental R_g values (discrepancies within 12.6% of the experimental EED value). In order to reduce the difference between the CG R_g and the experimental R_g , simulations had to be conducted at low ion concentrations (κ_0 to κ_6) together with the electrostatic energy term weights between $w_{\text{eel}} = 2$ and $w_{\text{eel}} = 6$. While reducing the percentage error of R_g values to a maximum of 2.2%, this increased the discrepancies in the EED values (a percentage error of the values of up to 17.3%). Consequently, the HP dp12 molecules simulated in settings better fitted to replicate

experimental R_g values were shorter than the HP molecules studied by ref 74.

The 10 HP dp24 conformations with EED values closest to the experimental lengths were obtained from CG simulations in a larger range of κ_i values than in the case of HP dp12 (κ_i ranging from κ_0 to κ_{10}). The simulated molecules were characterized by EED lying within 1.5% of the experimentally determined value. When the EED values of the simulated molecules were closer to the experimental EED values, the R_g values were below the experimentally determined R_g values (percentage error up to 26.1% of the experimental value). In order to achieve R_g values closer to those determined by ref 74, the modification of either the virtual bond-stretching energy term weight (w_{bond} from 2 to 4), the weight of the van der Waals interaction energy (w_{vdw} equal either 4 or 10), or the use of default energy term weights in a low ion concentration (κ_2) was necessary. As a consequence, however, the HP molecules were shorter than the experimentally determined conformations (up to 21.8% shorter than the experimental EED). It is important to note that while the EED and R_g are closely related measures; they describe different aspects of the size of polymer chains. For instance, a coiled polymer chain may be characterized by a small EED, owing to the close location of its ends, but a larger R_g value due to the overall volume of the coiled chain. Consequently, an inconsistency in the increase rate of the EED and R_g values of the simulated molecules, depending on the changes of the effective-energy weights and κ_i , could be observed.

Although a wide range of κ_i values enabled HP dp68 to achieve EED close to the experimental values, a predominance of higher κ_i among the 10 best combinations of κ_i and energy term weight could be observed. The conformations obtained from the CG simulations lied within the close range of the experimental EED (the percentage error of the EED lied at and below 0.8% for the 10 best combinations of energy weights and κ). The modification of the virtual bond stretching energy term weight was the most prevalent modification among those resulting in EED values closer to the experimental EED. The HP dp68 conformation characterized by a R_g closest to the experimental R_g , as determined by ref 75, was obtained by conducting CG simulations without the presence of ions in the implicit solvent together with a modification of the torsional energy term weight to $w_{\text{tor}} = 8$ (percentage error 0.6%). This resulted in a more compact and coiled chain, yet at the same time flexible in the context of the rotation of the virtual bonds. The remaining combinations of parameters among those resulting in a better agreement of experimental and theoretical R_g (percentage error within 1.6%) comprised modifications of the virtual bond stretching energy weights (w_{bond} from 6 to 10) in a wide range of κ values (κ_0 to κ_{10}). As was the case for HP dp12 and dp24, this resulted in a larger difference in EED values between the experimentally studied and simulated HP dp68 molecules (percentage error up to 21.1%).

Assuming that HP dp12 is a representative of shorter chains of HP, HP dp68—representative of long HP chains, and HP dp24 of medium-length chains, several observations can be made from the conducted analyses. Compared to medium-length and long HP, HP dp12 exhibits a greater tendency to adopt extended conformations. Unsurprisingly, the high importance of the repulsive electrostatic interactions between the CG sugar residues seems to be the most important energy term for HP dp12, causing the extension of the HP chain. While HP dp24 achieved R_g and EED close to experimental values for all tested κ_i values, provided the appropriate modification of the energy term

weights, HP dp68 showed a slight preference for higher ion concentrations, corresponding to more compact conformations. The energy term most important to achieve conformations of HP dp68 resemble the experimental ones was the virtual bond stretching energy. Taken together with the importance of the high torsional energy term weight in obtaining conformations of R_g comparable to experimental data, this suggests a more dynamic and flexible behavior of the long HP chains compared to the shorter HP molecules.

At the same time, it is important to highlight the possible bias toward more coiled conformations of the molecules studied due to the origin of the used parameters. The initial parameters used in the effective energy terms were determined based on all-atom MD simulations of HP using the GLYCAM06 force field and the TIP3P explicit water model.⁷³ However, the limitations of different water models in all-atom simulations of GAGs have been demonstrated in some studies: TIP3P may not reproduce the charge distribution with the same quality as more complex solvent models and therefore result in highly bent conformations of HP and other GAGs.^{95,96} This is likely caused by the accumulation of counterions near the GAG molecules, which enabled the relatively close contacts between sugar residues. Therefore, the tendency of HP to adopt coiled and bent conformations when simulated in the CG SUGRES-1P force field may be influenced by the origin of the parameters of the force field used to obtain them.⁷³

The length, flexibility, and overall shape of HP chains are important to ensure the formation of correct conformations and allow binding to proteins. The differences between experimental values of both EED and R_g and those of the CG molecules became more apparent with increasing chain length. The R_g of the tested HP molecules appeared to be more robust than the EED, i.e., it did not deviate as strongly from the experimental values for molecules simulated using combinations of energy term weights and κ optimized for a better agreement of the EED. Therefore, the parameters for subsequent analyses were chosen among those combinations of energy term weights and κ , which resulted in HP conformations characterized by EED close to the experimental chain lengths. As a result, three combinations of κ_i and energy term weights were chosen, each corresponding to the optimal parameters for HP dp12, dp24, and dp68: κ_2 and $w_{\text{eel}} = 7$, κ_7 and $w_{\text{eel}} = 7$, and finally κ_7 together with $w_{\text{bond}} = 4$.

3.2. CG Simulations of Short and Long Free HP Molecules for Selected Weights of Effective Energy Terms and Selected Values of κ . Simulations of HP of length dp6 to dp68 were conducted in the SUGRES-1P force field using the three combinations of κ_i and energy term weights mentioned in the previous subsection. The EED of the obtained HP molecules was plotted as a function of the degree of polymerization of HP chains (Figure 5). As could be expected, the parameter set, which previously was optimal for HP dp12, also proved to be better in the simulation of shorter HP chains. The difference in the experimental EED and the EED of CG HP was the smallest for this set of parameters compared to the other parameter sets for HP dp6 and HP dp12 (percentage errors of 20.4 and 0.8% for HP dp6 and dp12, respectively; Table S1). The conformations of medium-long HP molecules (dp18 and dp24) had EED values closest to the experimental ones with parameters optimal for HP dp24 (percentage error of 0.8% for HP dp18 and 0.4% for HP dp24; Table S5). The EED of HP dp30 was estimated with roughly the same percentage error by both the HP dp24-fitted and HP dp68-fitted parameter combinations, with an error of 16.5% for κ_7 with $w_{\text{eel}} = 7$ and

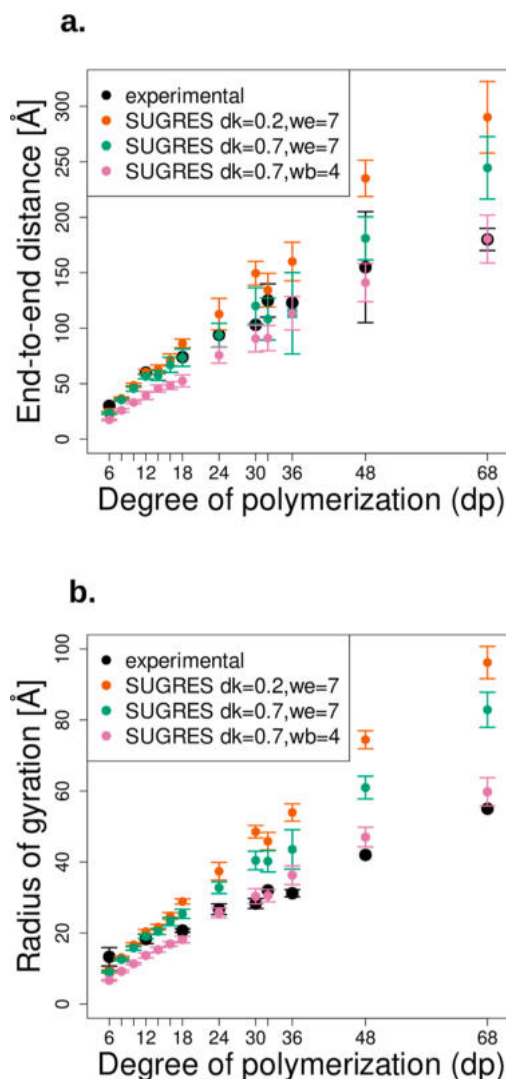


Figure 5. Dependence of the end-to-end distance (EED) (a) and the radius of gyration (R_g) (b) on the effective-energy term weight and ion concentration expressed by the κ parameter for HP dp from 6 to 68. Black points correspond to the experimental values from refs 74 and 75; standard deviation of the experimental values are represented by black bars where available. The EED and R_g values obtained by the modification of the weights and κ are shown as colored points, with a standard deviation represented by bars of corresponding color.

an error of 12.0% for κ_7 and $w_{\text{bond}} = 4$. This was also the case for HP dp36, where the percentage error between the experimental and theoretical EED was equal to 7.8% when using either of the aforementioned parameter combinations. Interestingly, the best parameter combination in terms of agreement of experimental and theoretical EED of HP dp32 was the HP dp12-fitted combination of κ_2 and $w_{\text{eel}} = 7$. Both HP dp48 and HP dp68 exhibited a better agreement with the experimentally determined EED when simulated using parameters estimated from tests using HP dp68 (percentage errors of 9.0 and 0.1% for HP dp48 and dp68, respectively).

The R_g of the short HP dp6 was underestimated in all three tested combinations of κ_i and weights (Table S6), with a minimum of 30.5% error in the value of R_g in the case of HP dp12-fitted parameters. In the case of HP dp12, the best parameter combination was κ_7 with $w_{\text{eel}} = 7$, resulting in only 3.9% of error above the experimentally determined value of R_g . For all other HP lengths (HP dp18 to dp68), the best agreement with experimental R_g was obtained when simulating the

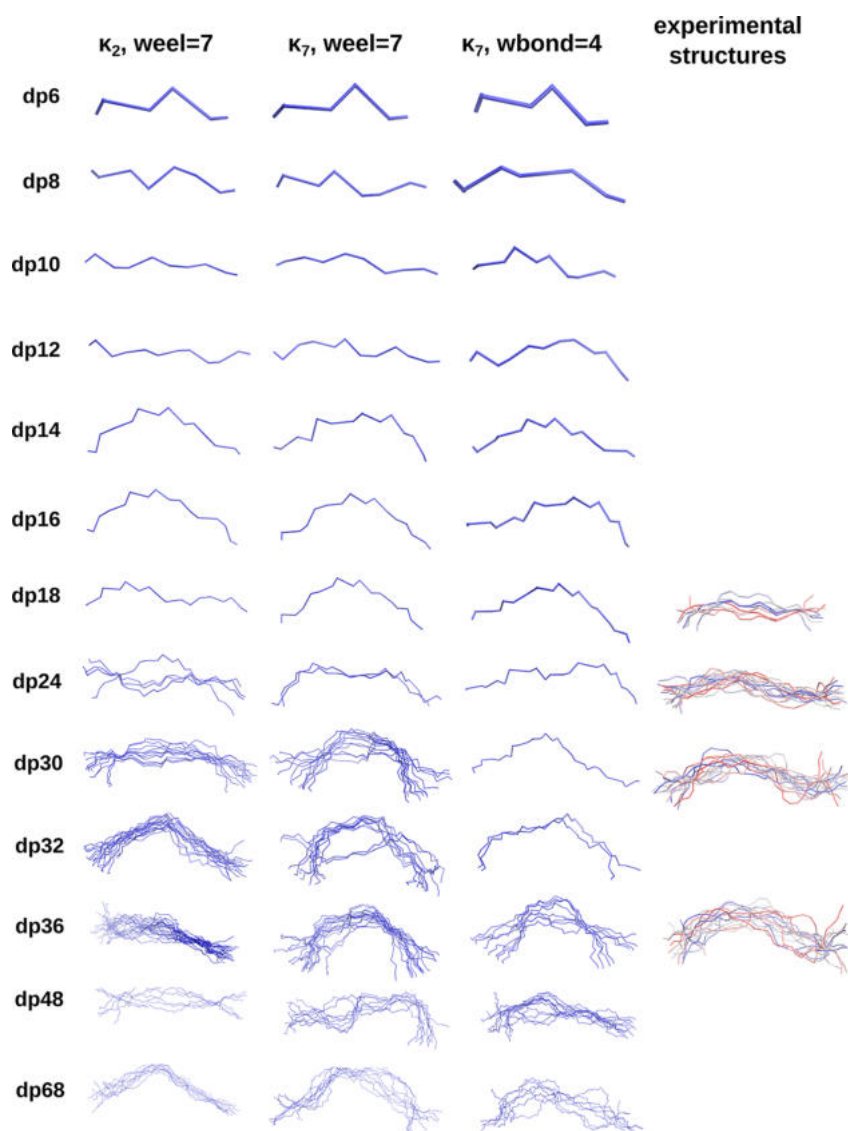


Figure 6. Representative conformations of the CG trajectories of HP conducted using κ_2 with $w_{\text{eel}} = 7$, κ_7 with $w_{\text{eel}} = 7$, and κ_7 and $w_{\text{bond}} = 4$, visualized in trace representations, colored in blue. Additionally, the CG SUGRES-1P representation of the experimentally determined conformations of HP⁷⁴ of HP dp18 (PDB ID: 1IRI), dp24 (PDB ID: 1IRJ), dp30 (PDB ID: 1IRK), and dp36 (PDB ID: 1IRL) in trace representation.

molecules using κ_7 and $w_{\text{bond}} = 4$. For HP dp18, dp24, and dp32, the optimal parameter combination underestimated the R_g of the simulated chains relative to the experimental R_g (percentage errors of 12.3, 3.8, and 7.2% for dp18, dp24, and dp32, respectively). In simulations of HP dp30, dp36, dp48, and dp68, the R_g of the obtained conformations lied above the experimentally determined values (percentage errors of 7.2% for HP dp30, 16.3% for HP dp36, 12.0% for HP dp48, and 8.7% for HP dp68). It seems, therefore, that the conformations of HP obtained by simulations using the CG SUGRES-1P force field would be able to reproduce either the degree of extension of the chains, as expressed by the EED, or their radius of gyration, but not both at the same time. In general, the HP dp12-fitted parameters caused the HP molecules to adopt more extended chains, characterized by a higher R_g . The shorter, more coiled chains obtained by conducting the CG simulations with the dp68-fitted parameters were a good choice for the longer chains (from HP dp18 to dp68). The parameters that seemed like a good “middle ground” when considering the agreement with experimental EED resulted in a small difference in the R_g values of experimental R_g and the R_g of the simulated molecules only for HP dp12.

The structures obtained by the CG MD simulations were clustered using the DBSCAN algorithm to identify representative conformations for particular κ_i and energy term weights, as visualized in Figure 6. The simulated HP molecules adapted coiled conformations independently of the parameter combination used in the CG simulations. In most of the simulations, only one conformation cluster could be identified due to the similarity of the sampled conformations. Importantly, in all three simulation settings and for all of the identified cluster representatives, a kink in the chain appeared for HP longer than dp14, corresponding to the kink observed experimentally, e.g., in ref 74. However, using κ_2 and $w_{\text{eel}} = 7$, fitted for shorter HP chains, the kink became less pronounced for longer HP molecules. At the same time, the overall shape of the other HP molecules did not seem qualitatively too different between the parameters fitted for medium-long and those for long HP chains (κ_7 and $w_{\text{eel}} = 7$). The representative conformations have been compared with the experimentally determined conformations of HP deposited in the RCSB PDB under the IDs: 1IRI, 1IRJ, 1IRK, and 1IRL,⁷⁴ visualized in the SUGRES-1P CG representation, as shown in Figure 6. A striking similarity in the overall shape of the chains of CG HP can be observed—

while the simulated chains appear to be slightly more extended than their experimental counterparts, the key feature of the chains, i.e., the kink in the structure, is well preserved in the CG molecules.

Table S7 details the root-mean-square deviation (RMSD) calculated for O4 atoms of the CG HP structures in reference to the first model in the conformations of HP dp18, dp24, dp30, and dp36 in PDB entries 3IRI, 3IRJ, 3IRK, and 3IRL.⁷⁴ The variety of structures obtained from CG simulations lies within the range of the experimentally determined diversity of the dynamic HP molecules. In the case of HP dp18, the RMSD between the ensemble of structures in PDB ID 3IRI⁷⁴ reached up to 7.6 Å, which is only 0.3 Å less than the mean RMSD between all of the frames of the CG simulation of HP in reference to the models in 3IRI. This difference in RMSD values increases with the increasing length of the HP chain; however, it remains relatively small. The RMSD obtained during CG simulations in reference to the models in PDB entry 3IRJ is 10.3 Å (Table S7), which is 2.7 Å larger than the RMSD between the experimentally determined conformations in PDB 3IRJ.⁷⁴ In the case of HP dp30 and dp36, the RMSD between the structure models in the respective PDB entries ranged up to 10.1 and 14.5 Å, while the mean RMSD of the HP structures obtained by CG simulations calculated in reference to the models in the respective PDB entries lies at 15.2 and 18.6 Å, respectively. This shows that the CG simulations of the HP chains in the SUGRES-1P force field were able to capture the dynamic behavior of the HP molecules approximately within the ranges of the experimental diversity of HP conformations.⁷⁴ It is important to note that the results of the CG simulation in the SUGRES-1P force field are based purely on theoretical approaches and no constraints based on experimentally determined EED were imposed on the simulated HP chains during the CG simulations. Taking all comparisons of the EED, R_g , and overall shape of the simulated chains into account, κ_7 together with $w_{\text{eel}} = 7$ is likely a good combination of parameters that can be used for the simulation of both short and long HP chains.

4. CONCLUSIONS

We have extended the SUGRES-1P CG model to the simulation of free HP chains. The effective interaction energy function has been modified compared to ref 73 by extracting the electrostatic interaction energy term from the sum containing all other interaction energies of two sugar residues to enable a direct modification of the corresponding electrostatic energy term weight. This enabled us to obtain more extended HP chain conformations characterized by remarkable similarity to the experimentally determined HP molecules. The estimation of the energy term weights and the κ_i parameter, describing the ion concentration of the simulated system, enabled us to identify the most important elements of the effective energy function specific to HP chains of different lengths. The results suggest that long HP molecules most likely adopt more coiled conformations, governed predominantly by the electrostatic interaction energy of the charged HP residues. The SUGRES-1P module is fully compatible with the UNICORN model of other biomacromolecule types, consequently providing the potential for simulating and analyzing their interactions with HP. The next step in our work is to thoroughly explore the interaction between HP and particular proteins, providing a significant approach to model large protein systems that contain GAGs and will prove useful in the simulation of the corresponding key biological phenomena,

such as collagen reorganization,^{97,98} the maintenance of protein gradients in the presence of GAGs,^{99–102} and GAG-induced amyloidogenesis.^{103,104} The inclusion of parameters for other types of GAGs in the CG SUGRES-1P model will be addressed in future works.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jctc.3c00511>.

Debye–Hückel screening factor κ , EED, R_g , and RMSD (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Adam K. Sieradzan – Faculty of Chemistry, University of Gdansk, 80-308 Gdańsk, Poland; orcid.org/0000-0002-2426-3644; Email: adam.sieradzan@ug.edu.pl

Authors

Annemarie Danielsson – Faculty of Chemistry, University of Gdansk, 80-308 Gdańsk, Poland; orcid.org/0000-0002-6151-4449

Sergey A. Samsonov – Faculty of Chemistry, University of Gdansk, 80-308 Gdańsk, Poland; orcid.org/0000-0002-5166-4849

Adam Liwo – Faculty of Chemistry, University of Gdansk, 80-308 Gdańsk, Poland; orcid.org/0000-0001-6942-2226

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acs.jctc.3c00511>

Notes

The authors declare no competing financial interest.

The SUGRES-1P package is available free of charge at: <https://unres.pl/downloads> together with example input files.

■ ACKNOWLEDGMENTS

This work was supported by the Polish National Science Centre (Narodowe Centrum Nauki), grants UMO-2018/31/G/ST4/00246 (to A.D. and S.A.S.), UMO-2021/40/Q/ST4/00035 (to A.L.), and UMO-2017/27/B/ST4/00926 (to A.K.S.) as well as the University of Gdansk [Research of Young Scientists grant (BMN) no. 539-T110-B984-23] (to A.D.). Calculations were conducted by using the resources of the cluster at the Faculty of Chemistry, University of Gdańsk.

■ REFERENCES

- (1) Esko, J. D.; Kimata, K.; Lindahl, U. *Proteoglycans and Sulfated Glycosaminoglycans*; Cold Spring Harbor Laboratory Press, 2009.
- (2) Rabenstein, D. L. Heparin and heparan sulfate: structure and function. *Nat. Prod. Rep.* **2002**, *19*, 312–331.
- (3) Folkman, J.; Taylor, S.; Spillberg, C. The role of heparin in angiogenesis. *Ciba Found. Symp.* **1983**, *100*, 132–149.
- (4) Folkman, J.; Shing, Y. Control of angiogenesis by heparin and other sulfated polysaccharides. *Adv. Exp. Med. Biol.* **1992**, *313*, 355–364.
- (5) Onishi, A.; St Ange, K.; Dordick, J. S.; Linhardt, R. J. Heparin and anticoagulation. *Front. Biosci.* **2016**, *21*, 1372–1392.
- (6) Khorana, A. A.; Sahni, A.; Altland, O. D.; Francis, C. W. Heparin inhibition of endothelial cell proliferation and organization is dependent on molecular weight. *Arterioscler., Thromb., Vasc. Biol.* **2003**, *23*, 2110–2115.

- (7) Lever, R.; Hoult, J. R.; Page, C. P. The effects of heparin and related molecules upon the adhesion of human polymorphonuclear leucocytes to vascular endothelium in vitro. *Br. J. Pharmacol.* **2000**, *129*, 533–540.
- (8) Bendas, G.; Borsig, L. Heparanase in Cancer Metastasis - Heparin as a Potential Inhibitor of Cell Adhesion Molecules. *Adv. Exp. Med. Biol.* **2020**, *1221*, 309–329.
- (9) Azizkhan, R. G.; Azizkhan, J. C.; Zetter, B. R.; Folkman, J. Mast cell heparin stimulates migration of capillary endothelial cells in vitro. *J. Exp. Med.* **1980**, *152*, 931–944.
- (10) Majack, R. A.; Clowes, A. W. Inhibition of vascular smooth muscle cell migration by heparin-like glycosaminoglycans. *J. Cell. Physiol.* **1984**, *118*, 253–256.
- (11) Zak-Nejmark, T.; Krasnowska, M.; Jankowska, R.; Jutel, M. Heparin modulates migration of human peripheral blood mononuclear cells and neutrophils. *Arch. Immunol. Ther. Exp.* **1999**, *47*, 245–249.
- (12) Falanga, A.; Marchetti, M. Heparin in tumor progression and metastatic dissemination. *Semin. Thromb. Hemostasis* **2007**, *33*, 688–694.
- (13) Chatzinikolaou, G.; Nikitovic, D.; Berdiaki, A.; Zafiropoulos, A.; Katonis, P.; Karamanos, N. K.; Tzanakakis, G. N. Heparin regulates colon cancer cell growth through p38 mitogen-activated protein kinase signalling. *Cell Proliferation* **2010**, *43*, 9–18.
- (14) Smorenburg, S. M.; Van Noorden, C. J. The complex effects of heparins on cancer progression and metastasis in experimental studies. *Pharmacol. Rev.* **2001**, *53*, 93–105.
- (15) Cohlberg, J. A.; Li, J.; Uversky, V. N.; Fink, A. L. Heparin and other glycosaminoglycans stimulate the formation of amyloid fibrils from alpha-synuclein in vitro. *Biochemistry* **2002**, *41*, 1502–1511.
- (16) Fichou, Y.; Vigers, M.; Goring, A. K.; Eschmann, N. A.; Han, S. Heparin-induced tau filaments are structurally heterogeneous and differ from Alzheimer's disease filaments. *Chem. Commun.* **2018**, *54*, 4573–4576.
- (17) Zhu, H.-L.; Fernández, C.; Fan, J.-B.; Shewmaker, F.; Chen, J.; Minton, A. P.; Liang, Y. Quantitative characterization of heparin binding to Tau protein: implication for inducer-mediated Tau filament formation. *J. Biol. Chem.* **2010**, *285*, 3592–3599.
- (18) Pérez, M.; Valpuesta, J. M.; Medina, M.; Montejo de Garcini, E.; Avila, J. Polymerization of τ into Filaments in the Presence of Heparin: The Minimal Sequence Required for τ - τ Interaction. *J. Neurochem.* **1996**, *67*, 1183–1190.
- (19) Vieira, T. C. R. G.; Cordeiro, Y.; Caughey, B.; Silva, J. L. Heparin binding confers prion stability and impairs its aggregation. *FASEB J.* **2014**, *28*, 2667–2676.
- (20) Imamura, M.; Tabeta, N.; Kato, N.; Matsuura, Y.; Iwamaru, Y.; Yokoyama, T.; Murayama, Y. Heparan Sulfate and Heparin Promote Faithful Prion Replication in Vitro by Binding to Normal and Abnormal Prion Proteins in Protein Misfolding Cyclic Amplification. *J. Biol. Chem.* **2016**, *291*, 26478–26486.
- (21) Iwahashi, N.; Ikezaki, M.; Nishikawa, T.; Namba, N.; Ohgita, T.; Saito, H.; Ihara, Y.; Shimanouchi, T.; Ino, K.; Uchimura, K.; Nishitsuji, K. Sulfated glycosaminoglycans mediate prion-like behavior of p53 aggregates. *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *117*, 33225–33234.
- (22) Simonaro, C. M.; D'Angelo, M.; He, X.; Eliyahu, E.; Shtraizent, N.; Haskins, M. E.; Schuchman, E. H. Mechanism of glycosaminoglycan-mediated bone and joint disease: implications for the mucopolysaccharidoses and other connective tissue diseases. *Am. J. Pathol.* **2008**, *172*, 112–122.
- (23) Greinacher, A.; Selleng, K.; Warkentin, T. E. Autoimmune heparin-induced thrombocytopenia. *J. Thromb. Haemostasis* **2017**, *15*, 2099–2114.
- (24) Hao, C.; Xu, H.; Yu, L.; Zhang, L. Heparin: An essential drug for modern medicine. *Prog. Mol. Biol. Transl. Sci.* **2019**, *163*, 1–19.
- (25) Mulloy, B.; Forster, M. J.; Jones, C.; Davies, D. B. N.m.r. and molecular-modelling studies of the solution conformation of heparin. *Biochem. J.* **1993**, *293*, 849–858.
- (26) García-Mayoral, M. F.; Canales, Á.; Díaz, D.; López-Prados, J.; Moussaoui, M.; de Paz, J. L.; Angulo, J.; Nieto, P. M.; Jiménez-Barbero, J.; Boix, E.; Bruix, M. Insights into the Glycosaminoglycan-Mediated Cytotoxic Mechanism of Eosinophil Cationic Protein Revealed by NMR. *ACS Chem. Biol.* **2013**, *8*, 144–151.
- (27) Sanderson, P. N.; Huckerby, T. N.; Nieduszynski, I. A. Conformational equilibria of alpha-L-iduronate residues in disaccharides derived from heparin. *Biochem. J.* **1987**, *243*, 175–181.
- (28) Künze, G.; Köhling, S.; Vogel, A.; Rademann, J.; Huster, D. Identification of the Glycosaminoglycan Binding Site of Interleukin-10 by NMR Spectroscopy. *J. Biol. Chem.* **2016**, *291*, 3100–3113.
- (29) Wu, L.; Viola, C. M.; Brzozowski, A. M.; Davies, G. J. Structural characterization of human heparanase reveals insights into substrate recognition. *Nat. Struct. Mol. Biol.* **2015**, *22*, 1016–1022.
- (30) Shaya, D.; Tocilj, A.; Li, Y.; Myette, J.; Venkataraman, G.; Sasisekharan, R.; Cygler, M. Crystal structure of heparinase II from *Pedobacter heparinus* and its complex with a disaccharide product. *J. Biol. Chem.* **2006**, *281*, 15525–15535.
- (31) Capila, I.; Hernáiz, M. J.; Mo, Y. D.; Mealy, T. R.; Campos, B.; Dedman, J. R.; Linhardt, R. J.; Seaton, B. A. Annexin V–Heparin Oligosaccharide Complex Suggests Heparan Sulfate–Mediated Assembly on Cell Surfaces. *Structure* **2001**, *9*, 57–64.
- (32) Shao, C.; Zhang, F.; Kemp, M. M.; Linhardt, R. J.; Waisman, D. M.; Head, J. F.; Seaton, B. A. Crystallographic analysis of calcium-dependent heparin binding to annexin A2. *J. Biol. Chem.* **2006**, *281*, 31689–31695.
- (33) Munakata, H.; Takagaki, K.; Majima, M.; Endo, M. Interaction between collagens and glycosaminoglycans investigated using a surface plasmon resonance biosensor. *Glycobiology* **1999**, *9*, 1023–1027.
- (34) Théoleyre, S.; Kwan Tat, S.; Vusio, P.; Blanchard, F.; Gallagher, J.; Ricard-Blum, S.; Fortun, Y.; Padrines, M.; Rédini, F.; Heymann, D. Characterization of osteoprotegerin binding to glycosaminoglycans by surface plasmon resonance: Role in the interactions with receptor activator of nuclear factor κ B ligand (RANKL) and RANK. *Biochem. Biophys. Res. Commun.* **2006**, *347*, 460–467.
- (35) Rusnati, M.; Bugatti, A. Surface Plasmon Resonance Analysis of Heparin-Binding Angiogenic Growth Factors. *Methods Mol. Biol.* **2016**, *1464*, 73–84.
- (36) Yu, Y.; Zhang, F.; Renois-Predelus, G.; Amster, I. J.; Linhardt, R. J. Filter-entrapped enrichment pull-down assay for glycosaminoglycan structural characterization and protein interaction. *Carbohydr. Polym.* **2020**, *245*, 116623.
- (37) Przybylski, C.; Gonnet, F.; Saesen, E.; Lortat-Jacob, H.; Daniel, R. Surface plasmon resonance imaging coupled to on-chip mass spectrometry: a new tool to probe protein-GAG interactions. *Anal. Bioanal. Chem.* **2020**, *412*, 507–519.
- (38) Liu, X.; Sun, C.; Zang, H.; Wang, W.; Guo, R.; Wang, F. Capillary electrophoresis for simultaneous analysis of heparin, chondroitin sulfate and hyaluronic acid and its application in preparations and synovial fluid. *J. Chromatogr. Sci.* **2012**, *50*, 373–379.
- (39) Loegel, T. N.; Trombley, J. D.; Taylor, R. T.; Danielson, N. D. Capillary electrophoresis of heparin and other glycosaminoglycans using a polyamine running electrolyte. *Anal. Chim. Acta* **2012**, *753*, 90–96.
- (40) Eldridge, S. L.; Higgins, L. A.; Dickey, B. J.; Larive, C. K. Insights into the capillary electrophoresis separation of heparin disaccharides from nuclear magnetic resonance, pKa, and electrophoretic mobility measurements. *Anal. Chem.* **2009**, *81*, 7406–7415.
- (41) Lin, L.; Liu, X.; Zhang, F.; Chi, L.; Amster, I. J.; Leach, F. E.; Xia, Q.; Linhardt, R. J. Analysis of heparin oligosaccharides by capillary electrophoresis-negative-ion electrospray ionization mass spectrometry. *Anal. Bioanal. Chem.* **2017**, *409*, 411–420.
- (42) Samsonov, S. A.; Pisabarro, M. T. Computational analysis of interactions in structurally available protein-glycosaminoglycan complexes. *Glycobiology* **2016**, *26*, 850–861.
- (43) Pepi, L. E.; Sanderson, P.; Stickney, M.; Amster, I. J. Developments in Mass Spectrometry for Glycosaminoglycan Analysis: A Review. *Mol. Cell. Proteomics* **2021**, *20*, 100025.
- (44) Zappe, A.; Miller, R. L.; Struwe, W. B.; Pagel, K. State-of-the-art glycosaminoglycan characterization. *Mass Spectrom. Rev.* **2022**, *41*, 1040–1071.

- (45) Pomin, V. H.; Wang, X. Glycosaminoglycan-Protein Interactions by Nuclear Magnetic Resonance (NMR) Spectroscopy. *Molecules* **2018**, *23*, 2314.
- (46) Skidmore, M. A.; Guimond, S. E.; Rudd, T. R.; Fernig, D. G.; Turnbull, J. E.; Yates, E. A. The activities of heparan sulfate and its analogue heparin are dictated by biosynthesis, sequence, and conformation. *Connect. Tissue Res.* **2008**, *49*, 140–144.
- (47) Karplus, M.; McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **2002**, *9*, 646–652.
- (48) Hollingsworth, S. A.; Dror, R. O. Molecular dynamics simulation for all. *Neuron* **2018**, *99*, 1129–1143.
- (49) Becker, C. F.; Guimarães, J. A.; Verli, H. Molecular dynamics and atomic charge calculations in the study of heparin conformation in aqueous solution. *Carbohydr. Res.* **2005**, *340*, 1499–1507.
- (50) Verli, H.; Guimarães, J. A. Molecular dynamics simulation of a decasaccharide fragment of heparin in aqueous solution. *Carbohydr. Res.* **2004**, *339*, 281–290.
- (51) Paiardi, G.; Richter, S.; Oreste, P.; Urbinati, C.; Rusnati, M.; Wade, R. C. The binding of heparin to spike glycoprotein inhibits SARS-CoV-2 infection by three mechanisms. *J. Biol. Chem.* **2022**, *298*, 101507.
- (52) Mycroft-West, C. J.; Su, D.; Pagani, I.; Rudd, T. R.; Elli, S.; Gandhi, N. S.; Guimond, S. E.; Miller, G. J.; Meneghetti, M. C. Z.; Nader, H. B.; Li, Y.; Nunes, Q. M.; Procter, P.; Mancini, N.; Clementi, M.; Bisio, A.; Forsyth, N. R.; Ferro, V.; Turnbull, J. E.; Guerrini, M.; Fernig, D. G.; Vicenzi, E.; Yates, E. A.; Lima, M. A.; Skidmore, M. A. Heparin Inhibits Cellular Invasion by SARS-CoV-2: Structural Dependence of the Interaction of the Spike S1 Receptor-Binding Domain with Heparin. *Thromb. Haemostasis* **2020**, *120*, 1700–1715.
- (53) Litov, L.; Petkov, P.; Rangelov, M.; Ilieva, N.; Lilkova, E.; Todorova, N.; Krachmarova, E.; Malinova, K.; Gospodinov, A.; Hristova, R.; Ivanov, I.; Nacheva, G. Molecular Mechanism of the Anti-Inflammatory Action of Heparin. *Int. J. Mol. Sci.* **2021**, *22*, 10730.
- (54) Uciechowska-Kaczmarzyk, U.; Babik, S.; Zsila, F.; Bojarski, K. K.; Beke-Somfai, T.; Samsonov, S. A. Molecular dynamics-based model of VEGF-A and its heparin interactions. *J. Mol. Graphics Modell.* **2018**, *82*, 157–166.
- (55) Gong, F.; Jemth, P.; Galvis, M. L. E.; Vlodavsky, I.; Horner, A.; Lindahl, U.; Li, J.-p. Processing of Macromolecular Heparin by Heparanase. *J. Biol. Chem.* **2003**, *278*, 35152–35158.
- (56) Künze, G.; Huster, D.; Samsonov, S. A. Investigation of the structure of regulatory proteins interacting with glycosaminoglycans by combining NMR spectroscopy and molecular modeling – the beginning of a wonderful friendship. *Biol. Chem.* **2021**, *402*, 1337–1355.
- (57) Lander, A. D. Targeting the glycosaminoglycan-binding sites on proteins. *Chem. Biol.* **1994**, *1*, 73–78.
- (58) Kogut, M. M.; Marcisz, M.; Samsonov, S. A. Modeling glycosaminoglycan-protein complexes. *Curr. Opin. Struct. Biol.* **2022**, *73*, 102332.
- (59) Atkovska, K.; Samsonov, S. A.; Paszkowski-Rogacz, M.; Pisabarro, M. T. Multipose Binding in Molecular Docking. *Int. J. Mol. Sci.* **2014**, *15*, 2622–2645.
- (60) Kmiecik, S.; Gront, D.; Kolinski, M.; Wieteska, L.; Dawid, A. E.; Kolinski, A. Coarse-Grained Protein Models and Their Applications. *Chem. Rev.* **2016**, *116*, 7898–7936.
- (61) Liwo, A.; Czaplewski, C.; Sieradzan, A. K.; Lipska, A. G.; Samsonov, S. A.; Murarka, R. K. Theory and Practice of Coarse-Grained Molecular Dynamics of Biologically Important Systems. *Biomolecules* **2021**, *11*, 1347.
- (62) Bathe, M.; Rutledge, G. C.; Grodzinsky, A. J.; Tidor, B. A Coarse-Grained Molecular Model for Glycosaminoglycans: Application to Chondroitin, Chondroitin Sulfate, and Hyaluronic Acid. *Biophys. J.* **2005**, *88*, 3870–3887.
- (63) Sattelle, B. M.; Shakeri, J.; Cliff, M. J.; Almond, A. Proteoglycans and Their Heterogeneous Glycosaminoglycans at the Atomic Scale. *Biomacromolecules* **2015**, *16*, 951–961.
- (64) Sattelle, B. M.; Shakeri, J.; Almond, A. Does microsecond sugar ring flexing encode 3D-shape and bioactivity in the heparanome? *Biomacromolecules* **2013**, *14*, 1149–1159.
- (65) Samsonov, S. A.; Bichmann, L.; Pisabarro, M. T. Coarse-Grained Model of Glycosaminoglycans. *J. Chem. Inf. Model.* **2015**, *55*, 114–124.
- (66) Liwo, A.; Czaplewski, C.; Sieradzan, A. K.; Lubecka, E. A.; Lipska, A. G.; Golon, L.; Karczyńska, A.; Krupa, P.; Mozolewska, M. A.; Makowski, M.; Gieldoń, A.; Maciejczyk, M.. In *Progress in Molecular Biology and Translational Science*; Strodel, B., Barz, B., Eds.; *Computational Approaches for Understanding Dynamical Systems: Protein Folding and Assembly*; Academic Press, 2020; Vol. 170, pp 73–122.
- (67) Lubecka, E. A.; Liwo, A. A general method for the derivation of the functional forms of the effective energy terms in coarse-grained energy functions of polymers. II. Backbone-local potentials of coarse-grained O1 → 4-bonded polyglucose chains. *J. Chem. Phys.* **2017**, *147*, 115101.
- (68) Liwo, A.; Baranowski, M.; Czaplewski, C.; Gołaś, E.; He, Y.; Jagiela, D.; Krupa, P.; Maciejczyk, M.; Makowski, M.; Mozolewska, M. A.; Niadzedzki, A.; Oldziej, S.; Scheraga, H. A.; Sieradzan, A. K.; Ślusarz, R.; Wirecki, T.; Yin, Y.; Zaborowski, B. A unified coarse-grained model of biological macromolecules based on mean-field multipole–multipole interactions. *J. Mol. Model.* **2014**, *20*, 2306.
- (69) He, Y.; Mozolewska, M. A.; Krupa, P.; Sieradzan, A. K.; Wirecki, T. K.; Liwo, A.; Kachlishvili, K.; Rackovsky, S.; Jagiela, D.; Ślusarz, R.; Czaplewski, C. R.; Oldziej, S.; Scheraga, H. A. Lessons from application of the UNRES force field to predictions of structures of CASP10 targets. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 14936–14941.
- (70) Zhou, R.; Maisuradze, G. G.; Suñol, D.; Todorovski, T.; Macias, M. J.; Xiao, Y.; Scheraga, H. A.; Czaplewski, C.; Liwo, A. Folding kinetics of WW domains with the united residue force field for bridging microscopic motions and experimental measurements. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 18243–18248.
- (71) Gołaś, E.; Maisuradze, G. G.; Senet, P.; Oldziej, S.; Czaplewski, C.; Scheraga, H. A.; Liwo, A. Simulation of the opening and closing of Hsp70 chaperones by coarse-grained molecular dynamics. *J. Chem. Theory Comput.* **2012**, *8*, 1750–1764.
- (72) He, Y.; Maciejczyk, M.; Oldziej, S.; Scheraga, H. A.; Liwo, A. Mean-field interactions between nucleic-acid-base dipoles can drive the formation of a double helix. *Phys. Rev. Lett.* **2013**, *110*, 098101.
- (73) Samsonov, S. A.; Lubecka, E. A.; Bojarski, K. K.; Ganzynkowicz, R.; Liwo, A. Local and long range potentials for heparin-protein systems for coarse-grained simulations. *Biopolymers* **2019**, *110*, No. e23269.
- (74) Khan, S.; Gor, J.; Mulloy, B.; Perkins, S. J. Semi-rigid solution structures of heparin by constrained X-ray scattering modelling: new insight into heparin-protein complexes. *J. Mol. Biol.* **2010**, *395*, 504–521.
- (75) Pavlov, G.; Finet, S.; Tatarenko, K.; Korneeva, E.; Ebel, C. Conformation of heparin studied with macromolecular hydrodynamic methods and X-ray scattering. *Eur. Biophys. J.* **2003**, *32*, 437–449.
- (76) Gay, J. G.; Berne, B. J. Modification of the overlap potential to mimic a linear site–site potential. *J. Chem. Phys.* **1981**, *74*, 3316–3319.
- (77) Case, D. A.; Betz, R. M.; Cerutti, D. S.; Cheatham, T. E., III.; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N. *AMBER*; University of California: San Francisco, 2016.
- (78) Debye, P.; Hückel, E. Zur Theorie der Elektrolyte. I. Gefrierpunktniedrigung und verwandte Erscheinungen. *Phys. Z.* **1923**, *9*, 185–206.
- (79) AL-Bazali, T. Insight into Debye Hückel length (κ^{-1}): smart gravimetric and swelling techniques reveals discrepancy of diffuse double layer theory at high ionic concentrations. *J. Pet. Explor. Prod. Technol.* **2022**, *12*, 461–471.
- (80) Makowski, M.; Liwo, A.; Sobolewski, E.; Scheraga, H. A. Simple Physics-Based Analytical Formulas for the Potentials of Mean Force of the Interaction of Amino-Acid Side Chains in Water. V. Like-Charged Side Chains. *J. Phys. Chem. B* **2011**, *115*, 6119–6129.
- (81) Sieradzan, A. K.; Sans-Duñó, J.; Lubecka, E. A.; Czaplewski, C.; Lipska, A. G.; Leszczyński, H.; Ocetkiewicz, K. M.; Prończ, J.; Czarnul, P.; Krawczyk, H.; Liwo, A. Optimization of parallel implementation of

UNRES package for coarse-grained simulations to treat large proteins. *J. Comput. Chem.* **2023**, *44*, 602–625.

(82) Khalili, M.; Liwo, A.; Rakowski, F.; Grochowski, P.; Scheraga, H. A. Molecular Dynamics with the United-Residue Model of Polypeptide Chains. I. Lagrange Equations of Motion and Tests of Numerical Stability in the Microcanonical Mode. *J. Phys. Chem. B* **2005**, *109*, 13785–13797.

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

(84) R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020.

(85) Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*; Portland, Oregon, 1996; pp 226–231.

(86) Empereur-mot, C.; Capelli, R.; Perrone, M.; Caruso, C.; Doni, G.; Pavan, G. M. Automatic multi-objective optimization of coarse-grained lipid force fields using SwarmCG. *J. Chem. Phys.* **2022**, *156*, 024801.

(87) Empereur-mot, C.; Pedersen, K. B.; Capelli, R.; Crippa, M.; Caruso, C.; Perrone, M.; Souza, P. C. T.; Marrink, S. J.; Pavan, G. M. Automatic Optimization of Lipid Models in the Martini Force Field Using SwarmCG. *J. Chem. Inf. Model.* **2023**, *63*, 3827–3838.

(88) Empereur-Mot, C.; Pesce, L.; Doni, G.; Bochicchio, D.; Capelli, R.; Perego, C.; Pavan, G. M. Swarm-CG: Automatic Parametrization of Bonded Terms in MARTINI-Based Coarse-Grained Models of Simple to Complex Molecules via Fuzzy Self-Tuning Particle Swarm Optimization. *ACS Omega* **2020**, *5*, 32823–32843.

(89) Stroh, K. S.; Souza, P. C. T.; Monticelli, L.; Risselada, H. J. CGCompiler: Automated coarse-grained molecule parameterization via noise-resistant mixed-variable optimization, (accessed 17.07.2023). <https://chemrxiv.org/engage/chemrxiv/article-details/64870e30be16ad5c57d26bd7>.

(90) Bejagam, K. K.; Singh, S.; An, Y.; Berry, C.; Deshmukh, S. A. PSO-Assisted Development of New Transferable Coarse-Grained Water Models. *J. Phys. Chem. B* **2018**, *122*, 1958–1971.

(91) Bejagam, K. K.; Singh, S.; An, Y.; Deshmukh, S. A. Machine-Learned Coarse-Grained Models. *J. Phys. Chem. Lett.* **2018**, *9*, 4667–4672.

(92) Conway, O.; An, Y.; Bejagam, K. K.; Deshmukh, S. A. Development of transferable coarse-grained models of amino acids. *Mol. Syst. Des. Eng.* **2020**, *5*, 675–685.

(93) Wang, J.; Olsson, S.; Wehmeyer, C.; Pérez, A.; Charron, N. E.; de Fabritiis, G.; Noé, F.; Clementi, C. Machine Learning of Coarse-Grained Molecular Dynamics Force Fields. *ACS Cent. Sci.* **2019**, *5*, 755–767.

(94) Husic, B. E.; Charron, N. E.; Lemm, D.; Wang, J.; Pérez, A.; Majewski, M.; Krämer, A.; Chen, Y.; Olsson, S.; de Fabritiis, G.; Noé, F.; Clementi, C. Coarse graining molecular dynamics with graph neural networks. *J. Chem. Phys.* **2020**, *153*, 194101.

(95) Marcisz, M.; Samsonov, S. A. Solvent Model Benchmark for Molecular Dynamics of Glycosaminoglycans. *J. Chem. Inf. Model.* **2023**, *63*, 2147–2157.

(96) Guvench, O.; Whitmore, E. K. Sulfation and Calcium Favor Compact Conformations of Chondroitin in Aqueous Solutions. *ACS Omega* **2021**, *6*, 13204–13217.

(97) Ricard-Blum, S.; Beraud, M.; Raynal, N.; Farndale, R. W.; Ruggiero, F. Structural Requirements for Heparin/Heparan Sulfate Binding to Type V Collagen. *J. Biol. Chem.* **2006**, *281*, 25195–25204.

(98) Bonnans, C.; Chou, J.; Werb, Z. Remodelling the extracellular matrix in development and disease. *Nat. Rev. Mol. Cell Biol.* **2014**, *15*, 786–801.

(99) Gude, F.; Froese, J.; Steffes, G.; Grobe, K. The role of glycosaminoglycan modification in Hedgehog regulated tissue morphogenesis. *Biochem. Soc. Trans.* **2023**, *51*, 983–993.

(100) Verkaar, F.; van Offenbeek, J.; van der Lee, M. M. C.; van Lith, L. H. C. J.; Watts, A. O.; Rops, A. L. W. M. M.; Aguilar, D. C.; Ziarek, J. J.; van der Vlag, J.; Handel, T. M.; Volkman, B. F.; Proudfoot, A. E. I;

Vischer, H. F.; Zaman, G. J. R.; Smit, M. J. Chemokine Cooperativity Is Caused by Competitive Glycosaminoglycan Binding. *J. Immunol.* **2014**, *192*, 3908–3914.

(101) Gschwandtnr, M.; Strutzmann, E.; Teixeira, M. M.; Anders, H. J.; Diedrichs-Möhring, M.; Gerlza, T.; Wildner, G.; Russo, R. C.; Adage, T.; Kungl, A. J. Glycosaminoglycans are important mediators of neutrophilic inflammation in vivo. *Cytokine* **2017**, *91*, 65–73.

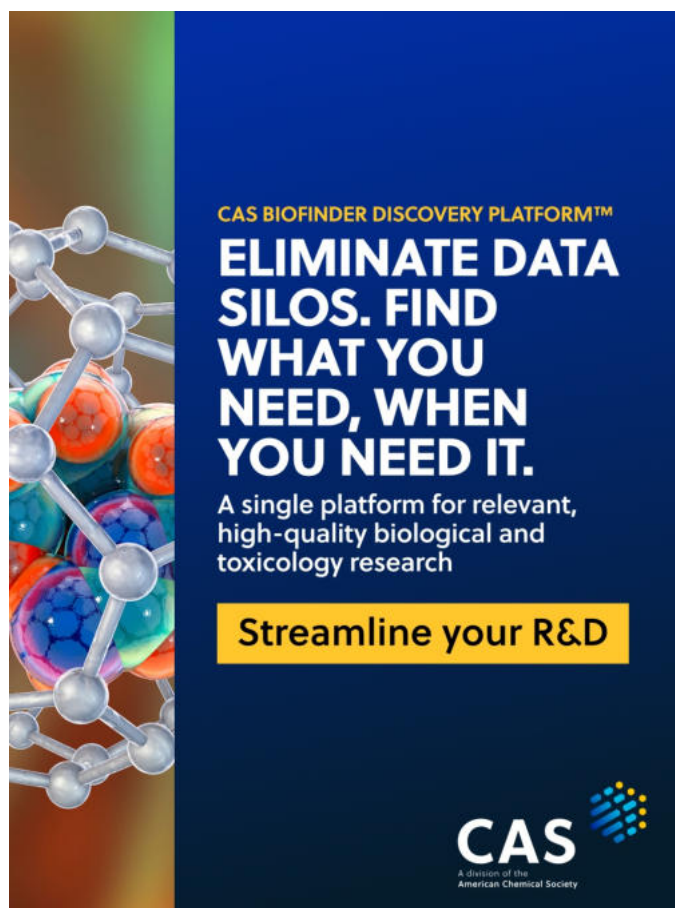
(102) Gangavarapu, P.; Rajagopalan, L.; Kolli, D.; Guerrero-Plata, A.; Garofalo, R. P.; Rajarathnam, K. The monomer-dimer equilibrium and glycosaminoglycan interactions of chemokine CXCL8 regulate tissue-specific neutrophil recruitment. *J. Leukocyte Biol.* **2012**, *91*, 259–265.

(103) Suk, J. Y.; Zhang, F.; Balch, W. E.; Linhardt, R. J.; Kelly, J. W. Heparin Accelerates Gelsolin Amyloidogenesis. *Biochemistry* **2006**, *45*, 2234–2242.

(104) Egashira, M.; Takase, H.; Yamamoto, I.; Tanaka, M.; Saito, H. Identification of regions responsible for heparin-induced amyloidogenesis of human serum amyloid A using its fragment peptides. *Arch. Biochem. Biophys.* **2011**, *511*, 101–106.

NOTE ADDED AFTER ASAP PUBLICATION

This paper was published ASAP on August 16, 2023, with errors in the Table of Contents/Abstract graphic. The corrected version was reposted on August 31, 2023.



CAS BIOFINDER DISCOVERY PLATFORM™

ELIMINATE DATA SILOS. FIND WHAT YOU NEED, WHEN YOU NEED IT.

A single platform for relevant, high-quality biological and toxicology research

Streamline your R&D

CAS
A division of the American Chemical Society

Publication 4. Implementation of the
UNRES/SUGRES-1P Coarse-Grained
Model of Heparin for Simulating
Protein/Heparin Interactions.

Implementation of the UNRES/SUGRES-1P Coarse-Grained Model of Heparin for Simulating Protein/Heparin Interactions

Annemarie Danielsson, Sergey A. Samsonov, and Adam K. Sieradzan*

Cite This: *J. Chem. Theory Comput.* 2024, 20, 10703–10715

Read Online

ACCESS |



Metrics & More

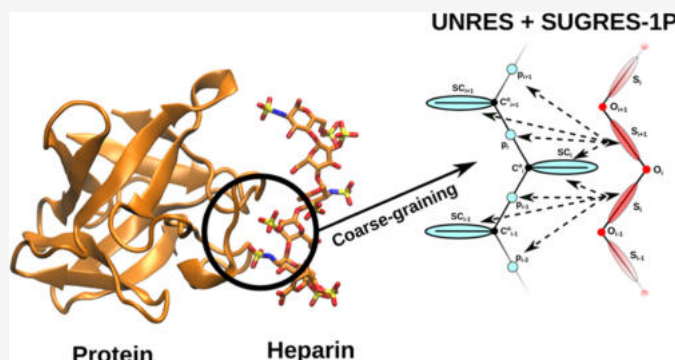


Article Recommendations



Supporting Information

ABSTRACT: Heparin is a natural highly sulfated unbranched periodic polysaccharide that plays a critical role in regulating various cellular events through interactions with its protein targets such as growth factors and cytokines. Although all-atom simulations of heparin-containing systems provide valuable insights into their structural and dynamical properties, long chains of heparin participate in many biologically relevant processes at much bigger scales and longer times than the ones which all-atom MD is able to effectively deal with. Among these processes is the establishment of chemokine gradients, amyloidogenesis, or collagen network organization. To address this limitation, coarse-grained models simplify these systems by reducing the number of degrees of freedom, allowing for the efficient exploration of structural changes within protein/heparin complexes. We introduce and validate the accuracy of a new coarse-grained physics-based model designed for studying protein/heparin interactions, which has been incorporated into the UNRES software package. The effective energy functions from UNRES and SUGRES-1P have been employed for the protein and heparin components, respectively. A good agreement between the obtained coarse-grained simulation results and experimental data confirms the suitability of the combined coarse-grained UNRES and SUGRES-1P model for *in silico* analysis of complex biological phenomena involving heparin, spanning time scales and molecular system sizes not attainable by conventional atomistic molecular dynamics simulations.



1. INTRODUCTION

Glycosaminoglycans (GAGs) are a class of long linear anionic polysaccharides composed of repeating disaccharide units.¹ GAGs are primarily located in the extracellular matrix (ECM), where they form a dynamic scaffold, regulating cell behavior by interacting with growth factors and cytokines.^{2,3} The self-organization of ECM components, facilitated by GAGs, is crucial for tissue development and wound healing.⁴ GAGs also play a vital role in maintaining protein gradients in living organisms, influencing embryonic development and other processes through interactions with growth factors.^{5–9} Additionally, they contribute to biomolecular organization, impacting protein assembly and self-organization at cell surfaces,¹⁰ particularly in cell signaling¹¹ and the modulation of protein aggregation in neurodegenerative diseases.^{12–15}

Heparin (HP), the most negatively charged GAG,¹ is recognized for its anticoagulant and antithrombotic effects.^{16,17} Through interactions with a wide range of protein classes, HP also exhibits anti-inflammatory,¹⁸ antiviral,¹⁹ and anticancer effects.²⁰

Computational analysis and simulation of protein/HP complexes is crucial for understanding their dynamic behavior and interactions, alongside experimental techniques. All-atom (AA) molecular dynamics (MD) simulations, which capture the interactions between individual atoms,²¹ have been

successfully applied for modeling biomolecular systems involving HP polysaccharide chains.^{22–25} However, HP chains can achieve sizes of on average 5–20 kDa *in vivo*,^{26,27} corresponding to a degree of polymerization (dp) of approximately 10–100.²⁸ Their length and flexibility result in a large conformational space that may be undersampled using classical AA MD approaches.²⁹ For example in AA simulations of various protein/GAG complexes, the average RMSD of the GAG between the experimental and simulated structure equals 7.9 ± 3.3 Å for simulations performed using the TIP3P explicit water model.³⁰ This also indicates that protein/GAG systems should be simulated without any restraints as they typically undergo large conformational changes. Additionally, many biological processes involving protein/GAG complexes occur over extended time scales, which necessitates the use of coarse-grained (CG) approaches.

Received: April 29, 2024

Revised: November 6, 2024

Accepted: November 6, 2024

Published: November 21, 2024



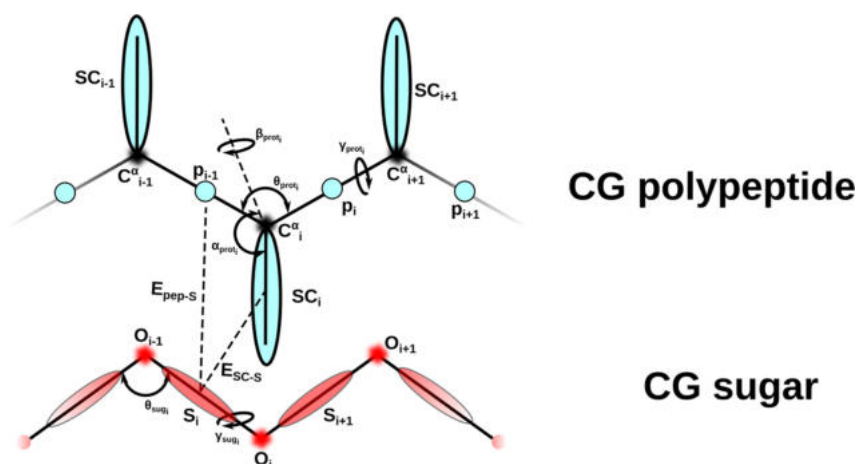


Figure 1. Coarse-grained representation of polypeptide and polysaccharide chains in the UNRES and SUGRES-1P models, respectively.⁵⁸ The unified CG peptide groups (**p**) are represented by a blue filled circle, the unified side-chain beads (**SC**) by a blue filled ellipsoid. The unified CG sugar rings of the polysaccharide (**S**) are shown as red filled ellipsoids. The C^{α} atoms of the polypeptide and the **O** atoms of the polysaccharide are used to define the geometry of the polymer chains and do not function as interaction sites. The virtual bonds connecting the CG beads to the C^{α} and **O** atoms are shown as thick black lines.

In CG methods, groups of atoms are merged into larger pseudoatoms representing individual interaction sites. The group selection is based on criteria such as chemical similarity or spatial proximity. This simplification allows to effectively capture essential system features while minimizing computational costs, thereby enabling the extensive analysis of biologically relevant processes, such as protein binding and unbinding.

Despite the biological importance of GAGs, only a few CG models exist for these complex molecules, with the first one introduced by Bathe et al. in 2005 for modeling chondroitin sulfate and hyaluronic acid.³¹ Sattelle et al. successfully applied their CG model of heparan sulfate to study proteoglycans.³² Samsonov et al.³³ proposed a more detailed CG model for 17 different GAG types, utilizing 28 pseudoatoms to accurately represent various functional groups and achieving good results in modeling GAG chains of various lengths and characteristics. Recently, the MARTINI CG force field has been extended to the inclusion of several types of GAGs and was successful in the reproduction of the thermodynamic and conformational properties of GAGs.³⁴

In this work, we present the integration of protein/HP interaction potentials into the UNICORN (unified coarse-grained) physics-based model,³⁵ combining the UNRES CG model for proteins and SUGRES-1P model for HP. The reduction of the interacting residues to either two (in the case of proteins) or one (in the case of polysaccharides) CG interaction sites per monomeric unit significantly reduces the number of degrees of freedom in the studied systems. We analyzed the dynamics of three protein/HP complexes using 1 μ s CG MD simulations in UNICORN, followed by comparison of the results with the corresponding crystallographic structures: (i) basic fibroblast growth factor (bFGF) with HP dp6, (ii) acidic fibroblast growth factor (aFGF) dimer with HP of varying degrees of polymerization, and (iii) a hepatocyte growth factor/scatter factor (HGF/SF) splice variant with HP dp6. We mainly focused on crystal structures as the other types of structural data are not available.²⁹

Both bFGF and aFGF, known HP-binding proteins, regulate various cellular processes, including cell proliferation, differentiation, and angiogenesis, by interacting with cell surface receptors.^{36–40} HGF/SF, a multifunctional protein, is involved

in cell growth,^{41,42} cell motility,⁴³ and morphogenesis^{44–46} and binds HP, which influences its dimerization and activity.^{47–49} NK1 is a natural splicing variant of HGF/SF, which includes its N-terminal (N) and the first kringle (K1) domains.⁵⁰ It has been observed to exist in a monomeric state in solution in the absence of HP and to adopt a dimeric structure in the presence of HP.^{48,51}

Our CG simulations of the protein/HP complexes highlighted the flexibility of the HP oligosaccharide and the HP-binding residues. The identified binding sites aligned well with those characterized by experiments, suggesting the predictive power of the UNRES and SUGRES-1P models for protein/HP complexes over extended time scales. The presented CG force field can be effectively utilized for analyzing protein/HP systems across a broad spectrum of sizes. It can facilitate the simulation of critical biological processes, including the reorganization of collagen,^{52,53} maintenance of protein gradients,^{5,7–9} and the formation of amyloid aggregates induced by GAGs.^{54,55}

2. METHODS

2.1. Structure of the Coarse-Grained Protein/HP Systems. The protein/HP model utilized in this study incorporates the UNRES⁵⁶ representation for proteins and the SUGRES-1P⁵⁷ representation for the HP component of the simulated systems. The solvent is implicitly considered, influencing the effective potentials governing interactions between CG sites. The energy function integrates UNRES ($U_{\text{prot}}^{\text{UNRES}}$) and SUGRES-1P ($U_{\text{sugar}}^{\text{SUGRES-1P}}$) energies for protein and HP components, along with interaction energies, as defined in eq 1. Subsequent sections elaborate on the specific energy functions employed.

$$U = U_{\text{prot}}^{\text{UNRES}} + U_{\text{sugar}}^{\text{SUGRES-1P}} + U_{\text{prot-sugar}} \quad (1)$$

For a detailed description of the UNRES and SUGRES-1P models within the CG UNICORN force field, including a helpful comparison of the CG structures with the corresponding all-atom representations, we refer the reader to Figure 1 in Liwo et al.⁵⁶

2.2. UNRES Model of Polypeptide Chains. In the UNRES model, the polypeptide chain is reduced to unified

side chains (SC) and peptide groups (p), which function as interaction sites.⁵⁶ The p group is located halfway between consecutive α -carbon atoms (C^α), which are linked by virtual bonds and used only to determine the geometry of the polypeptide chain. The unified SCs are represented by an ellipsoid of revolution connected to the C^α atom by virtual bonds (Figure 1). Each SC is composed of two interaction sites, corresponding to a nonpolar part located in the middle of the SC ellipsoid, and a charged or polar site located at the “head” of the ellipsoid.

The backbone virtual bond angle between the C_{i-1}^α , C_i^α and C_{i+1}^α atoms is represented by $\theta_{\text{prot},i}$ while the backbone virtual bond dihedral angle is symbolized by $\gamma_{\text{prot},i}$ defined by the C_{i-1}^α , C_i^α , C_{i+1}^α and C_{i+2}^α atoms. The location of the center of the united SC_{*i*} group with respect to the C_{i-1}^α , C_i^α and C_{i+1}^α atoms is defined by the angles $\alpha_{\text{prot},i}$ and $\beta_{\text{prot},i}$.

The UNRES force field employs an effective energy function expressed as a cluster-cumulant expansion of the potential of mean force for a protein in an aqueous environment⁵⁹ (eq 2):

$$\begin{aligned}
 U_{\text{prot}}^{\text{UNRES}} = & w_{\text{SCSC}} \sum_{i < j} U_{\text{SC}_i \text{SC}_j} + w_{\text{SCp}} \sum_{i \neq j} U_{\text{SC}_i \text{p}_j} \\
 & + w_{\text{PP}}^{\text{VDW}} \sum_{i < j-1} U_{\text{p}_i \text{p}_j}^{\text{VDW}} + w_{\text{PP}2}^{\text{el}} f_2(T) \sum_{i < j-1} U_{\text{p}_i \text{p}_j}^{\text{el}} \\
 & + w_{\text{tor}2} f_2(T) \sum_i U_{\text{tor}}(\gamma_i, \theta_i, \theta_{i+1}) + w_b \sum_i U_b(\theta_i) \\
 & ++ w_{\text{rot}} \sum_i U_{\text{rot}}(\theta_i, \alpha_{\text{SC}i}, \beta_{\text{SC}i}) \\
 & + w_{\text{bond}} \sum_i U_{\text{bond}}(d_i) + w_{\text{ssbond}} \\
 & \sum_{\text{nss}} U_{\text{ssbond}}(d_{\text{ssbond}}) ++ w_{\text{cor}3}^{(3)} f_3(T) U_{\text{cor}}^{(3)} \\
 & + w_{\text{turn}3}^{(3)} f_3(T) U_{\text{turn}}^{(3)} \quad (2)
 \end{aligned}$$

The energy terms are represented by the symbol U , where $U_{\text{SC}_i \text{SC}_j}$ is the mean free energy of the hydrophobic and hydrophilic interactions between SCs and interactions of SCs with the implicit solvent. Due to the two aforementioned interaction sites comprising each SC, $U_{\text{SC}_i \text{SC}_j}$ encompasses a sum of energy terms: the interaction of the charged or polar parts of the interacting SCs, the interaction of the nonpolar interaction sites, and energy stemming from the interaction of the charged/polar and nonpolar sites with the solvent. Depending on the chemical character of the interacting SCs (same or oppositely charged, polar, hydrophobic), $U_{\text{SC}_i \text{SC}_j}$ contains the following energy terms: E_{GBerne} (van der Waals interactions of the nonpolar interaction sites described by the Gay-Berne potential), E_{LJ} (van der Waals interaction between charged sites), E_{el} (electrostatic interactions of charged sites described by the Coulomb potential), E_{pol} (interaction of charged/polar with nonpolar sites described by the Generalized Born model), $E_{\text{pol}}^{\text{GB}}$ (polarization of the solvent by the SCs), $\Delta F_{\text{cav}}^{\text{iso}}$ (cavity term for charged interaction sites),^{60–63} and ΔF_{cav} (cavity term for nonpolar interaction sites).

$U_{\text{SC}_i \text{p}_j}$ is the excluded-volume interaction between SCs and the peptide groups p. The interactions between peptide groups are split into the Lennard-Jones interaction energy ($U_{\text{p}_i \text{p}_j}^{\text{VDW}}$) and the peptide-group-dipole interactions, ($U_{\text{p}_i \text{p}_j}^{\text{el}}$), which accounts

for the hydrogen bonds formed between peptide groups. The dynamic breaking and formation of disulfide bonds is accounted for by the U_{ssbond} term. The local properties of the polypeptide chain are accounted for by the backbone torsional, virtual bond angle bending, side-chain rotamer, and virtual bond-deformation terms, i.e., $U_{\text{tor}}(\gamma_i, \theta_i, \theta_{i+1})$, $U_b(\theta_i)$, $U_{\text{rot}}(\theta_i, \alpha_{\text{SC}i}, \beta_{\text{SC}i})$ and $U_{\text{bond}}(d_i)$, respectively. The correct reproduction of regular α -helical and β -sheet structures is ensured by the multibody terms, $U_{\text{cor}}^{(3)}$ and $U_{\text{turn}}^{(3)}$ which describe the multibody contributions from the coupling between backbone-local and backbone-electrostatic interactions, and the correlation contributions that involve 3 consecutive peptide groups, respectively.

The weight of each energy term is given by w_x , and the terms corresponding to factors of order higher than 1 are additionally multiplied by the respective temperature factors, defined by eq 3, which reflect the dependence of the first generalized-cumulant term in those factors on temperature:⁶⁴

$$f_n(T) = \frac{\ln[\exp(1) + \exp(-1)]}{\ln\{\exp[(T/T_0)^{n-1}] + \exp[-(T/T_0)^{n-1}]\}} \quad (3)$$

where $T_0 = 300$ K.

In the presented work, all parameters and energy terms relating to the UNRES representation of proteins were adopted from previous calibration studies involving proteins of different structural classes.⁶⁵

2.3. SUGRES-1P Model of the Polysaccharide Chains.

In the SUGRES-1P model, the polysaccharide chain representation is reduced to only one CG bead representing the unified sugar rings (S) (Figure 1), which function as interaction sites of the sugar molecule. The S residue is positioned halfway between consecutive glycosidic linkage oxygen atoms (O) connected by virtual bonds, that serve only to define the geometry of the virtual chain, as do the C^α s in the UNRES model. Each residue is composed of a charged and uncharged interaction site to account for the anisotropy of the CG HP residues, as the center of charge is located off the geometrical center of the residue. The current implementation of SUGRES-1P allows modeling HP molecules composed of N-sulfated and 6-O-sulfated glucosamine (GlcNS6S) and 2-O-sulfated iduronic acid (IdoA2S). The virtual bond angle θ_{sug} corresponds to the angle between three consecutive O atoms, while the virtual bond dihedral angle γ is defined by two planes formed by four consecutive O atoms. The effective energy function of the SUGRES-1P force field is expressed by eq 4.

$$\begin{aligned}
 U = & w_{\text{SiSj}}^{\text{vdW}} \sum_i \sum_{i < j} U_{\text{SiSj}}^{\text{vdW}} + w_{\text{SiSj}2}^{\text{el}} f_2(T) \sum_i \sum_{j < i} U_{\text{SiSj}}^{\text{el}} \\
 & ++ w_{\text{bond}} \sum_i U_{\text{bond}}(d_i) + w_b \sum_i U_b(\theta_i) \\
 & ++ w_{\text{tor}} \sum_i f_2(T) U_{\text{tor}}(\gamma_i, \theta_{i-1}, \theta_i) \quad (4)
 \end{aligned}$$

The term $U_{\text{bond}}(d_i)$ accounts for the energetics of the virtual bond-deformation, where d_i corresponds to the length of the i -th virtual bond, $U_b(\theta_i)$ is the virtual bond angle-deformation term for angle θ_i and the virtual bond-torsional energy term for virtual bond dihedral angle γ_i is given by $U_{\text{tor}}(\gamma_i, \theta_{i-1}, \theta_i)$, where angles θ_{i-1} and θ_i are the adjacent virtual bond angles. The Coulomb (electrostatic) interaction energy between the charged parts of the interaction sites is expressed by $U_{\text{SiSj}}^{\text{el}}$.

The term U_{S,S_j}^{vdW} represents the interaction of the polar and charged parts of the residue, excluding the Coulombic charge–charge interactions, as expressed by eq 5:

$$U_{S,S_j}^{\text{vdW}} = E_{\text{GBerne}} + E_{\text{pol}}^{\text{GB}} + E_{\text{pol}} + \Delta F_{\text{cav}} + \Delta F_{\text{cav}}^{\text{iso}} + E_{\text{LJ}} \quad (5)$$

E_{GBerne} corresponds to the interaction energy between the uncharged parts of the interaction sites, modeled by the Gay-Berne potential.⁶⁶ $E_{\text{pol}}^{\text{GB}}$ gives the contribution to the energy arising from the polarization of the solvent by charged parts of the interaction sites, computed using the generalized Born model. E_{pol} is the polarization energy corresponding to the interactions between the charged and uncharged parts of the interaction sites of two sugar residues. The cavity term of the isotropic charged parts of the interaction sites is expressed by $\Delta F_{\text{cav}}^{\text{iso}}$, while the cavity term of the uncharged parts is calculated by ΔF_{cav} . The isotropic Lennard-Jones potential, E_{LJ} , is used to model the van der Waals interaction energy between two polar parts.

The energy terms are multiplied by the appropriate weights w as well as, in the case of U_{S,S_j}^{el} , temperature factors $f_n(T)$ that reflect the temperature-dependence of the appropriate effective energy terms. The energy term weights and parameters relating to the CG model of HP were adopted from our previous work, wherein parameters were calibrated using HP molecules of length ranging from dp6 to dp68.⁶⁷

2.4. Effective Potentials for the Polypeptide/HP Interactions. The energy function for protein/HP interactions ($U_{\text{prot-sugar}}$) is based on the potentials determined in our earlier work⁵⁸ and is expressed by eq 6:

$$U_{\text{prot-sugar}} = w_{\text{SC-S}} \sum_i \sum_j U_{\text{SC}_i\text{S}_j} + w_{\text{p-S}} \sum_i \sum_j U_{\text{p}_i\text{S}_j} \quad (6)$$

where $U_{\text{SC}_i\text{S}_j}$ and $U_{\text{p}_i\text{S}_j}$ are protein side-chain/sugar, and protein peptide-group/sugar effective interaction potentials, respectively. Each term is multiplied by the appropriate weight (w_x), whose values are in the current study optimized in a systematic manner. As CG force field require calibration,⁶⁸ to accurately model the interactions of proteins with HP, we have performed an adjustment of the energy term weights referring to the interaction between the CG HP residues and the SC and peptide groups of the CG protein residues. To this end, we have incrementally modified the weights by 0.1 within a range of 0.5 to 1.5. The weight optimization was performed for the bFGF/HP dp6 complex.

The effective energy function for protein/HP interactions adopts different forms depending on the nature of the interacting side chains: for charged SCs (both of the opposite and the same charge as the GAG residues), the interaction energy equals the sum of the van der Waals, polarization, and electrostatic interaction energy together with the cavity term for charged parts of the interaction sites (eq 7)

$$U_{\text{SC}_i\text{-S}_j} = U_{\text{SC}_i\text{S}_j}^{\text{vdW}} + \Delta F_{\text{cav}} + \Delta F_{\text{cav}}^{\text{iso}} + U_{\text{SC}_i\text{S}_j}^{\text{el}} + U_{\text{polSC}_i\text{S}_j}^{\text{GB}} + U_{\text{SC}_i\text{S}_j}^{\text{pol}} + U_{\text{SC}_i\text{S}_j}^{\text{LJ}} \quad (7)$$

while for uncharged polar and nonpolar SCs, the interaction energy is reduced to the van der Waals term only:

$$U_{\text{SC}_i\text{-S}_j} = U_{\text{SC}_i\text{S}_j}^{\text{vdW}} + \Delta F_{\text{cav}} \quad (8)$$

The parameters used (Supporting Information Tables S1–S3) in the energy function modeling polypeptide/HP interactions were adopted from Samsonov et al.⁵⁸ Due to unphysical values describing the minimum distance between the charged parts of the interacting beads, these parameters used in the $\Delta F_{\text{cav}}^{\text{iso}}$ energy term were adjusted empirically to obtain realistic results, matching sugar bead sizes comparable to sizes of HP monosaccharide units as obtained using crystallographic experiments.⁶⁹

The stability of MD simulations for protein/HP interactions was assessed in terms of energy conservation in microcanonical MD simulations using the bFGF/HP dp6 complex. Temperature conservation was verified by conducting canonical simulations with the Berendsen thermostat⁷⁰ implemented with UNRES at $T = 300$ K, using the coupling parameter $\tau = 48.9$ fs.

2.5. Coarse-Grained Molecular Dynamics Simulations of Protein/HP Complexes. To study the dynamics of protein/HP complexes, we carried out CG MD simulations of three different proteins complexed with short HP oligosaccharide chains: (i) monomeric bFGF (PDB ID: 1BFC⁶⁹), (ii) dimeric aFGF (PDB ID: 2AXM⁷¹), and (iii) the NK1 protein, a splicing variant of the HGF protein consisting of its N-terminal (N) and first kringle domain (K1) (PDB ID: 3MKP⁵⁰). In the case of aFGF, three different ligands were used: HP dp6 (as in the crystal structure), HP dp8, and HP dp10. The dp8 and dp10 structures were constructed by elongating the HP dp6 molecule from the 2AXM crystal structure by two and four residues using the tleap module of the AMBER package.⁷²

Energy minimization of each system was carried out using the SUMSL algorithm⁷³ before performing simulations. For each system, a series of ten canonical CG MD simulations were conducted at $T = 300$ K in implicit solvent. All weights of the protein/HP interaction energy terms $U_{\text{SC}_i\text{S}_j}$ and $U_{\text{p}_i\text{S}_j}$ were kept at the default values of 1.0, while the weights of the effective energy function terms describing the protein and sugar part of the system were taken from previous studies.^{65,67} Each MD run consisted of 2×10^6 steps with a 0.498 fs time step length (0.01 MTU). The typical time step in UNRES ranges from 0.498 to 4.98 fs (0.01–0.1 MTU). However, based on additional simulations of the bFGF/HP dp6 system, we determined that using a 4.98 fs time step led to excessively high values of electrostatic forces, triggering an automatic algorithm to reduce the time step and indicating potential instability in the system. To avoid such instability, we opted for a conservative 0.498 fs time step in this study. For other timesteps (0.498, 0.996, and 2.49 fs), no qualitative differences were observed in the contact maps between HP and bFGF. We thus recommend using timesteps within the 0.498–2.49 fs range for similar systems. Due to the reduction of the number of explicitly treated degrees of freedom and smoothing of the free energy landscape in the CG model, there is approximately 1000 times speed-up in comparison to the AA explicit water time scales⁷⁴ for the protein system. Therefore, the CG simulation time per each simulation run corresponds to approximately 1 μs AA simulation time. It should be noted that event based speed up for sugar systems has not yet been determined. Snapshots were collected every 1000 steps, yielding a total of 2000 snapshots per trajectory.

2.6. Analysis and Visualization. The CG trajectories were analyzed by in-house scripts using the Python

programming language⁷⁵ as well as the cpptraj module of the AMBER MD package.⁷²

In order to identify representative structures of the protein/HP complexes, CG snapshots with dissociated HP ligands were filtered out. The structures were clustered using the DBSCAN algorithm,⁷⁶ with the same clustering parameters determined from all systems within the study. The minimum number of data points per cluster (minPts) was set to 10, and the ϵ parameter, defining the cluster radius, was set to 2.0 Å. Cluster centers were considered representative structures and were compared to the crystallographic structures of the analyzed protein/HP systems using VMD⁷⁷ and Pymol⁷⁸ software. The representative structures of proteins were converted to AA representations using the PULCHRA software⁷⁹ for visualization and comparison purposes. Calculations of root-mean-square deviation (RMSD), root-mean-square fluctuation (RMSF), and eigenvectors of motion were conducted using the cpptraj module in the AMBER software.⁷² Contacts formed between the proteins and HP were evaluated using the *nativecontacts* command of the cpptraj module using a distance cutoff of 8.0 Å between the CG protein and HP pseudoatoms. Heatmaps depicting contacts were visualized using Python and the matplotlib library.⁸⁰ Venn diagrams were constructed with the aid of the BioVenn⁸¹ and Venny⁸² tools. The mobility of the proteins and HP chains during the CG simulations has been analyzed using Principal Component Analysis (PCA) and in VMD using the ProDy interface⁸³ with default parameters. For visualization purposes the first eigenvector is visualized. The scree plots were drawn.

3. RESULTS

3.1. Energy and Temperature Conservation in MD Simulations. The UNRES/SUGRES-1P CG MD simulations were assessed for energy and temperature conservation using the bFGF/HP dp6 system. The total energy during microcanonical MD simulations remained stable (Supporting Information (SI) Figure S1a), while the kinetic and potential energies fluctuated but maintained a relatively stable sum, demonstrating the conservation of total energy in the microcanonical MD simulation.

In temperature conservation tests, the average bath temperature of the CG simulations is narrower than the theoretical distribution (SI Figure S1b), as expected for the Berendsen thermostat, yet it maintained the system's temperature near the desired value, indicating effective temperature control during the MD simulation.

3.2. Optimization of Energy Term Weights. The optimal weights were determined by evaluating which settings decreased both the number of dissociation events of HP from bFGF and any unbiological structural changes in bFGF assessed through visual inspection and RMSD analysis (SI Table S4). These optimized weights were then applied consistently across all other protein/HP systems analyzed in the study. The optimal weights were determined to be 1.1 for both $w_{\text{SC-S}}$ and $w_{\text{P-S}}$.

3.3. Basic Fibroblast Growth Factor in Complex with HP dp6. From the performed CG simulations, an ensemble of structures was derived from trajectories wherein the HP did not dissociate from the protein, constituting all of the 10 CG simulations. The system reached convergence in those trajectories, as evidenced by the change in the RMSD over time in reference to the initial structure (SI Figures S2 and S3).

The CG structures of the bFGF/HP complex were clustered, resulting in 5 clusters (SI Figure S4). The binding sites identified in the cluster representatives 1 through 4 are similar in shape and size, encompassing largely overlapping sets of residues (SI Table S3). However, the orientation of HP within the binding site varies across the different representatives, which underscores the flexibility and adaptability of the binding site in accommodating the ligand in distinct conformations. To evaluate the similarity of the obtained protein structures, the cluster representatives were compared with the crystal structure using the TM-score⁸⁴ and GDT-TS⁸⁵ for C α atoms (Table 1).

Table 1. TM-Score and GDT-TS Values of the Cluster Representatives in Reference to the Crystal Structure of bFGF

	TM-score	GDT-TS
cluster 1	0.56	0.97
cluster 2	0.32	0.27
cluster 3	0.34	0.32
cluster 4	0.32	0.29
cluster 5	0.35	0.32

The CG bFGF structure represented by the cluster centroid of the most populated cluster (65.7% of the CG MD trajectory), closely resembles the crystal structure of bFGF in both 3D arrangement and as measured by the TM-score and GDT-TS. The less populated clusters exhibit lower similarity, likely due to the flexibility of specific regions of the bFGF protein. These findings indicate that the CG MD simulation effectively captures the correct structure of the bFGF protein in complex with HP, along with the dynamic flexibility observed throughout the simulation and captured by the clustering in the form of the less populated clusters.

The centroid frame of the most populated cluster was chosen as the representative CG structure and superimposed onto the experimental structure of bFGF/HP (PDB ID: 1BFC). To quantify the deviation between the experimental and the representative CG structure, RMSD calculations were applied to the protein and HP part, yielding an RMSD of 4.6 and 11.7 Å, respectively.

The HP-binding residues in the representative CG and crystal structures (Figure 2) comprises 15 polar and charged residues in both, although the exact composition of the HP-binding site varies (SI Figure S5). Several experimentally observed HP-binding residues (Arg-82, Lys-120, Gln-124, Lys-126, Lys-136)^{69,86,87} are observed to be in close proximity to HP in the representative CG frame.

To comprehensively characterize the HP-binding region established by bFGF, we conducted an analysis on the frequency of contacts formed between the pseudoatoms of HP and those of the bFGF protein for the entire trajectory (SI Figures S6 and S7). While minor differences can be observed between the replicates, all show frequent contacts established between HP and residues 117 to 138 of bFGF, corresponding to the same region in close proximity to HP as in the crystal structure. In most of the MD replicates, contacts could also be observed for residues 71 to 85, although differing in frequency. These residues form a β sheet not directly interacting with HP in the crystal structure. Additionally, patches of bFGF residues ranging from 21 to 51 were found to establish contacts with HP in a lower intensity for only some of the replicates,

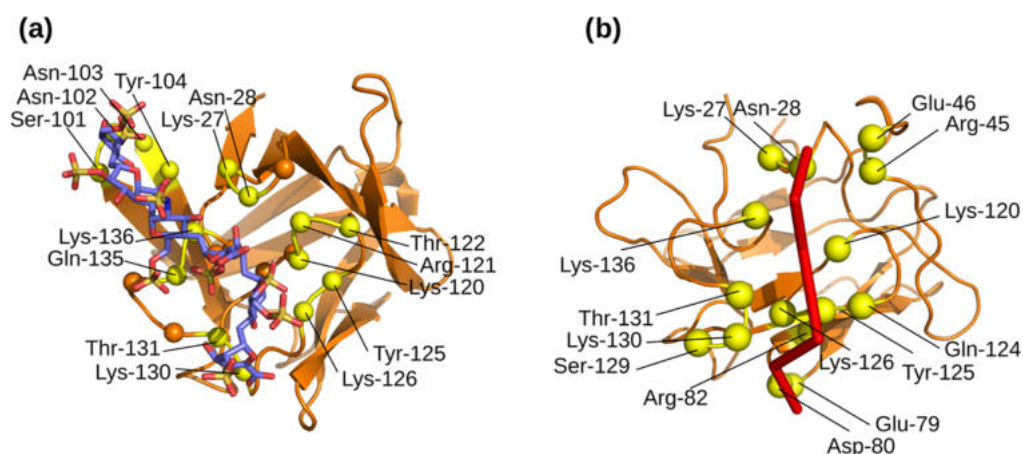


Figure 2. (a) Crystal structure of bFGF/HP dp6. (b) The representative frame of the CG trajectory of bFGF/HP dp6. The protein is shown in cartoon representation, and C α atoms of charged and polar residues within an 8 Å radius of HP are shown as labeled spheres. HP is shown in licorice representation.

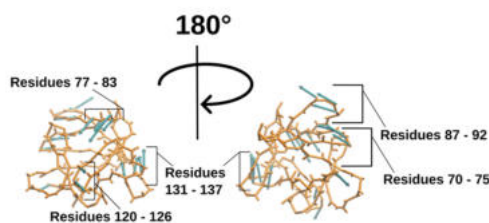
corresponding to regions of bFGF exhibiting the greatest diversity in contact formation during the MD among the replicates. While contacts between HP and the experimentally established binding site were maintained in all of the analyzed CG trajectories, the considerable flexibility exhibited by both the HP ligand and the protein itself could be captured by the CG simulations.

The flexibility of the protein observed in our simulation was particularly pronounced for the N- and C-termini as well as for residues Pro-37 to Pro-50, Tyr-74 to Lys-87, and Arg-121 to Gln-124 (SI Figure S8). In the crystal structure only residues Arg-124 to Gln-124 and the C-terminus are located near the HP chain, while in the representative CG structure all four of the regions can be seen to establish contacts with HP, indicating increased mobility of those residues to increase the contact with the GAG.

Reports in the literature present varying evidence on the conformational changes occurring in the bFGF protein upon HP binding. Some studies indicate no discernible changes upon the binding of HP to (dimeric) bFGF,⁸⁸ while others suggest small yet consistent alterations in the conformation of bFGF upon the binding to HP.⁸⁹ In order to identify the key motions occurring during the CG simulations, we have performed an analysis of the eigenvectors associated with the normal modes of motion of the bFGF/HP dp6 complex using PCA. This was juxtaposed with a similar investigation performed on a CG trajectory of the bFGF protein in its apo form (Figure 3). The bFGF protein was characterized by significantly less mobility in its *apo* form compared to the bFGF/HP dp6 complex, manifesting only a limited set of motions attributable to a select group of residues. Since the eigenvalues corresponding to these movements represent nearly 20% of the variance, the conclusions could be not fully representative and should be interpreted rather qualitatively.

This HP-induced mobility of bFGF fragments potentially contributes to the formation of bFGF complexes with other proteins.^{88,90} In the crystal structure of the ternary complex involving bFGF, FGFR1, and HP, the loop region between residues Gly-43 and Lys-53 of bFGF is located in close proximity to and likely interacts with FGFR1.⁹¹ Analysis of cluster representatives reveals that in the second most populated cluster (SI Figure S4), the loop does not interact with HP. This observation underscores the loop's significant

(a) bFGF *apo*



(b) bFGF/HP dp6

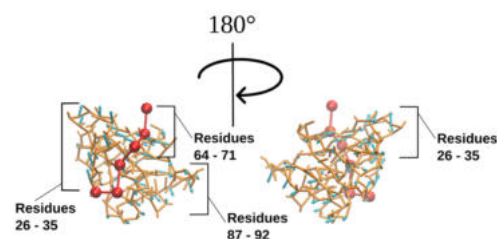


Figure 3. Motions of bFGF/HP dp6 during the CG simulation obtained by PCA. The protein is shown in orange licorice representation, the HP chain is shown in red licorice and sphere representation. Arrows in cyan represent the motions of every fourth residue in the dimension of the first principal component (PC1).

flexibility during simulation. Furthermore, residues Ser-101 to Tyr-104 along with Ser-109 to Tyr-116 are observed to engage in interactions with HP within the bFGF/FGFR1/HP complex, enclosing the HP chain from both sides.

Our findings highlight the dynamic nature of the bFGF/HP complex, which appear pivotal in facilitating multimerization and the assembly of higher-order complexes. Previous research has elucidated how bFGFs form homodimers and tetramers supported by HP chains.^{92–94} The dynamic characteristics observed in our simulations might not be evident in the crystal structure due to the constraints imposed by crystal packing, however they are in line with the documented need for structural flexibility to support the integration of bFGF into complex assemblies.

3.3.1. Influence of the Starting Position on Complex Stability. In order to examine the influence of the initial position of the complex in the CG trajectory on the stability of

the complex, we have performed CG simulations of the bFGF/HP complex with the same setup, i.e., parameters and weights, as previously, but modified the initial frame by shifting the position of HP (Figure 4). The displacement of the HP

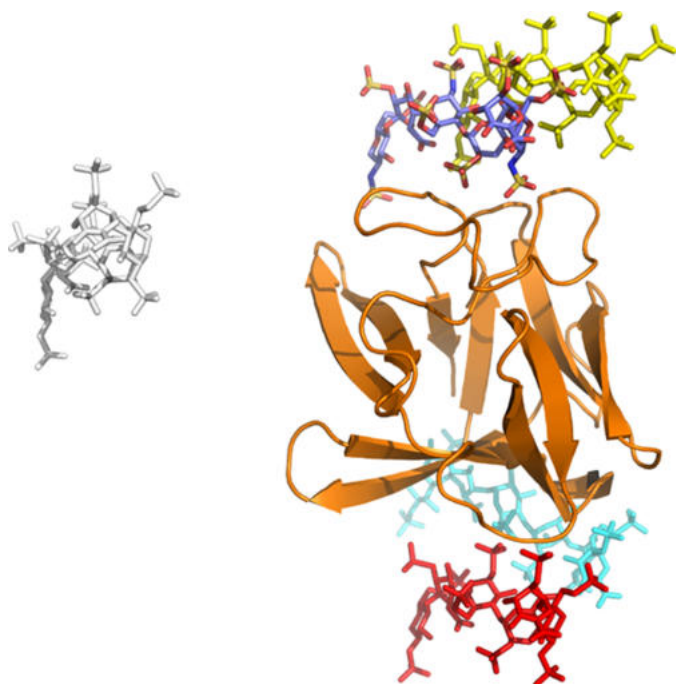


Figure 4. Initial positions in reference to the crystal structure of bFGF/HP, with the shifted HP molecule shown in yellow (position 1), red (position 2), cyan (position 3), and white (position 4).

molecule in reference to the crystal structure was measured using RMSD and equaled 9.2 Å for position 1, 39.9 Å for position 2, 38.7 Å for position 3, and 38.5 Å for position 4.

Assuming the HP position in the crystal structure corresponds to the optimal binding pose of HP, we found that a suboptimal position affected the stability of the system. While in all CG MD trajectories, irrespective of the initial binding pose, HP was observed to establish contacts with bFGF, the number of MD trajectories in which it later dissociated from bFGF increased, as evidenced in Table 2.

Table 2. Number of CG MD Simulations, in Which HP was Found to Dissociate from bFGF

initial position	no. of MD with dissociation events
1	2 out of 10
2	2 out of 10
3	3 out of 10
4	2 out of 10

3.3.2. bFGF Dimer Stability in the Presence of HP. Multiple literature sources identify HP as a stabilizing factor for the bFGF dimer. To investigate the ability of our CG model to reproduce this phenomenon, we have performed 10 CG simulation replicates for the bFGF dimer without HP and 10 CG simulation replicates in the presence of HP (starting structures shown in Figure 5).

The bFGF dimer was observed to spontaneously dissociate and bind again over the course of the MD trajectories for all of the replicates. In contrast, in the presence of HP, the two

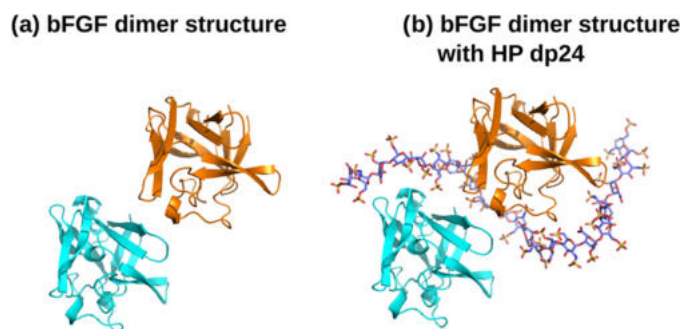


Figure 5. Initial positions of bFGF dimer (a) without HP, (b) with HP.

bFGF monomers bound tightly and the complex was stable once formed throughout each of the CG trajectories ().

3.4. Acidic Fibroblast Growth Factor in Complex with HP Chains of Different Lengths. Interactions between HP and FGF proteins are affected by HP chain length, where specific lengths favor the formation of a stable, functional FGF/HP complex, underscoring the importance of understanding HP length's role in their interaction for elucidating molecular intricacies in physiological processes.^{95–98} We have conducted CG MD simulations of dimeric aFGF in the UNICORN force field with HP of three different lengths: dp6, dp8, and dp10. The obtained structures were compared to experimental data.⁷¹

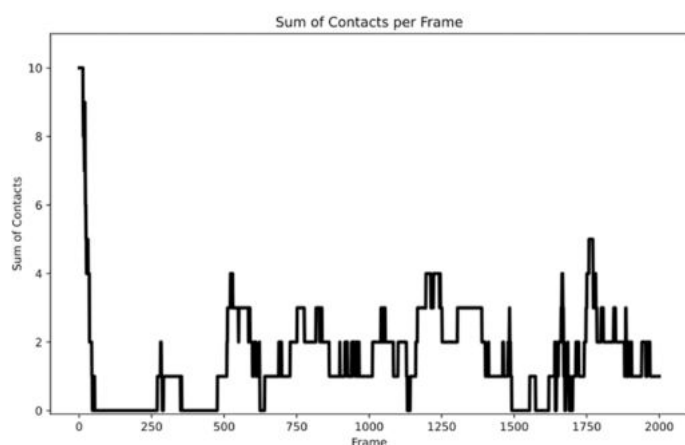
CG MD trajectories in which HP did not dissociate from the aFGF dimer were analyzed for stability. The amount of MD simulations in which HP was found to dissociate from the aFGF dimer was comparable for all three analyzed HP chain lengths. In the case of aFGF/HP dp6, 8 out of 10 simulations showed no dissociation event and were taken for further analysis. For aFGF/HP dp8, this equaled to 6 out of 10 MD simulations. For aFGF/HP dp10, 6 out of 10 simulations showed no dissociation event, however the reassociation of the aFGF/HP dp10 complex could be observed toward the end of another CG MD simulation. The flexibility of the HP chain in the complexes was overall similar, as evidenced by the RMSD values of the oligosaccharide chains in reference to the first frame of the MD simulations (SI Figures S10–S15).

Analysis of the three aFGF/HP complexes using PCA revealed a consistent twisting and squeezing motion, with monomers rotating in opposite directions and converging toward the HP chain (SI Figure S17). The magnitude of this motion was reduced by longer HP chains.

While the overall structure of the complex was maintained in the CG trajectories, the dimer appeared more compact in the representative CG frames compared to the crystal structure. The radii of gyration for the complexes in the CG simulations were slightly smaller yet still comparable to experimental values: 19.8 ± 1.3 Å for aFGF/HP dp6, 21.0 ± 1.5 Å for aFGF/HP dp8, and 20.3 ± 0.9 Å for aFGF/HP dp10, as opposed to 23.4 Å for the experimental structure of aFGF/HP dp6. This compaction may result from stronger interactions between the monomers and the interposed HP chain. This may be attributable to an enhanced attraction between the monomers and the HP chain positioned in-between (Figure 7).

The HP-binding regions were comparable in size for the crystal (11 polar and charged residues) and the CG representative structures (9 polar or charged residues in aFGF/HP dp6, 12 in aFGF/HP dp8, and 14 in aFGF/HP

(a) bFGF dimer without HP



(b) bFGF dimer with HP dp24

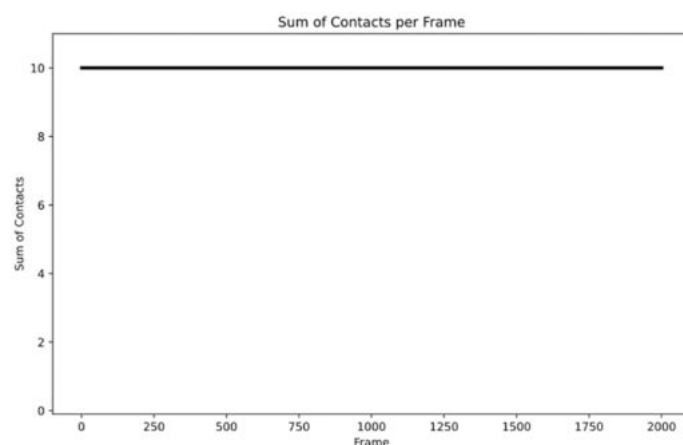
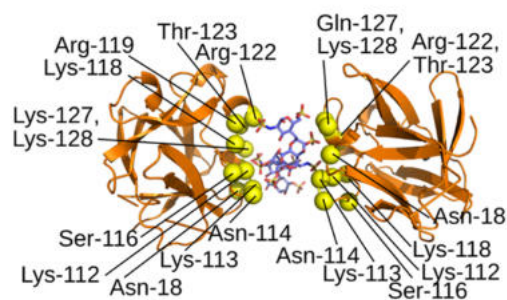
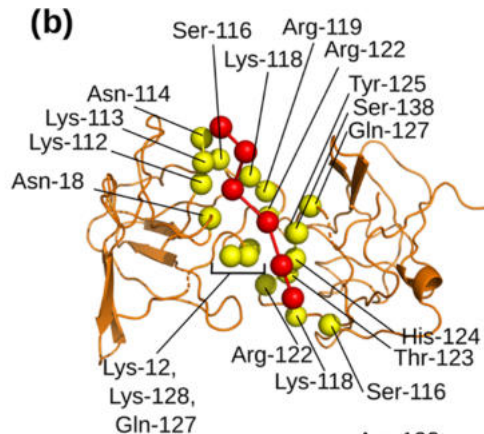


Figure 6. Sum of the contacts performed from CG MD trajectories for (a) the bFGF dimer without HP, and (b) bFGF dimer in the presence of HP dp24, in which the bFGF monomers establish contacts with each other within 8 Å.

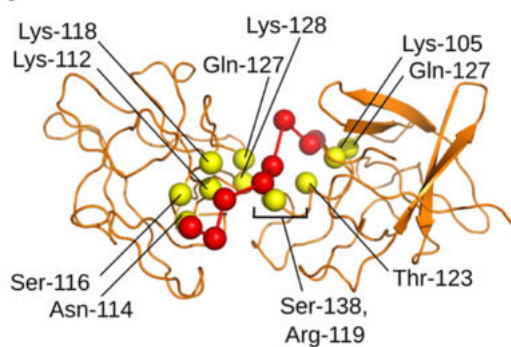
(a)



(b)



(c)



(d)

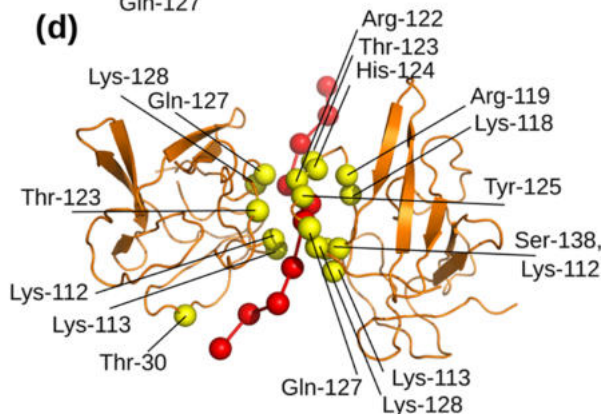


Figure 7. (a) Crystal structure of aFGF/HP dp6. (b) Representative CG structure of FGF/HP dp6. (c) The representative CG structure of FGF/HP dp8. (d) The representative CG structure of FGF/HP dp10. The protein is shown in cartoon representation, with C_{α} atoms of HP-binding residues shown as labeled spheres. HP is shown in licorice representation.

dp10). The experimentally determined HP-binding residues (Asn-18, Lys-112, Lys-113, Asn-114, Lys-118, Arg-119, and Arg-122)^{86,99,100} were identified in all of the CG representatives (SI Figure S18), indicating strong electrostatic interactions between the protein and HP.

Irrespective of the HP chain length in a complex, a prevalence of contacts with HP could be observed for one of the monomers of aFGF, indicating a difference in contact intensity across the MD trajectories between the two aFGF monomers (SI Figure S20). The main region of frequent

contacts encompasses residues Lys-101 to Pro-121, corresponding roughly to the same binding site as observed in the crystal structure. Interestingly, for aFGF/HP dp6 contacts between the oligosaccharide and the N-terminus could be observed, which decreased in frequency with increasing HP chain length and almost no contacts established between aFGF and HP dp10 for this region (SI Figures S21–S23). This could possibly hint at an increased stability of the complex for longer HP chains both for the protein and oligosaccharide components.

3.5. NK1 Splicing Variant of the Hepatocyte Growth Factor with HP dp6. Among its various functions, HP is recognized for stabilizing the structures of multimeric proteins and enhancing dimerization of the NK1 splicing variant of HGF/SF. To analyze the binding site of HP to monomeric NK1, ten CG MD simulations were conducted, with HP remaining bound to NK1 in five trajectories, which were chosen for further analysis. System stability was assessed by RMSD calculations (SI Figures S33, S34).

Three representative binding poses could be identified by clustering the CG trajectory frames via DBSCAN (SI Figure S35). In the representative trajectory frame of the most populated cluster, the HP chain maintains its attachment to the experimentally determined binding site, but adopts a shifted binding pose compared to the crystal structure (Figure 8).

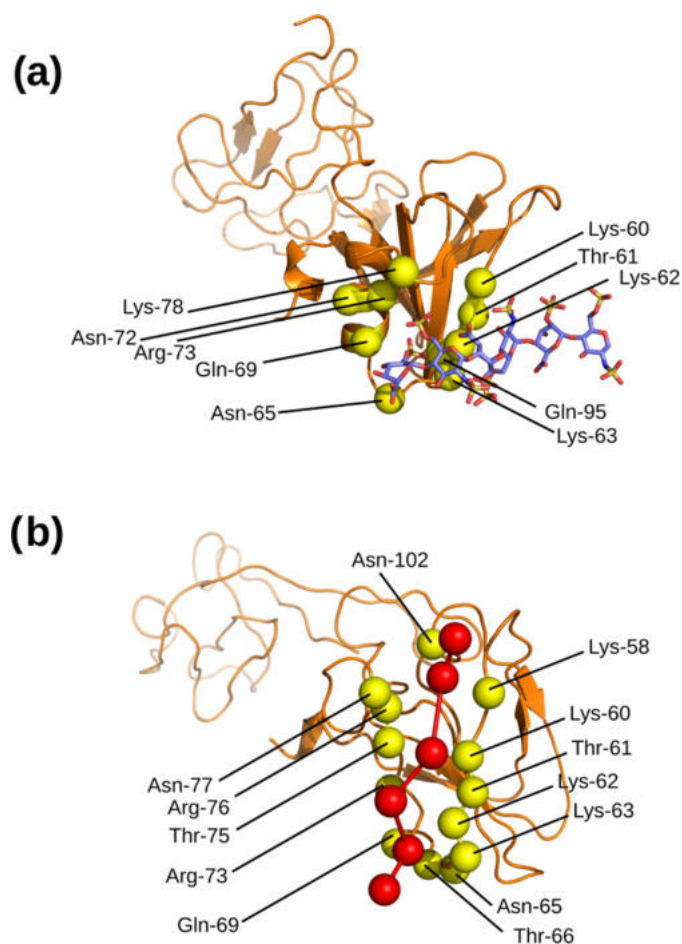


Figure 8. (a) Representative frame of the CG trajectory of NK1/HP dp6. (b) Crystal structure of bFGF/HP dp6. The protein is shown in cartoon representation, and C α atoms of charged and polar residues within an 8 significant flexibility radius of HP are shown as labeled spheres. HP is shown in licorice representation.

PCA analysis of the CG MD trajectories showed relatively little movement of the complex (SI Figure S37), as the complex appears to be stable when no dissociation of the HP chain is observed. The greatest mobility of the NK1 protein can be observed predominantly for the N-terminal residues as evidenced by the RMSF analysis (SI Figure S39), while the fluctuation of the HP residues over the course of the MD simulation appear to be greatly reduced in comparison to the protein, again highlighting the stability of NK1/HP binding (SI Figure S40). The binding pose in the CG structure involves

contacts with experimentally confirmed HP-binding residues (Lys-58, Lys-60, Thr-61, Lys-62, Lys-63, Arg-73)^{51,101,102} (SI Figure S23). An analysis of contacts formed between NK1 and HP the CG MD trajectory (SI Figure S28) reveals that a broad range of NK1 residues establishes frequent contacts with the oligosaccharide. For most of the replicates, a consistent region of contact encompasses residues Lys-58 to Val-64, Cys-70 to Lys-78, and Lys-94 to Phe-101. However, significant differences between the replicates can be observed indicating a very flexible system and mobility of the protein, facilitating a variety of contacts with HP (SI Figure S29).

The observed interactions are consistent with scientific literature emphasizing the role of HP in promoting the oligomerization of NK1,^{47,48,101} particularly through binding to the N-terminal domain,^{51,103} as a prevalence of the frequent contacts across MD simulation replicates can be observed for the N-terminus and adjacent regions of NK1 (SI Figure S28).

4. CONCLUSIONS

In this work, we present the implementation and evaluation of the combined UNRES and SUGRES-1P CG model designed for MD simulations of protein/HP complexes. This study marks the first ever application of a physics-based CG model for analyses of GAG-containing biomolecular systems.

We predict the interactions maintained in protein/HP complexes as well as native contacts established between the proteins and HP oligosaccharides in agreement with the experimental data. The high natural flexibility of protein/HP complexes is also captured by the presented CG MD simulations.

The combined UNRES and SUGRES-1P model demonstrated robust performance, evidenced by the interaction of the experimentally detected residues with HP and the maintenance of stable binding poses throughout repeated MD simulations in the tested complexes. Moreover, we were able to correctly reproduce the geometry of these analyzed complexes in terms of the radius of gyration, which closely matched that of the experimental structures. The presented approach could be applied for simulating complex biological processes across long time scales in large biomolecular systems, thereby improving the understanding of the extracellular matrix function.

We plan to develop backmapping software for the GAG system to reconstruct the atomistic details of protein/carbohydrate interactions. Additionally, we aim to use a machine learning approach to parametrize all types of GAGs and to establish structure-parameter relationships.

■ ASSOCIATED CONTENT

Data Availability Statement

The protein/HP interaction energy function was implemented in FORTRAN90 and is available free of charge at: <https://unres.pl/downloads>.

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jctc.4c00575>.

Tables S1–S4 - force field interaction parameters; Table S5 - weight optimization evaluation; Tables S6, S7 - charged and polar residues in contact with HP; Figure S1 - energy as a time function and temperature distribution; Figures S2, S3, S11, S12, S13, S14, S15, S16, S31, S32 - RMSD vs time for simulated systems; Figures S4, S17, S18, S19, S33 - cluster centroids for

simulated systems; Figures S5, S21, S34 - Venn diagrams showing the overlap in HP-binding residues of the representative CG frame and the crystal structure for simulated systems; Figures S6, S7, S22, S23, S24, S25, S35, S36 - contact maps for simulated systems; Figures S8, S9, S29, S30, S39, S40 - RMSF plots for simulated systems; Figures S10, S26, S27, S28, S38 - Scree plots for PCA analysis; Figures S20, S37 - motions obtained from PCA analysis (PDF)

AUTHOR INFORMATION

Corresponding Author

Adam K. Sieradzan – Faculty of Chemistry, University of Gdansk, 80-308 Gdansk, Poland; orcid.org/0000-0002-2426-3644; Email: adam.sieradzan@ug.edu.pl

Authors

Annemarie Danielsson – Faculty of Chemistry, University of Gdansk, 80-308 Gdansk, Poland; orcid.org/0000-0002-6151-4449

Sergey A. Samsonov – Faculty of Chemistry, University of Gdansk, 80-308 Gdansk, Poland; orcid.org/0000-0002-5166-4849

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.jctc.4c00575>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the Polish National Science Centre (Narodowe Centrum Nauki), grants UMO-2018/31/G/ST4/00246 (to A.D. and S.A.S.) and UMO-2017/27/B/ST4/00926 (to A.K.S.). Calculations were conducted by using the resources of the cluster at the Faculty of Chemistry, University of Gdańsk.

REFERENCES

- (1) Esko, J. D.; Kimata, K.; Lindahl, U. *Essentials of Glycobiology*, 2nd ed.; Varki, A.; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Bertozzi, C. R.; Hart, G. W.; Etzler, M. E., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor (NY), 2009.
- (2) Mitsi, M.; Forsten-Williams, K.; Gopalakrishnan, M.; Nugent, M. A. A Catalytic Role of Heparin within the Extracellular Matrix. *J. Biol. Chem.* **2008**, *283*, 34796–34807.
- (3) Schönherr, E.; Hausser, H. J. Extracellular matrix and cytokines: a functional unit. *Dev. Immunol.* **2000**, *7*, 89–101.
- (4) Goh, M.; Hwang, Y.; Tae, G. Epidermal growth factor loaded heparin-based hydrogel sheet for skin wound healing. *Carbohydr. Polym.* **2016**, *147*, 251–260.
- (5) Gude, F.; Froese, J.; Steffes, G.; Grobe, K. The role of glycosaminoglycan modification in Hedgehog regulated tissue morphogenesis. *Biochem. Soc. Trans.* **2023**, *51*, 983–993.
- (6) Gude, F.; Froese, J.; Manikowski, D.; Di Iorio, D.; Grad, J.-N.; Wegner, S.; Hoffmann, D.; Kennedy, M.; Richter, R. P.; Steffes, G.; Grobe, K. Hedgehog is relayed through dynamic heparan sulfate interactions to shape its gradient. *Nat. Commun.* **2023**, *14* (1), No. 758.
- (7) Verkaar, F.; van Offenbeek, J.; van der Lee, M. M. C.; van Lith, L. H. C. J.; Watts, A. O.; Rops, A. L. W. M. M.; Aguilar, D. C.; Ziarek, J. J.; van der Vlag, J.; Handel, T. M.; Volkman, B. F.; Proudfoot, A. E. I.; Vischer, H. F.; Zaman, G. J. R.; Smit, M. J. Chemokine Cooperativity Is Caused by Competitive Glycosaminoglycan Binding. *J. Immunol.* **2014**, *192*, 3908–3914.
- (8) Gschwandtner, M.; Strutzmann, E.; Teixeira, M. M.; Anders, H. J.; Diedrichs-Möhning, M.; Gerlza, T.; Wildner, G.; Russo, R. C.; Adage, T.; Kungl, A. J. Glycosaminoglycans are important mediators of neutrophilic inflammation in vivo. *Cytokine* **2017**, *91*, 65–73.
- (9) Gangavarapu, P.; Rajagopalan, L.; Kolli, D.; Guerrero-Plata, A.; Garofalo, R. P.; Rajarathnam, K. The monomer-dimer equilibrium and glycosaminoglycan interactions of chemokine CXCL8 regulate tissue-specific neutrophil recruitment. *J. Leukocyte Biol.* **2011**, *91*, 259–265.
- (10) Wang, C.-H.; Davamani, F.; Sue, S.-C.; Lee, S.-C.; Wu, P.-I.; Tang, F.-M.; Shih, C.; Huang, T.-h.; Wu, W.-g. Cell surface heparan sulfates mediate internalization of the PWWP/HATH domain of HDGF via macropinocytosis to fine-tune cell signalling processes involved in fibroblast cell migration. *Biochem. J.* **2011**, *433*, 127–138.
- (11) Xue, S.; Zhou, F.; Zhao, T.; Zhao, H.; Wang, X.; Chen, L.; Li, J.-p.; Luo, S.-Z. Phase separation on cell surface facilitates bFGF signal transduction with heparan sulphate. *Nat. Commun.* **2022**, *13*, No. 1112.
- (12) Ramachandran, G.; Udgaonkar, J. B. Understanding the kinetic roles of the inducer heparin and of rod-like protofibrils during amyloid fibril formation by Tau protein. *J. Biol. Chem.* **2011**, *286*, 38948–38959.
- (13) Klajnert, B.; Cortijo-Arellano, M.; Bryszewska, M.; Cladera, J. Influence of heparin and dendrimers on the aggregation of two amyloid peptides related to Alzheimer's and prion diseases. *Biochem. Biophys. Res. Commun.* **2006**, *339*, 577–582.
- (14) Stopschinski, B. E.; Thomas, T. L.; Nadj, S.; Darvish, E.; Fan, L.; Holmes, B. B.; Modi, A. R.; Finnell, J. G.; Kashmer, O. M.; Estill-Terpack, S.; Mirbaha, H.; Luu, H. S.; Diamond, M. I. A synthetic heparinoid blocks Tau aggregate cell uptake and amplification. *J. Biol. Chem.* **2020**, *295*, 2974–2983.
- (15) Vieira, T. C. R. G.; Cordeiro, Y.; Caughey, B.; Silva, J. L. Heparin binding confers prion stability and impairs its aggregation. *FASEB J.* **2014**, *28*, 2667–2676.
- (16) Linhardt, R. J.; St; Ange, K.; Dordick, J. S.; Linhardt, R. J. Heparin and anticoagulation. *Front. Biosci.* **2016**, *21*, 1372–1392.
- (17) Gray, E.; Hogwood, J.; Mulloy, B. The anticoagulant and antithrombotic mechanisms of heparin. *Handb. Exp. Pharmacol.* **2012**, *207*, 43–61.
- (18) Young, E. The anti-inflammatory effects of heparin and related compounds. *Thromb. Res.* **2008**, *122*, 743–752.
- (19) Braz-de Melo, H. A.; Faria, S. S.; Pasquarelli-do Nascimento, G.; Santos, I. d. O.; Kobinger, G. P.; Magalhaes, K. G. The Use of the Anticoagulant Heparin and Corticosteroid Dexamethasone as Prominent Treatments for COVID-19. *Front. Med.* **2021**, *8*, No. 615333.
- (20) Ma, S.-N.; Mao, Z.-X.; Wu, Y.; Liang, M.-X.; Wang, D.-D.; Chen, X.; Chang, P.-A.; Zhang, W.; Tang, J.-H. The anti-cancer properties of heparin and its derivatives: a review and prospect. *Cell Adhes. Migr.* **2020**, *14*, 118–128.
- (21) Lipp, V. P.; Rethfeld, B.; Garcia, M. E.; Ivanov, D. S. Atomistic-continuum modeling of short laser pulse melting of Si targets. *Phys. Rev. B* **2014**, *90*, No. 245306.
- (22) Paiardi, G.; Richter, S.; Oreste, P.; Urbinati, C.; Rusnati, M.; Wade, R. C. The binding of heparin to spike glycoprotein inhibits SARS-CoV-2 infection by three mechanisms. *J. Biol. Chem.* **2022**, *298*, No. 101507.
- (23) Mycroft-West, C. J.; Su, D.; Pagani, I.; Rudd, T. R.; Elli, S.; Gandhi, N. S.; Guimond, S. E.; Miller, G. J.; Meneghetti, M. C. Z.; Nader, H. B.; Li, Y.; Nunes, Q. M.; Procter, P.; Mancini, N.; Clementi, M.; Bisio, A.; Forsyth, N. R.; Ferro, V.; Turnbull, J. E.; Guerrini, M.; Fernig, D. G.; Vicenzi, E.; Yates, E. A.; Lima, M. A.; Skidmore, M. A. Heparin Inhibits Cellular Invasion by SARS-CoV-2: Structural Dependence of the Interaction of the Spike S1 Receptor-Binding Domain with Heparin. *Thromb. Haemostasis* **2020**, *120*, 1700–1715.
- (24) Uciechowska-Kaczmarzyk, U.; Babik, S.; Zsila, F.; Bojarski, K. K.; Beke-Somfai, T.; Samsonov, S. A. Molecular dynamics-based

model of VEGF-A and its heparin interactions. *J. Mol. Graphics Modell.* **2018**, *82*, 157–166.

(25) Litov, L.; Petkov, P.; Rangelov, M.; Ilieva, N.; Lilkova, E.; Todorova, N.; Krachmarova, E.; Malinova, K.; Gospodinov, A.; Hristova, R.; Ivanov, I.; Nacheva, G. Molecular Mechanism of the Anti-Inflammatory Action of Heparin. *Int. J. Mol. Sci.* **2021**, *22*, No. 10730.

(26) Connors, M. S.; Money, S. R. The New Heparins. *Ochsner J.* **2002**, *4*, 41–47.

(27) Shute, J. K. Heparin, Low Molecular Weight Heparin, and Non-Anticoagulant Derivatives for the Treatment of Inflammatory Lung Disease. *Pharmaceuticals* **2023**, *16*, No. 584.

(28) Zang, L.; Zhu, H.; Wang, K.; Liu, Y.; Yu, F.; Zhao, W. Not Just Anticoagulation-New and Old Applications of Heparin. *Molecules* **2022**, *27* (20), 6968.

(29) Perez, S.; Makshakova, O.; Angulo, J.; Bedini, E.; Bisio, A.; de Paz, J. L.; Fadda, E.; Guerrini, M.; Hricovini, M.; Hricovini, M.; Lisacek, F.; Nieto, P. M.; Pagel, K.; Paiardi, G.; Richter, R.; Samsonov, S. A.; Vives, R. R.; Nikitovic, D.; Ricard Blum, S. Glycosaminoglycans: What Remains To Be Deciphered? *JACS Au* **2023**, *3*, 628–656.

(30) Anila, S.; Samsonov, S. A. Benchmarking Water Models in Molecular Dynamics of Protein-Glycosaminoglycan Complexes. *J. Chem. Inf. Model.* **2024**, *64*, 1691–1703.

(31) Bathe, M.; Rutledge, G. C.; Grodzinsky, A. J.; Tidor, B. A Coarse-Grained Molecular Model for Glycosaminoglycans: Application to Chondroitin, Chondroitin Sulfate, and Hyaluronic Acid. *Biophys. J.* **2005**, *88*, 3870–3887.

(32) Sattelle, B. M.; Shakeri, J.; Cliff, M. J.; Almond, A. Proteoglycans and Their Heterogeneous Glycosaminoglycans at the Atomic Scale. *Biomacromolecules* **2015**, *16*, 951–961.

(33) Samsonov, S. A.; Bichmann, L.; Pisabarro, M. T. Coarse-Grained Model of Glycosaminoglycans. *J. Chem. Inf. Model.* **2015**, *55*, 114–124.

(34) Shivgan, A. T.; Marzinek, J. K.; Krah, A.; Matsudaira, P.; Verma, C. S.; Bond, P. J. Coarse-Grained Model of Glycosaminoglycans for Biomolecular Simulations. *J. Chem. Theory Comput.* **2024**, *20* (8), 3308–3321, DOI: 10.1021/acs.jctc.3c01088.

(35) Liwo, A.; Czaplewski, C.; Sieradzan, A. K.; Lubecka, E. A.; Lipska, A. G.; Golon, L.; Karczynska, A.; Krupa, P.; Mozolewska, M. A.; Makowski, M.; Ganzynkowicz, R.; Gieldon, A.; Maciejczyk, M. Scale-consistent approach to the derivation of coarse-grained force fields for simulating structure, dynamics, and thermodynamics of biopolymers. *Prog. Mol. Biol. Transl. Sci.* **2020**, *170*, 73–122.

(36) Engele, J.; Bohn, M. C. Effects of acidic and basic fibroblast growth factors (aFGF, bFGF) on glial precursor cell proliferation: age dependency and brain region specificity. *Dev. Biol.* **1992**, *152*, 363–372.

(37) Olwin, B. B.; Rapraeger, A. Repression of myogenic differentiation by aFGF, bFGF, and K-FGF is dependent on cellular heparan sulfate. *J. Cell Biol.* **1992**, *118*, 631–639.

(38) Schweigerer, L. Fibroblast growth factor and angiogenesis. *Z. Kardiol.* **1989**, *78 Suppl 6*, 12–15.

(39) Yang, H. S.; Bhang, S. H.; Hwang, J. W.; Kim, D.-I.; Kim, B.-S. Delivery of Basic Fibroblast Growth Factor Using Heparin-Conjugated Fibrin for Therapeutic Angiogenesis. *Tissue Eng., Part A* **2010**, *16*, 2113–2119.

(40) Coutu, D. L.; Galipeau, J. Roles of FGF signaling in stem cell self-renewal, senescence and aging. *Aging* **2011**, *3*, 920–933.

(41) Ohba, N.; Funatomi, H.; Seki, T.; Makino, R.; Mitamura, K. Hepatocyte growth factor stimulates cell growth and enhances the expression of transforming growth factor alpha mRNA in AsPC-1 human pancreatic cancer cells. *J. Gastroenterol.* **1999**, *34*, 498–504.

(42) Kiehne, K.; Herzig, K. H.; Fölsch, U. R. c-met expression in pancreatic cancer and effects of hepatocyte growth factor on pancreatic cancer cell growth. *Pancreas* **1997**, *15*, 35–40.

(43) Tajima, H.; Matsumoto, K.; Nakamura, T. Regulation of cell growth and motility by hepatocyte growth factor and receptor expression in various cell species. *Exp. Cell Res.* **1992**, *202*, 423–431.

(44) Kato, T.; Oka, K.; Nakamura, T.; Ito, A. Bronchioalveolar morphogenesis of human bronchial epithelial cells depending upon hepatocyte growth factor. *J. Cell. Mol. Med.* **2015**, *19*, 2818–2826.

(45) Anderson, R. M.; Delous, M.; Bosch, J. A.; Ye, L.; Robertson, M. A.; Hesselson, D.; Stainier, D. Y. R. Hepatocyte growth factor signaling in intrapancreatic ductal cells drives pancreatic morphogenesis. *PLoS Genet.* **2013**, *9*, No. e1003650.

(46) Felici, A.; Giubellino, A.; Bottaro, D. P. Gab1 mediates hepatocyte growth factor-stimulated mitogenicity and morphogenesis in multipotent myeloid cells. *J. Cell. Biochem.* **2010**, *111*, 310–321.

(47) Schwall, R. H.; Chang, L. Y.; Godowski, P. J.; Kahn, D. W.; Hillan, K. J.; Bauer, K. D.; Zioncheck, T. F. Heparin induces dimerization and confers proliferative activity onto the hepatocyte growth factor antagonists NK1 and NK2. *J. Cell Biol.* **1996**, *133*, 709–718.

(48) Chirgadze, D. Y.; Hepple, J. P.; Zhou, H.; Byrd, R. A.; Blundell, T. L.; Gherardi, E. Crystal structure of the NK1 fragment of HGF/SF suggests a novel mode for growth factor dimerization and receptor binding. *Nat. Struct. Biol.* **1999**, *6*, 72–79.

(49) Naka, D.; Ishii, T.; Shimomura, T.; Hishida, T.; Hara, H. Heparin modulates the receptor-binding and mitogenic activity of hepatocyte growth factor on hepatocytes. *Exp. Cell Res.* **1993**, *209*, 317–324.

(50) Youles, M.; Holmes, O.; Petoukhov, M. V.; Nessen, M. A.; Stivala, S.; Svergun, D. I.; Gherardi, E. Engineering the NK1 fragment of hepatocyte growth factor/scatter factor as a MET receptor antagonist. *J. Mol. Biol.* **2008**, *377*, 616–622.

(51) Lietha, D.; Chirgadze, D. Y.; Mulloy, B.; Blundell, T. L.; Gherardi, E. Crystal structures of NK1-heparin complexes reveal the basis for NK1 activity and enable engineering of potent agonists of the MET receptor. *EMBO J.* **2001**, *20*, 5543–5555.

(52) Ricard-Blum, S.; Beraud, M.; Raynal, N.; Farndale, R. W.; Ruggiero, F. Structural Requirements for Heparin/Heparan Sulfate Binding to Type V Collagen*. *J. Biol. Chem.* **2006**, *281*, 25195–25204.

(53) Bonnans, C.; Chou, J.; Werb, Z. Remodelling the extracellular matrix in development and disease. *Nat. Rev. Mol. Cell Biol.* **2014**, *15* (12), 786–801. Publisher: Nature Publishing Group.

(54) Suk, J. Y.; Zhang, F.; Balch, W. E.; Linhardt, R. J.; Kelly, J. W. Heparin Accelerates Gelsolin Amyloidogenesis. *Biochemistry* **2006**, *45*, 2234–2242.

(55) Egashira, M.; Takase, H.; Yamamoto, I.; Tanaka, M.; Saito, H. Identification of regions responsible for heparin-induced amyloidogenesis of human serum amyloid A using its fragment peptides. *Arch. Biochem. Biophys.* **2011**, *511*, 101–106.

(56) Liwo, A.; Baranowski, M.; Czaplewski, C.; et al. A unified coarse-grained model of biological macromolecules based on mean-field multipole-multipole interactions. *J. Mol. Model.* **2014**, *20* (8), No. 2306, DOI: 10.1007/s00894-014-2306-5.

(57) Lubecka, E. A.; Liwo, A. A general method for the derivation of the functional forms of the effective energy terms in coarse-grained energy functions of polymers. II. Backbone-local potentials of coarse-grained O1–4-bonded polyglucose chains. *J. Chem. Phys.* **2017**, *147*, No. 115101.

(58) Samsonov, S. A.; Lubecka, E. A.; Bojarski, K. K.; Ganzynkowicz, R.; Liwo, A. Local and long range potentials for heparin-protein systems for coarse-grained simulations. *Biopolymers* **2019**, *110*, No. e23269.

(59) Liwo, A.; Sieradzan, A. K.; Lipska, A. G.; Czaplewski, C.; Joung, I.; Zmudzinska, W.; Halabis, A.; Oldziej, S. A general method for the derivation of the functional forms of the effective energy terms in coarse-grained energy functions of polymers. III. Determination of scale-consistent backbone-local and correlation potentials in the UNRES force field and force-field calibration and validation. *J. Chem. Phys.* **2019**, *150*, No. 155104.

(60) Makowski, M.; Sobolewski, E.; Czaplewski, C.; Liwo, A.; Oldziej, S.; No, J. H.; Scheraga, H. A. Simple physics-based analytical formulas for the potentials of mean force for the interaction of amino acid side chains in water. 3. Calculation and parameterization of the

potentials of mean force of pairs of identical hydrophobic side chains. *J. Phys. Chem. B* **2007**, *111*, 2925–2931.

(61) Makowski, M.; Sobolewski, E.; Czaplewski, C.; Oldziej, S.; Liwo, A.; Scheraga, H. A. Simple physics-based analytical formulas for the potentials of mean force for the interaction of amino acid side chains in water. IV. Pairs of different hydrophobic side chains. *J. Phys. Chem. B* **2008**, *112*, 11385–11395.

(62) Makowski, M.; Liwo, A.; Scheraga, H. A. Simple physics-based analytical formulas for the potentials of mean force of the interaction of amino-acid side chains in water. VI. Oppositely charged side chains. *J. Phys. Chem. B* **2011**, *115*, 6130–6137.

(63) Makowski, M.; Liwo, A.; Scheraga, H. A. Simple Physics-Based Analytical Formulas for the Potentials of Mean Force of the Interaction of Amino Acid Side Chains in Water. VII. Charged-Hydrophobic/Polar and Polar-Hydrophobic/Polar Side Chains. *J. Phys. Chem. B* **2017**, *121*, 379–390.

(64) Liwo, A.; Khalili, M.; Czaplewski, C.; Kalinowski, S.; Oldziej, S.; Wachucik, K.; Scheraga, H. A. Modification and optimization of the united-residue (UNRES) potential-energy function for canonical simulations. I. Temperature dependence of the effective energy function and tests of the optimization method with single training proteins. *J. Phys. Chem. B* **2007**, *111*, 260–285.

(65) Krupa, P.; Halabis, A.; Zmudzinska, W.; Oldziej, S.; Scheraga, H. A.; Liwo, A. Maximum Likelihood Calibration of the UNRES Force Field for Simulation of Protein Structure and Dynamics. *J. Chem. Inf. Model.* **2017**, *57*, 2364–2377.

(66) Gay, J. G.; Berne, B. J. Modification of the overlap potential to mimic a linear site-site potential. *J. Chem. Phys.* **1981**, *74*, 3316–3319.

(67) Danielsson, A.; Samsonov, S. A.; Liwo, A.; Sieradzan, A. K. Extension of the SUGRES-1P Coarse-Grained Model of Polysaccharides to Heparin. *J. Chem. Theory Comput.* **2023**, *19*, 6023–6036.

(68) Chakraborty, D.; Mondal, B.; Thirumalai, D. Brewing COFFEE: A Sequence-Specific Coarse-Grained Energy Function for Simulations of DNA-Protein Complexes. *J. Chem. Theory Comput.* **2024**, *20*, 1398–1413. Publisher: American Chemical Society.

(69) Faham, S.; Hileman, R. E.; Fromm, J. R.; Linhardt, R. J.; Rees, D. C. Heparin Structure and Interactions with Basic Fibroblast Growth Factor. *Science* **1996**, *271*, 1116–1120.

(70) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(71) DiGabriele, A. D.; Lax, I.; Chen, D. I.; Svahn, C. M.; Jaye, M.; Schlessinger, J.; Hendrickson, W. A. Structure of a heparin-linked biologically active dimer of fibroblast growth factor. *Nature* **1998**, *393* (6687), 812–817.

(72) Case, D. A.; Betz, R. M.; Cerutti, D. S.; Cheatham, T. E., III; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N. AMBER; University of California: San Francisco, 2016; p 2016.

(73) Gay, D. M. Algorithm 611: Subroutines for unconstrained minimization using a model/trust-region approach. *ACM Trans. Math. Software* **1983**, *9*, 503–524.

(74) Khalili, M.; Liwo, A.; Rakowski, F.; Grochowski, P.; Scheraga, H. A. Molecular Dynamics with the United-Residue Model of Polypeptide Chains. I. Lagrange Equations of Motion and Tests of Numerical Stability in the Microcanonical Mode. *J. Phys. Chem. B* **2005**, *109*, 13785–13797.

(75) Van Rossum, G.; Drake, F. L. *Python 3 Reference Manual*; CreateSpace, 2009.

(76) Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining* Portland, Oregon, 1996; pp 226–231.

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

(78) The PyMOL Molecular Graphics System.

(79) Rotkiewicz, P.; Skolnick, J. Fast procedure for reconstruction of full-atom protein models from reduced representations. *J. Comput. Chem.* **2008**, *29*, 1460–1465.

(80) Hunter, J. D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95.

(81) Hulsen, T.; de Vlieg, J.; Alkema, W. BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics* **2008**, *9*, 488.

(82) Oliveros, J. C. Venny. An interactive tool for comparing lists with Venn's diagrams.

(83) Bakan, A.; Meireles, L. M.; Bahar, I. ProDy: Protein Dynamics Inferred from Theory and Experiments. *Bioinformatics* **2011**, *27*, 1575–1577.

(84) Zhang, Y.; Skolnick, J. Scoring function for automated assessment of protein structure template quality. *Proteins* **2004**, *57*, 702–710.

(85) Wang, Z.; Tegge, A. N.; Cheng, J. Evaluating the absolute quality of a single protein model using structural features and support vector machines. *Proteins* **2009**, *75*, 638–647.

(86) Xu, D.; Esko, J. D. Demystifying Heparan Sulfate-Protein Interactions. *Annu. Rev. Biochem.* **2014**, *83*, 129–157.

(87) Thompson, L. D.; Pantoliano, M. W.; Springer, B. A. Energetic Characterization of the Basic Fibroblast Growth Factor-Heparin Interaction: Identification of the Heparin Binding Domain. *Biochemistry* **1994**, *33*, 3831–3840.

(88) Venkataraman, G.; Sasisekharan, V.; Herr, A. B.; Ornitz, D. M.; Waksman, G.; Cooney, C. L.; Langer, R.; Sasisekharan, R. Preferential self-association of basic fibroblast growth factor is stabilized by heparin during receptor dimerization and activation. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 845–850.

(89) Prestrelski, S. J.; Fox, G. M.; Arakawa, T. Binding of heparin to basic fibroblast growth factor induces a conformational change. *Arch. Biochem. Biophys.* **1992**, *293*, 314–319.

(90) Roghani, M.; Mansukhani, A.; Dell'Era, P.; Bellosta, P.; Basilico, C.; Rifkin, D. B.; Moscatelli, D. Heparin increases the affinity of basic fibroblast growth factor for its receptor but is not required for binding. *J. Biol. Chem.* **1994**, *269*, 3976–3984.

(91) Schlessinger, J.; Plotnikov, A. N.; Ibrahimi, O. A.; Eliseenkova, A. V.; Yeh, B. K.; Yayon, A.; Linhardt, R. J.; Mohammadi, M. Crystal structure of a ternary FGF-FGFR-heparin complex reveals a dual role for heparin in FGFR binding and dimerization. *Mol. Cell* **2000**, *6*, 743–750.

(92) Ornitz, D. M.; Yayon, A.; Flanagan, J. G.; Svahn, C. M.; 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*, 240–247.

(93) Arakawa, T.; Wen, J.; Philo, J. S. Stoichiometry of Heparin Binding to Basic Fibroblast Growth Factor. *Arch. Biochem. Biophys.* **1994**, *308*, 267–273.

(94) Moy, F. J.; Safran, M.; Seddon, A. P.; Kitchen, D.; Böhlen, P.; Aviezer, D.; Yayon, A.; Powers, R. Properly oriented heparin-decasaccharide-induced dimers are the biologically active form of basic fibroblast growth factor. *Biochemistry* **1997**, *36*, 4782–4791.

(95) Delehedde, M.; Lyon, M.; Gallagher, J. T.; Rudland, P. S.; Fernig, D. G. Fibroblast growth factor-2 binds to small heparin-derived oligosaccharides and stimulates a sustained phosphorylation of p42/44 mitogen-activated protein kinase and proliferation of rat mammary fibroblasts. *Biochem. J.* **2002**, *366*, 235–244.

(96) Guglieri, S.; Hricovini, M.; Raman, R.; Polito, L.; Torri, G.; Casu, B.; Sasisekharan, R.; Guerrini, M. Minimum FGF2 binding structural requirements of heparin and heparan sulfate oligosaccharides as determined by NMR spectroscopy. *Biochemistry* **2008**, *47*, 13862–13869.

(97) Robinson, C. J.; Harmer, N. J.; Goodger, S. J.; Blundell, T. L.; Gallagher, J. T. Cooperative dimerization of fibroblast growth factor 1 (FGF1) upon a single heparin saccharide may drive the formation of 2:2:1 FGF1.FGFR2c.heparin ternary complexes. *J. Biol. Chem.* **2005**, *280*, 42274–42282.

(98) Zhang, F.; Zheng, L.; Cheng, S.; Peng, Y.; Fu, L.; Zhang, X.; Linhardt, R. J. Comparison of the Interactions of Different Growth Factors and Glycosaminoglycans. *Molecules* **2019**, *24*, 3360.

(99) Fromm, J. R.; Hileman, R. E.; Caldwell, E. E.; Weiler, J. M.; Linhardt, R. J. Differences in the interaction of heparin with arginine and lysine and the importance of these basic amino acids in the binding of heparin to acidic fibroblast growth factor. *Arch. Biochem. Biophys.* **1995**, *323*, 279–287.

(100) Ogura, K.; Nagata, K.; Hatanaka, H.; Habuchi, H.; Kimata, K.; Tate, S.-i.; Ravera, M. W.; Jaye, M.; Schlessinger, J.; Inagaki, F. Solution structure of human acidic fibroblast growth factor and interaction with heparin-derived hexasaccharide. *J. Biomol. NMR* **1999**, *13*, 11–24.

(101) Kemp, L. E.; Mulloy, B.; Gherardi, E. Signalling by HGF/SF and Met: the role of heparan sulphate co-receptors. *Biochem. Soc. Trans.* **2006**, *34*, 414–417.

(102) Zhou, H.; Casas-Finet, J. R.; Heath Coats, R.; Kaufman, J. D.; Stahl, S. J.; Wingfield, P. T.; Rubin, J. S.; Bottaro, D. P.; Byrd, R. A. Identification and dynamics of a heparin-binding site in hepatocyte growth factor. *Biochemistry* **1999**, *38*, 14793–14802.

(103) Sakata, H.; Stahl, S. J.; Taylor, W. G.; Rosenberg, J. M.; Sakaguchi, K.; Wingfield, P. T.; Rubin, J. S. Heparin binding and oligomerization of hepatocyte growth factor/scatter factor isoforms. Heparan sulfate glycosaminoglycan requirement for Met binding and signaling. *J. Biol. Chem.* **1997**, *272*, 9457–9463.



CAS BIOFINDER DISCOVERY PLATFORM™

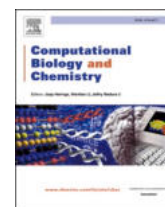
**CAS BIOFINDER
HELPS YOU FIND
YOUR NEXT
BREAKTHROUGH
FASTER**

Navigate pathways, targets, and
diseases with precision

Explore CAS BioFinder



Publication 5. Molecular dynamics-based
descriptors of 3-O-Sulfated Heparan
sulfate as contributors of protein binding
specificity.



Molecular dynamics-based descriptors of 3-O-Sulfated Heparan sulfate as contributors of protein binding specificity

Annemarie Danielsson^a, Małgorzata M. Kogut^{a,b}, Martyna Maszota-Zieleniak^a,
Pradeep Chopra^c, Geert-Jan Boons^{c,d,e}, Sergey A. Samsonov^{a,*}

^a Faculty of Chemistry, University of Gdańsk, ul. Wita Stwosza 63, 80-308 Gdańsk, Poland

^b Intercollegiate Faculty of Biotechnology of University of Gdańsk and Medical University of Gdańsk, ul. Abrahama 58, 80-307 Gdańsk, Poland

^c Complex Carbohydrate Research Center, University of Georgia, Athens, GA 30602, USA

^d Department of Chemistry, University of Georgia, Athens, GA 30602, USA

^e Department of Chemical Biology and Drug Discovery, Utrecht Institute for Pharmaceutical Sciences and Bijvoet Center for Biomolecular Research, Utrecht University, 3584 CG Utrecht, the Netherlands

ARTICLE INFO

Keywords:

Glycosaminoglycans
Sugar binding specificity
Molecular dynamics
Principal component analysis

ABSTRACT

Glycosaminoglycans are linear periodic and anionic polysaccharides found in the extracellular matrix, involved in a range of key biochemical processes as a result of their interactions with a variety of protein partners. Due to the template-less synthesis, high flexibility and charge of GAGs, as well as the multipose binding of GAG ligands to receptors, the specificity of GAG-protein interactions can be difficult to elucidate. In this study we propose a set of MD-based descriptors of unbound Heparan Sulfate hexasaccharides that can be used to characterize GAGs and explain their binding affinity to a set of protein receptors. With the help of experimental data on GAG-protein binding affinity, we were able to further characterize the nature of this interaction in addition to providing a basis for predictor functions of GAG-protein binding specificity.

1. Introduction

Glycosaminoglycans (GAGs) are a family of naturally-occurring linear periodic and anionic polysaccharides whose building blocks are repeating disaccharide units composed of an amino sugar and an uronic acid or hexose (Esko et al., 2009). Depending on their exact dimeric unit composition and their glycosidic linkages, several classes of GAGs can be distinguished: Heparan Sulfate (HS), Heparin (HP), Dermatan Sulfate (DS), Chondroitin Sulfate (CS), Keratan Sulfate (KS), and Hyaluronic Acid (HA). Heparan Sulfate (HS), composed of alternating N-acetylglucosamine (GlcNAc) and glucuronic acid (GlcA) residues (Esko et al., 2009), is found often covalently attached to proteins of the extracellular matrix and plasma membrane, forming proteoglycans (PGs). HS is known to be involved in key biological processes, including cell division (Ugby et al., 2019) and differentiation (Kraushaar et al., 2012; Patel et al., 2008; Yokoyama et al., 2020), angiogenesis (Zhang et al., 2014), coagulation (Ho et al., 1997), viral infection (Clausen et al., 2020; Yue

et al., 2021), neuron growth (Brickman et al., 1998; Johnson et al., 2007), as well as tumor proliferation (Hendriks et al., 2005) and metastasis (Qazi et al., 2016).

HS can undergo (selective) de-N-acetylation and N-sulfation of its GlcNAc residues, epimerization of GlcA to iduronic acid (IdoA), as well as sulfation by O-sulfotransferases on the 2-O position of its uronic acid (C2 of the uronic acid) and of the 6-O and 3-O positions of GlcNAc residues (Esko et al., 2009). While 3-O-sulfation is the rarest of all HS modifications, constituting only about 0.5% of the total sulfation (Huang et al., 2015; Pejler et al., 1987), 3-O-sulfotransferases (3-OSTs) represent the largest family of enzymes modifying HS polysaccharide chains (Thacker et al., 2014). This may indicate that 3-OSTs in cooperation with other HS-modifying enzymes produce HS molecules that are unique and therefore engage in highly-specific interactions with proteins (Chopra et al., 2021).

The specificity of GAG-protein binding has been studied extensively before for a variety of complexes (Gama et al., 2006; Gama and

Abbreviations: GAG, Glycosaminoglycan; HP, Heparin; HS, Heparan Sulfate; DS, Dermatan Sulfate; CS, Chondroitin Sulfate; KS, Keratan Sulfate; HA, Hyaluronic Acid; ATIII, Antithrombin III; PCA, Principal Component Analysis; PC, Principal Component; MD, Molecular Dynamics; RMSD, Root Mean Square Deviation; H-bond, Hydrogen Bond; HC, Hierarchical Clustering; RFU, Relative Fluorescent Unit.

* Corresponding author.

E-mail address: sergey.samsonov@ug.edu.pl (S.A. Samsonov).

<https://doi.org/10.1016/j.compbiolchem.2022.107716>

Received 14 April 2022; Received in revised form 3 June 2022; Accepted 20 June 2022

Available online 23 June 2022

1476-9271/© 2022 Elsevier Ltd. All rights reserved.

Hsieh-Wilson, 2005; Joseph et al., 2015; Künze et al., 2021; Liu et al., 2002; Pichert et al., 2012; Rogers et al., 2011; Sankaranarayanan et al., 2017; Schlorke et al., 2012). In some cases, the only parameter that affects the strength of such interactions is GAG charge, which makes them entirely electrostatically-driven and, therefore, unspecific (Hintze et al., 2014; Koehler et al., 2017; Panitz et al., 2016; Rother et al., 2016). However, a considerable number of GAG-protein complexes rely on specific interactions (Almond, 2018; Guerrini et al., 2008; Künze et al., 2021; Pomin and Mulloy, 2015; Sage et al., 2013). A number of proteins have been identified as receptors of 3-O-sulfated HS molecules participating in specific interactions, including Antithrombin III (Guerrini et al., 2008), Neuropilin 1 (Thacker et al., 2016), Stabilin 2 (Pempe et al., 2012), Advanced Glycosylation End-Product Receptor (Thacker et al., 2014) and growth factors such as Fibroblast Growth Factor 7 (Luo et al., 2006). Anomalies of 3-O-Sulfation have been implicated in tumorigenesis (Denys and Allain, 2019), renal fibrosis (Ferrerias et al., 2019), and tauopathies relating to Alzheimer's Disease (Zhao et al., 2020). 3-O-sulfated HS molecules on cell surfaces also participate in viral infections (O'Donnell and Shukla, 2008). Taken together, the characteristics of HS and their interactions with proteins, as well as involvement in essential biological pathways render them promising and interesting targets in medicine.

Although significant research has been conducted on this topic, the template-less synthesis and resulting structural complexity of HS molecules complicates studies of their interaction with proteins. While research on GAG-protein specificity concentrates mostly on the GAG-binding sites of proteins (Jokiranta et al., 2005; Li et al., 2016; Morgan et al., 2015; Mosier et al., 2012; Multhaupt, 1994; Pratt and Church, 1992; Sarkar and Desai, 2015; Sun et al., 2001; Taylor et al., 1995; Witt and Lander, 1994), it has been also observed that the sequence (Irie et al., 2002; Raghuraman et al., 2010, 2006; Sankaranarayanan et al., 2015; Sankaranarayanan and Desai, 2014; Shworak et al., 1994), sulfation degree and pattern (Ashikari-Hada et al., 2009; Irie et al., 2002; Kinnunen et al., 1996; Nakato and Kimata, 2002; Patel et al., 2008; Pye et al., 2000; Stringer and Gallagher, 1997; Viviano et al., 2004), chain length (Patel et al., 2008; Viviano et al., 2004) and, to a lesser degree, the conformation (Guglier et al., 2008) of GAGs generally and HS specifically can contribute to binding specificity.

The aim of this study was to inspect and analyze the specificity of HS-protein interactions employing computational approaches. To this end, we applied molecular dynamics (MD) simulations as well as a standard Machine Learning Algorithm (MLA) to examine 27 3-, 6-, and 3,6-O-sulfated HS molecules characterized and studied by (Chopra et al., 2021), followed by linear regression and cluster analysis in order to assess the connection between characteristics of unbound HS molecules and the specificity of binding to their protein partners. The analyzed dataset of 27 HS molecules with clearly defined sequences and sulfation patterns presents an unique opportunity to gain insight regarding the influence of specific differences in sulfation modifications on the 3D structure and physico-chemical characteristics of these molecules during Molecular Dynamics simulations. Those, in turn, can be linked to the affinity between the analyzed HS molecules and a set of proteins known to bind HS with different specificity. The application of a simple and well-known unsupervised MLA, Principal Component Analysis (PCA), enabled us to characterize the unbound HS molecules individually, but also the interplay between the proposed molecular descriptors in the context of GAG-protein binding. The investigation of the behavior of unbound HS molecules during MD simulations at atomistic level, that is not accessible using most experimental approaches, was supported and complemented by information gained from experiments (Chopra et al., 2021). Taken together, both approaches helped to overcome the limitations of either of them when applied alone and allowed us to identify characteristics of unbound HS molecules that could be linked with confidence to differences in binding affinity for a set of HS-protein complexes. This knowledge may prove useful in drug design of HS molecules exhibiting traits in line with our findings. In particular it

could suggest modifications of the GAG sequence that would lead to the appropriate conformations of the unbound HS with a higher propensity of binding their protein partners. Hence, our data assist in enabling the design and synthesis of specific HS molecules for potential use in regenerative medicine, cancer treatment and the prevention and treatment of viral infections.

2. Materials and methods

2.1. Heparan sulfate and protein dataset

In our study, we considered the library of 27 HS hexasaccharides previously synthesized and analyzed by (Chopra et al., 2021). The synthetic HS molecules, constructed from 9 different backbone templates, differed in the modification of their central GlcNAc residue (3-O-Sulfation, 6-O-Sulfation, 3,6-O-Sulfation) (sequences detailed in Table 1). In order to conduct simulations via computational methods, the 27 HS molecules were built in the xLeap module of AMBER16 (Case et al., 2016) and described with the GLYCAM06 force field parameters (Kirschner et al., 2008). The two building blocks of the analyzed HS hexasaccharides are shown in Fig. 1. The exact placement of the sulfate groups along the GAG influences the strength of electrostatic interaction in GAG-protein complexes and represents a “sulfation code” assumed to regulate the specificity of interactions with proteins (Gama and Hsieh-Wilson, 2005).

The analyzed HS molecules were named 1A, 1B, 1C to 9A, 9B, 9C in accordance with the naming scheme in (Chopra et al., 2021). The names, detailed in Table 1, reflect the position of the sulfate group on the GlcNAc residue (“A”: 3,6-O-sulfation, “B”: 3-O-sulfation, “C”: 6-O-Sulfation) and the sequence of the 3 residues varying between the 27 oligosaccharides.

Binding affinity data between the 27 HS hexasaccharides and a set of nine functionally diverse proteins known to recognize 3-O-sulfation to varying degrees was taken from (Chopra et al., 2021). The nine proteins considered are: Antithrombin III (ATIII), Heparin Cofactor 2 (HC-II), Fibroblast Growth Factor 7 and 9 (FGF-7, FGF-9), Fibroblast Growth Factor Receptor I (FGFR-I), Neuropilin 1 (Nrp-1), Bone Morphogenic Protein 2 (BMP-2), Stabilin 2 (Stab-2), Advanced Glycosylation End-Product Receptor (RAGE). The interaction study between HS molecules and the nine proteins presented in (Chopra et al., 2021) is the first such study examining the interaction selectivity for the set of HS hexasaccharides of clearly defined sequences and specific sulfation patterns along the GAG. All nine proteins are known to interact with HS and are involved in key biochemical processes connected to cell-cell signalling, e.g. angiogenesis, immunity, axon guidance, embryonic development. The study by (Chopra et al., 2021) is thus a particularly exclusive dataset that can be studied in order to elucidate the variability in binding affinity and selectivity of specific HS molecules and a wide range of their protein partners.

2.2. Molecular dynamics

Molecular dynamics (MD) simulations of the 27 unbound HS hexasaccharides were carried out in AMBER16 (Case et al., 2016; Götz et al., 2012) with the GLYCAM06 (Kirschner et al., 2008) force field in order to obtain descriptors of the oligosaccharides with possible predictive power in relation to their binding affinity and specificity.

Each of the 27 molecules was solvated in a TIP3P octahedral periodic box with minimum distance between solute and box edge of 15.0 Å and neutralized with counterions (Na^+). For every hexasaccharide, two energy minimization steps were carried out (first 1.5×10^3 steepest descent cycles and 10^3 conjugate gradient cycles with harmonic force restraints on solute atoms, followed by 6×10^3 steepest descent cycles and 3×10^3 conjugate gradient cycles without restraints). Subsequently, the system was heated up to 300 K for 10 ps with harmonic force restraints of 100 kcal/mol Å⁻² on solute atoms, and equilibration for 50 ps

Table 1

Sequences of the 27 Heparan Sulfate molecules; differences in sequence between the HS types are marked in bold.

1A: GlcA–GlcNS6S–GlcA–GlcNS3S6S–GlcA–GlcNS6S	4A: GlcA–GlcNS6S–GlcA–GlcNS3S6S–IdoA–GlcNS6S	7A: GlcA–GlcNS6S–GlcA–GlcNS3S6S–IdoA2S–GlcNS6S
1B: GlcA–GlcNS6S–GlcA–GlcNS3S–GlcA–GlcNS6S	4B: GlcA–GlcNS6S–GlcA–GlcNS3S–IdoA–GlcNS6S	7B: GlcA–GlcNS6S–GlcA–GlcNS3S–IdoA2S–GlcNS6S
1C: GlcA–GlcNS6S–GlcA–GlcNS6S–GlcA–GlcNS6S	4C: GlcA–GlcNS6S–GlcA–GlcNS6S–IdoA–GlcNS6S	7C: GlcA–GlcNS6S–GlcA–GlcNS6S–IdoA2S–GlcNS6S
2A: GlcA–GlcNS6S–IdoA–GlcNS3S6S–GlcA–GlcNS6S	5A: GlcA–GlcNS6S–IdoA–GlcNS3S6S–IdoA–GlcNS6S	8A: GlcA–GlcNS6S–IdoA–GlcNS3S6S–IdoA2S–GlcNS6S
2B: GlcA–GlcNS6S–IdoA–GlcNS3S–GlcA–GlcNS6S	5B: GlcA–GlcNS6S–IdoA–GlcNS3S–IdoA–GlcNS6S	8B: GlcA–GlcNS6S–IdoA–GlcNS3S–IdoA2S–GlcNS6S
2C: GlcA–GlcNS6S–IdoA–GlcNS6S–GlcA–GlcNS6S	5C: GlcA–GlcNS6S–IdoA–GlcNS6S–IdoA–GlcNS6S	8C: GlcA–GlcNS6S–IdoA–GlcNS6S–IdoA2S–GlcNS6S
3A: GlcA–GlcNS6S–IdoA2S–GlcNS3S6S–GlcA–GlcNS6S	6A: GlcA–GlcNS6S–IdoA2S–GlcNS3S6S–IdoA–GlcNS6S	9A: GlcA–GlcNS6S–IdoA2S–GlcNS3S6S–IdoA2S–GlcNS6S
3B: GlcA–GlcNS6S–IdoA2S–GlcNS3S–GlcA–GlcNS6S	6B: GlcA–GlcNS6S–IdoA2S–GlcNS3S–IdoA–GlcNS6S	9B: GlcA–GlcNS6S–IdoA2S–GlcNS3S–IdoA2S–GlcNS6S
3C: GlcA–GlcNS6S–IdoA2S–GlcNS6S–GlcA–GlcNS6S	6C: GlcA–GlcNS6S–IdoA2S–GlcNS6S–IdoA–GlcNS6S	9C: GlcA–GlcNS6S–IdoA2S–GlcNS6S–IdoA2S–GlcNS6S

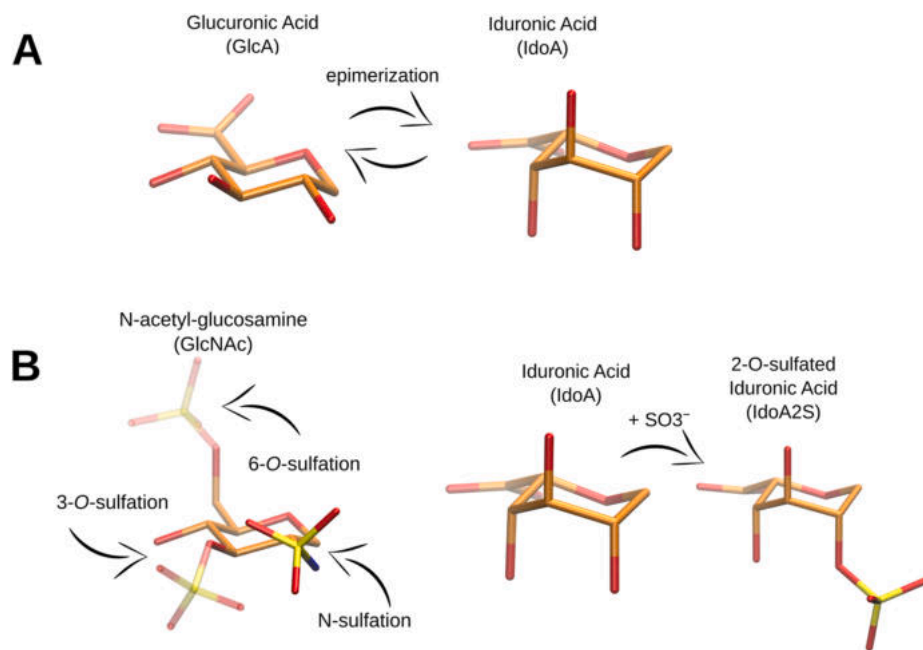


Fig. 1. The possible modifications of the HS building blocks in this study. All the HS molecules are made up of repetitive disaccharide units containing an uronic acid (glucuronic acid, GlcA or iduronic acid, IdoA) and N-acetyl-glucosamine, GlcNAc derivative. A) Glucuronic acid can undergo reversible epimerization to iduronic acid. B) Possible sulfation modifications of N-acetyl-glucosamine on its 3-O and 6-O positions (semi-transparent licorice representation) as well as the 2-O sulfation position of iduronic acid.

at 300 K and 10^5 Pa in isothermal isobaric ensemble (NPT). A 100 ns productive MD run was carried out in an NPT ensemble. The SHAKE algorithm (Ryckaert et al., 1977), 2 fs time integration step, 8 Å cutoff for non-bonded interactions, and the Particle Mesh Ewald method (Darden et al., 1993) were used.

2.3. Descriptors of unbound HS molecules

Based on the MD trajectories of each unbound HS, a set of physico-chemical properties (“descriptors”) was analyzed in the PTRAJ (Roe and Cheatham, 2013) module of AMBER16 using default parameters (Table 2). Restricting the descriptors to only those relating to intra-molecular characteristics of the HS was done with the goal to discover what properties of the unbound molecules are putatively linked to the binding affinity. In case those descriptors are identified, one can argue that while the synthesis of HS molecules in vivo is template-less, the cellular machinery is required to ensure certain characteristics and modifications of HS molecules in vivo in order for the molecules to be able to interact specifically with their protein partners. The inclusion of other descriptors to be obtained from MD simulation of HS complexes with proteins as well as descriptors of the HS-binding sites on the proteins should also be considered, as these would likely enhance the information gained from computational studies of these unbound HS oligosaccharides. Nevertheless, it can be assumed that the characteristics of unbound HS molecules are likely to affect the binding affinity to proteins in the extracellular matrix, as the long GAGs naturally occurring in cells would possibly have substantially limited flexibility in vivo to ensure correct complex formation with proteins (Sattelle et al., 2015; Spencer et al., 2010). Therefore, it is likely that certain characteristics of HS have to be already encoded and ensured in their unbound state, thus

likely to be observed in the preference of certain values of the physico-chemical descriptors analyzed in this study.

To detect groups of highly correlated descriptors, Pearson product-moment correlation coefficients were computed for the descriptor dataset and plotted as a heatmap. The correlation analysis was used as a guide in feature selection.

2.4. Machine learning and correlation analysis

The descriptor dataset was standardized prior to applying MLAs. The feature extraction and selection approaches were used to reduce the dimensionality of the original dataset and to uncover patterns in the descriptor dataset that could later be linked to the binding affinity of HS-protein complexes.

PCA of the descriptor dataset of unbound HS molecules was carried out in python 3.8.5 using the scikit-learn library (ver. 0.24.2) (Pedregosa et al., 2011). PCA is a type of unsupervised machine learning technique used in a variety of fields to reduce the dimensionality of large datasets as well as to reduce noise and extract patterns from the data. The approach works by transforming a matrix of correlated variables into a new coordinate system of uncorrelated variables called Principal Components (PCs) which are linear combinations of the initial ones and capture the essential information from the original dataset. We applied PCA to the descriptor dataset to create a lower-dimensional dataset of PCs that explain most of the variability in the MD-derived descriptors. Each PC can be seen as a summary of a set of variables that captures the physico-chemical character shared by the descriptors that contribute the most to the given PC. Because the construction of the PCs did not include any information on the HS-protein binding affinity, the obtained PCs only describe the character of the unbound HS oligosaccharides

Table 2
MD-derived descriptors of unbound HS molecules used in the study.

No.	Descriptor	Symbol (Units)	Explanation
1.–8.	Fraction of formed H-bonds	Hbond[0/1/2/3/4/5/6/7] (no unit)	Sum of fraction of MD-simulation frames where 0/1/2/3/4/5/6/7 intramolecular H-bonds were formed by atoms of the HS molecule
9.	Radius of gyration	R_gyration (Å)	A measure of elastic stability (resistance to deformation of shape and conformation) of the HS molecule; compactness of the molecule
10.	Dipole moment	Dipole (e * Å)	The dipole moment: a measure of the polarity of a molecule
11.	Molecule length	Length (Å)	Length of the molecule in Å defined as the distance between two terminal atoms
12.–19.	Dihedral angles of glycosidic linkages	glycosidic [1min1/1min2/3min1/3min2/5min1/5min1/2 min/4 min] (no unit)	The distribution of dihedral angles of each glycosidic linkages defined as O5 _{n+1} -C1 _{n+1} -O4 _n -C4 _n and C1 _{n+1} -O4 _n -C4 _n -C5 _n , where <i>n</i> is the sequential number of the sugar monomeric unit; Specifically, the percentage of points in the distribution of the dihedral angles that belong to a minimum
20.		glycosidic percent (%)	Total percentage of points of the dihedral angle distribution that belong to minima
21.–26.	Fluctuation of HS monosaccharide units	fluct [r1/r2/r3/r4/r5/r5] (Å)	Root mean square fluctuation analysis of atoms of HS residues
27.–28.	Ring pucker	pucker [1C4/4C1] (no unit)	Fraction of ¹ C ₄ and ⁴ C ₁ ring pucker conformations of the sugar monomeric units; The conformations are defined as: ¹ C ₄ : $\gamma = 48^\circ \pm 30^\circ$, $\delta = -64^\circ \pm 30^\circ$, ⁴ C ₁ : $\gamma = -54^\circ \pm 30^\circ$, $\delta = 62^\circ \pm 30^\circ$, where angle γ is defined by atoms C1–C2–C3–C4, and angle δ by atoms C1–O5–C5–C4
29.	Free energy components from the MM-GBSA model	VDW (kcal/mol)	Van der Waals energy
30.	Electrostatic energy	EEL (kcal/mol)	Electrostatic energy in vacuo
31.	EGB	EGB (kcal/mol)	Generalized Born electrostatic solvation energy
32.	ESURF	ESURF (kcal/mol)	Non-polar solvation energy
33.	Total free energy	TOTAL (kcal/mol)	Total free energy
34.	QH Entropy	S_APPROX (kcal/mol)	Quasi-Harmonic configurational entropy
35.	NM Entropy	S_nm (kcal/mol)	Normal mode entropy

simulated using MD.

As an alternative approach, Pearson correlation was used to assess the relationship between each individual descriptor and the HS-protein binding affinity. The values of each descriptor measured during the MD simulation were correlated with the binding affinity values for each complex individually to investigate which descriptors can be assumed to be most strongly linked with HS-protein binding strength and whether differences can be observed between complexes that can be attributed to the character and specificity of the binding. A descriptor was considered to be significantly associated with the binding affinity data if the *p*-value of the *t*-test for the correlation was below the $\alpha = 0.05$ significance level.

2.5. Linear regression analysis

Linear regression analysis was carried out in R (*stats* base package) with the purpose of determining the strength of the link between the computed PCs and HS-protein binding affinity data as given in (Chopra et al., 2021). To evaluate the usefulness of the PCA-based approach, for each HS-binding protein, the first four PCs of the dataset were used in a matching linear regression analysis.

Statistical significance and comparisons of regression models were conducted using the *t*-test and F-test. The cutoff for statistical significance used equals $\alpha = 0.05$; *p*-values of the *t*-test for the independent variables (PCs) below the α level were considered statistically significant. The F-tests were used to assess how well the regression models fit the data. R² (coefficient of determination) values were reported for each linear regression model to describe the amount of variance in the target variable (HS-binding affinity for each protein) accounted for by the independent variables.

2.6. Statistical and cluster analysis

Statistical analyses were carried out in R version 3.6.3 (R Core Team, 2020) (descriptive statistics of the descriptor dataset, Pearson correlation calculations using the “psych” package version 2.1.9 (Revelle, 2021)) as well as python 3.8.5 using numpy 1.19.5 (Harris et al., 2020) and pandas 1.2.2 (McKinney et al., 2010). Cluster analysis via

Hierarchical Clustering was carried out for the Pearson correlation coefficients between PCs (“scoring function”) and the binding affinity data in python using seaborn 0.11.1 (Waskom, 2021).

2.7. Visualization

Visualization was carried out in python using matplotlib 3.3.4 (Hunter, 2007) and seaborn 0.11.1 (Waskom, 2021).

3. Results

The descriptors of unbound HS hexasaccharides analyzed during the MD simulation are summarized in Supplementary Table 1.

3.1. MD simulation analysis

The properties of the 27 HS molecules throughout the MD simulation were analyzed and summarized in terms of the major descriptors. The evolution over time for Root Mean Square Deviation (RMSD) with respect to the starting conformation, radius of gyration and molecule length (end-to-end distance) are shown in Fig. 2 for molecules 1A, 1B, 1C, which were chosen as representatives for the analysis of these descriptors for two reasons. First, all the three different analyzed modes of sulfation (3,6-O-, 3-O-, and 6-O-sulfation) can be found in 1A, 1B and 1C, respectively. Moreover, the majority of the analyzed hexasaccharides behaved throughout the MD simulation in a manner qualitatively indistinguishable from these three HS molecules. Therefore, the evaluation whether the MD simulation reached convergence and the analysis of the descriptors over time is described in greater detail for 1A, 1B and 1C, as well as for any of the other molecules if their behavior over time during the MD was clearly different than for the majority of analyzed HS hexasaccharides. Supplementary Figs. 1–3 show the behavior over time for RMSD, radius of gyration, and end-to-end distance for all 27 simulated HS hexasaccharides. Upon inspecting the changes of descriptors over time for the 27 HS (Fig. 2 and Supplementary Figs. 1–3), it can be seen that the MD simulation converged, i.e. the simulated system reached its equilibrium state.

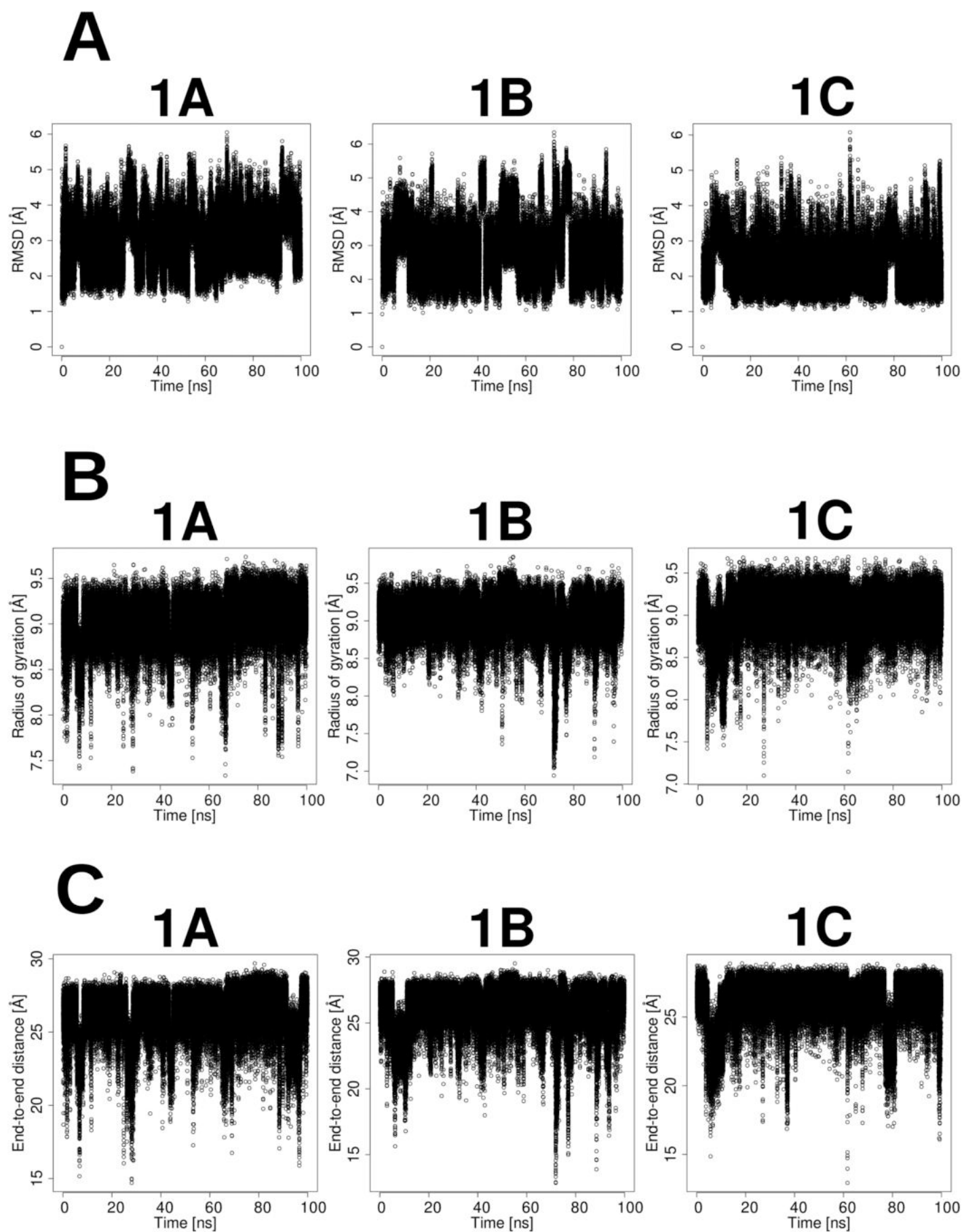


Fig. 2. A) Root Mean Square Deviation (RMSD) [Å] of HS molecules 1A, 1B, 1C calculated in reference to the starting position of the simulation. B) Radius of gyration [Å] of HS molecules 1A, 1B, 1C. C) Length of the simulated hexasaccharide, calculated as the end-to-end distance [Å], of HS molecules 1A, 1B, 1C. All the data are obtained from 100 ns of the MD simulation.

3.1.1. Root mean square deviation

RMSD analysis indicated that most of the HS molecules exhibit large fluctuations in the RMSD value across the MD simulation, as can be seen for molecules 1A, 1B, 1C in Fig. 2, chosen as representatives for the HS dataset. This indicates a high mobility of the unbound molecules. Exceptions can be observed for molecules 7B and 9A (Supplementary Fig. 1), which reach a stable state at a relatively early point of the MD simulation, as well as for 7C, for which RMSD reaches convergence towards the end of the simulation.

3.1.2. Radius of gyration

An analysis of the radius of gyration (R_{gyration}) of the 27 HS molecules shows trends in line with those observed for RMSD. As can be seen in the example of HS 1A, 1B, 1C (Fig. 2), the hexasaccharides are very flexible showing a high degree of mobility. As with RMSD analysis, molecule 9A, and to a lesser degree 7B and 7C (Supplementary Fig. 2) behave differently than the other molecules by exhibiting a decrease in R_{gyration} and thereby achieving a more stable and compact conformation towards the end of the MD simulation. 6C also exhibited a low radius of gyration for a significant portion of the MD simulation, but towards the end the values of the radius of gyration increased again, signifying a switching from an extended conformation in the beginning of the MD, to a more compact conformation in the middle of the simulation, and back to a more extended and flexible conformation towards the end.

3.1.3. HS length

The hexasaccharide molecules exhibit rapid changes of molecule length (measured as end-to-end distance) throughout the MD

simulation, with most molecules not converging to either an extended nor a compact state at the end of the simulation, as can be seen in Fig. 2 for the representative molecules 1A, 1B, 1C. However, HS 9A is an exception in the dataset, converging to a clearly compact conformation, with a significant decrease in the end-to-end distance seen throughout the simulation (Supplementary Fig. 3). Similarly, molecules 7C and 9C are seen to shorten towards the very end of the MD simulation.

3.1.4. Intramolecular H-bonds

Fig. 3 shows the distribution of frequencies of intramolecular hydrogen bonds (H-bonds) for the 27 HS molecules. For the vast majority of the analyzed molecules, the median value of H-bonds formed during the MD simulation equals 3 H-bonds, with some exceptions (median of 2 H-bonds for 7A, 8A, 9A). Conformations in which either no intramolecular H-bonds are formed during the MD simulation as well as those with the maximum possible amount of intramolecular H-bonds are unlikely to observe for all of the analyzed molecules. Nevertheless, the 3,6-sulfated 9A, 8A, and 7A hexasaccharides seem to have a preference of conformations with less intramolecular H-bonds compared to the other HS molecules.

3.1.5. Dipole moment

All of the molecules seem to be similar in terms of polarity, with values of the dipole moment measured across the MD simulation ranging from 420.6 e $\cdot$ Å (for 4B) to 615.6 e $\cdot$ Å (for 9C) (Supplementary Table 1). The most polar molecules are 9C (615.8 e $\cdot$ Å) and 9A (600.0 e $\cdot$ Å), substantially separated from the third-most polar HS hexasaccharide (3A with 562.4 e $\cdot$ Å) by over 30 e $\cdot$ Å units.

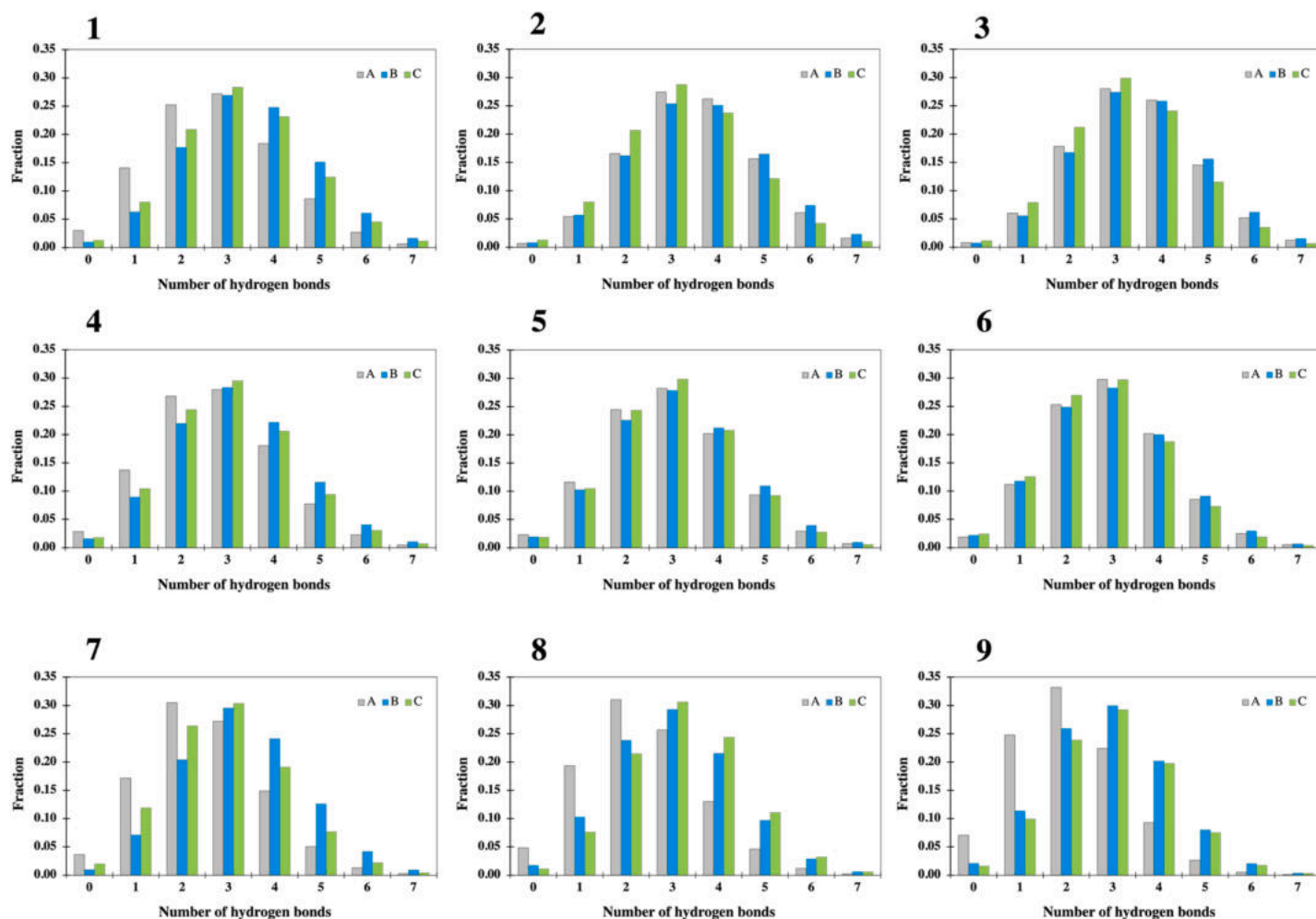


Fig. 3. Distribution of frequencies of intramolecular H-bonds formed by the HS molecules during MD simulations.

3.1.6. Glycosidic linkage dihedral angles

In terms of dihedral angles of the glycosidic linkages (Supplementary Table 1), the HS molecules do not seem to display significant differences between each other. The percentage of observations belonging to minima of the glycosidic linkage dihedral angles are comparable between the analyzed hexasaccharides.

3.1.7. Sugar ring fluctuation

For all of the analyzed HS molecules, the highest values of sugar ring fluctuations throughout the MD simulation were observed for the sugar ring of the first residue (Supplementary Table 1) and to a much lesser degree for the sugar ring of the last residue. Hexasaccharides 6C, 7C and 9A exhibit the highest fluctuation of their first sugar ring (8.8 Å, 8.3 Å and 8.8 Å, respectively, versus a range of 4.7–6.8 Å for the remaining HS molecules of the dataset). The same trend can be observed for the last sugar residue (fluctuation values of 4.6 Å for 6C, 3.8 Å for 7C and 5.5 Å for 9A, compared to a range of 2.0–3.6 Å for the rest of the hexasaccharides), but not for the 2nd, 3rd, 4th and 5th sugar rings.

3.1.8. Ring pucker conformation

A strong preference of the 4C_1 sugar ring pucker conformation can be seen for all 27 HS hexasaccharides (Supplementary Table 1). Among the 27 molecules, sugar rings of 1A, 6A and 7B show the highest propensity for the 1C_4 conformation, with a fraction of MD time in which such a conformation could be observed equal to 0.1, 0.19 and 0.11, respectively.

3.1.9. Enthalpy and entropy

In terms of enthalpy and entropy components of the unbound HS molecules throughout the MD simulation, most of the analyzed hexasaccharides have comparable values (Supplementary Table 1). A stark difference can be only observed for the electrostatic energy in vacuo for sugar 9A in comparison to the other HS molecules (1568 kcal/mol, versus a range of 356 kcal/mol to 1065 kcal/mol for the remaining sugars), indicating highly unfavorable electrostatic energy values of the unbound 9A hexasaccharide in solvent environment.

In summary, the 3,6-O-Sulfated HS hexasaccharide 9A, where the central sugar residue is flanked by IdoA2S residues on both sides, shows behavior clearly different than the other sugar hexasaccharides throughout the MD simulation, especially in terms of its RMSD calculated in reference to the starting conformation, frequency of intramolecular H-bonds, radius of gyration, fluctuations of the first and sixth sugar rings and electrostatic energy in vacuo. Noteworthy, sugar 9A was found to bind with high affinity to all proteins except for ATIII (Chopra et al., 2021).

3.2. Descriptor analysis

Pairwise Pearson correlation coefficients were calculated for the 35 descriptors summarized in Table 2, in order to examine the relationships between the descriptors. The correlation matrix visualized as a heatmap (Fig. 4) shows that some descriptors form clusters of strong positive and negative correlation, which may point to mutual relationships/dependencies and therefore possible redundancy in the primary descriptor dataset. A strong positive correlation (high positive Pearson correlation coefficient) is depicted by a red color, a strong negative correlation (negative Pearson correlation coefficient) – by violet, while no correlation corresponds to green colors on Fig. 4.

H-bond-related descriptors form three clusters based on the correlation coefficients: low amount of or no intramolecular H-bonds (Hbond0, Hbond1, Hbond2), high amount of intramolecular H-bonds (Hbond4, Hbond5, Hbond6, Hbond7) and exactly 3 H-bonds formed during the MD simulation (Hbond3). These descriptors correspond to the fraction of MD trajectory frames in which the simulated HS molecule formed the given amount of H-bonds. Therefore, the distribution of those frequencies, shown in Fig. 3, hints at preferences of certain HS

types regarding the formation of intramolecular H-bonds. While the most frequent amount of H-bonds for the majority of HS molecules was 3 H-bonds, hexasaccharides 7A, 8A and 9A are characterized by right-skewed distributions with the mode equal to 2 intramolecular H-bonds. The differences in sequence and sulfation pattern between the HS molecules dictates the preference for H-bond formation within the molecule. This information can in turn shed light on the formation of intermolecular H-bonds between the given HS and proteins, because the preference for forming many intramolecular H-bonds may indicate also a high likelihood of forming intermolecular H-bonds upon binding to proteins.

The closely related descriptors R_gyration and Length, corresponding to the mean radius of gyration of the HS molecule and its end-to-end distance, respectively, show a strong positive correlation to each other. The more compact a molecule is and thus the smaller its radius of gyration, the smaller is its end-to-end distance. Conversely, an extended molecule will display a higher radius of gyration and a greater length. Interestingly, among all other descriptors, both R_gyration and Length are strongly negatively correlated to glycosidic_3min2. A stiff hexasaccharide structure, i.e. only a narrow range of values possible for glycosidic bond dihedral angles, would influence the level of compactness of a molecule. The negative correlation coefficient value therefore links a preference of lower dihedral angle values with a less compact structure. Among the 27 HS hexasaccharides, 6C, 7B and 9A display a more compact structure during the MD simulations (lower radius of gyration and shorter end-to-end distance) compared to the other oligosaccharides.

All descriptors of sugar ring fluctuation are positively correlated to each other, however the fluctuation of the second sugar ring (fluct_r2) only shows a strong correlation to the first sugar ring and a weak correlation to the other fluctuation descriptors. For all studied HS molecules, the first sugar ring shows the highest fluctuations, with the 2nd, 3rd, 4th, 5th and 6th rings being considerably less mobile throughout the MD simulation. Molecules 6C, 7B, 7C and 9A have the most fluctuating 1st sugar residue among the HS hexasaccharides and this trend, although weaker, is also seen for residues 3, 4, 5 and 6. However, the 2nd sugar residue for 6C, 7B, 7C, 9A shows fluctuation levels comparable to the other HS molecules. Therefore, the distinction between fluct_r2 and the remaining fluctuation descriptors can be attributed to the behavior of those four HS hexasaccharides. At the same time, all fluctuation parameters are strongly correlated to R_gyration and Length (negative correlation) as well as to the second dihedral angle minimum of the third glycosidic bond (glycosidic_3min2, positive correlation). The fluctuation of a molecule is strongly linked to the molecule's compactness as well as the rigidity of its structure, with less fluctuation corresponding to an extended structure with a higher radius of gyration.

All of the 27 HS show a strong preference for the 4C_1 geometry. Ring pucker descriptors (pucker_1C4 and pucker_4C1) are strongly anti-correlated to each other and similarly correlated to most other descriptors. However, while pucker_1C4 is positively correlated to the first dihedral angle minimum of the first glycosidic linkage (glycosidic_1min1), pucker_4C1 shows a negative correlation to this descriptor.

The descriptors of the glycosidic bond dihedral angle minima do not form clusters as distinct and clear as the other descriptors. Nevertheless, some of the glycosidic linkage descriptors can be seen to be strongly correlated to certain other descriptors in the dataset, e.g. the strong negative correlation between the second minimum of the 3rd linkage and R_gyration and Length, as well as the aforementioned positive correlation to fluctuation descriptors.

The enthalpy and entropy terms, as well as the dipole moment descriptor establish two separate clusters. Descriptors relating to the dipole moment, electrostatic energy, the non-polar solvation energy, the total free energy and the normal mode-calculated entropy are positively correlated to each other (correlation coefficient above 0.78). At the same time, the van der Waals energy term, the polar solvation energy term,

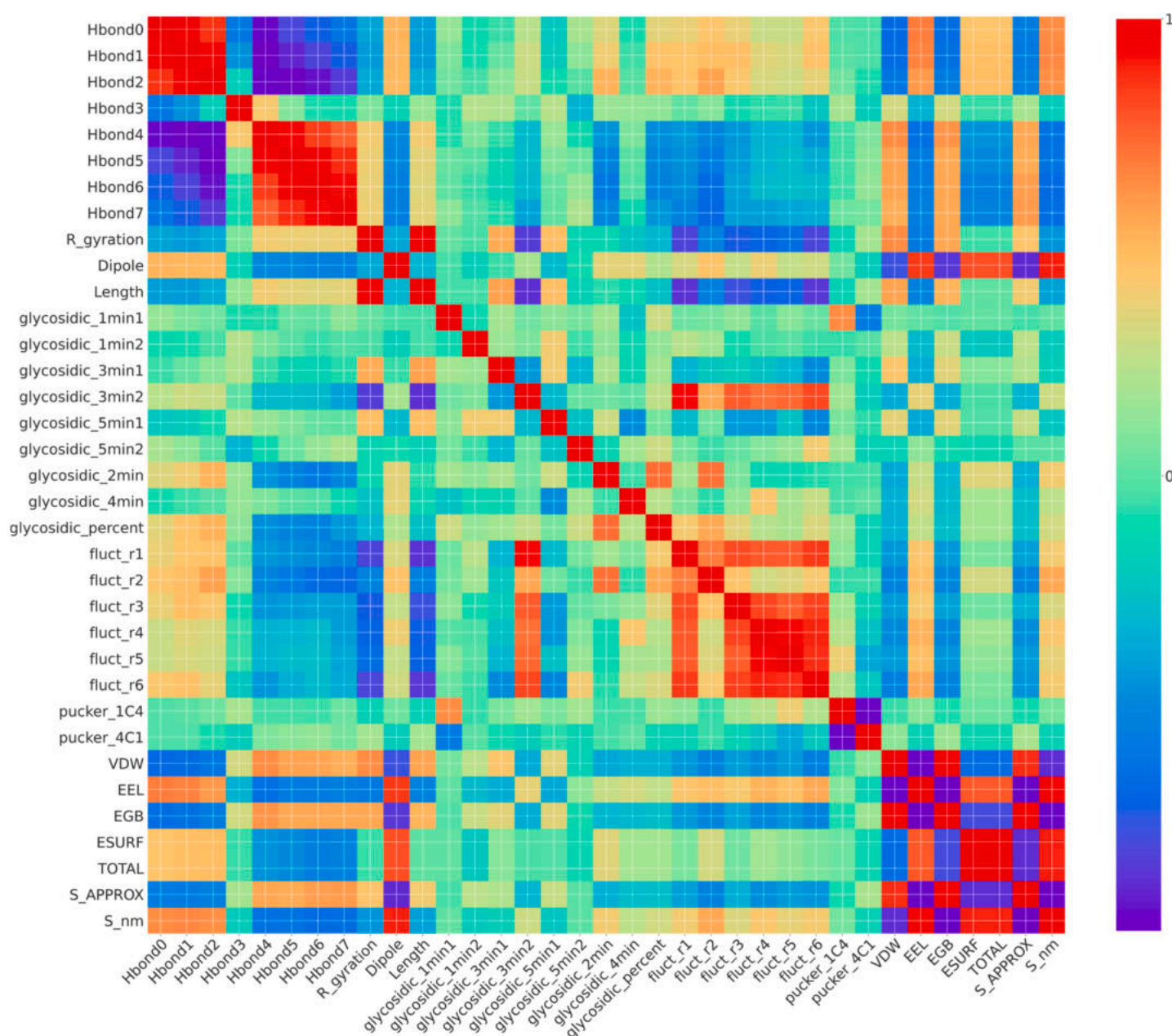


Fig. 4. Heatmap of Pearson correlation coefficients between the analyzed HS descriptors obtained from MD simulations.

and the quasi-harmonic entropy form another group of highly-correlated descriptors (correlation coefficients above 0.87). The negative correlation of the two entropy descriptors can seem surprising. What is important to consider for these descriptions of entropy, in this study they correspond to unbound ligand molecules and not, as may usually be the case when calculating entropy, to the protein-ligand complex formation. This negative correlation can be rooted in the difference in approaches used to calculate the entropy of a system. Quasi-harmonic entropy, corresponding to descriptor S_{approx} , is better at describing the translational and rotational changes in a molecule's conformation, closely related to the lengths, angles and torsions of bonds within the molecule. This approach also implicitly includes the effect of solvent on the entropy of the analyzed system. Normal mode entropy calculations, corresponding to S_{nm} in the descriptor data set, is linked to vibrational entropy, i.e. the bending and stretching of the molecule. Hence, both approaches describe different characteristics of the HS hexasaccharides and do not have to necessarily be positively correlated. The two groups formed by descriptors of enthalpy and entropy can be roughly classified as a group of "electrostatics" descriptors and a group of "shape" or "conformation" descriptors. The electrostatic

energy part of the total energy of the unbound HS molecules is the highest for 6A, 8A and 9A. The same oligosaccharides also exhibit very low van der Waals interaction energy, polar energy and quasi-harmonic entropy values. This could mean that the relatively high (i.e. unfavorable) electrostatic energy caused by the conformation of those molecules is balanced out by the favorable shape adapted by the molecules during the MD simulation.

The MD analysis of the 27 HS molecules provides invaluable information about the conformational landscape of those GAG molecules, uncovering preferences among the different HS types that may ultimately influence the affinity and specificity of binding to proteins.

3.3. Correlation analysis using Pearson product-moment correlation

The correlation between each of the 35 descriptors and the binding affinity data for 9 HS-protein complexes was assessed using Pearson product-moment correlation. [Supplementary Table 2](#) contains the correlation coefficient values for all the descriptors. One immediate observation is that for different proteins, particular descriptors are correlated with binding affinity of respective HS-protein complexes.

ATIII is the most distinct from all other considered proteins – only a small number of descriptors can be found to be linked to the binding affinity: the fluctuation of the second sugar ring and the dihedral angle descriptor for the 4th glycosidic linkage.

3.3.1. Intramolecular H-bonds

H-bond descriptors do not seem to be strongly correlated with the binding affinity of the remaining proteins, except for BMP2 (strong correlation for Hbond0 and Hbond3), FGFR1 (strong correlation for Hbond0, Hbond1, Hbond3) and Stab-2 (strong correlation for Hbond0, Hbond1, Hbond3).

3.3.2. Radius of gyration

The radius of gyration and length of the HS molecules are similarly linked only to the binding affinity for complexes containing BMP2, FGFR1 and Stab-2. The dipole moment seems to be strongly correlated for all proteins except for ATIII.

3.3.3. Glycosidic linkage dihedral angles

In the case of glycosidic linkage dihedral angle descriptors, the descriptor relating to the second minimum of the first glycosidic linkage is strongly negatively correlated only with HCII and FGF-7. The descriptor of the first minimum of the 3rd glycosidic linkage as well as the first minimum of the 5th glycosidic linkage exhibit strong anti-correlation with the binding affinity of all protein complexes except for ATIII. The descriptor of the dihedral angle of the 4th glycosidic linkage positively correlated with complexes containing HCII, BMP2, Nrp1, and negatively correlated with ATIII.

3.3.4. Sugar ring fluctuation

The fluctuation of the sugar residues is linked to the binding affinities only in the cases of: the second sugar ring and ATIII, the fourth sugar ring and BMP2, FGFR1, Stab-2, Nrp1, the fifth sugar ring and FGFR1, Stab-2, and the sixth sugar ring and BMP2, FGFR1, Stab-2, Nrp1.

3.3.5. Enthalpy and entropy

The enthalpy and entropy strongly correlated for all proteins except for ATIII as well as in the case of the solvation descriptor (ESURF) and the total energy descriptor (TOTAL) and proteins FGFR1 and Stab-2.

Taken together, this analysis indicates that there is no clear unique pattern of correlation across the complexes, with each exhibiting differences in correlation sign and value. Therefore, no one elegant analytical function, linking the descriptors of intramolecular physico-chemical properties of the HS molecules with the binding affinity of complexes, could be constructed even for groups of HS-protein complexes. Instead, any further analysis incorporating these correlation values would necessarily have to be tailored to each HS-protein complex.

3.4. Principal component analysis

Based on the conclusion drawn in the previous section, PCA seems, therefore, to be a appropriate next step for the investigation of the descriptors set. This alternative approach to the calculation of correlation coefficients for each pair of descriptor and HS-protein complex binding affinity is feature extraction, by which initial properties (here: the physico-chemical descriptors of HS hexasaccharides) are transformed into a set of new, independent descriptors (Principal Components, PCs). The constructed PCs are linear combinations of the original descriptors and as such combine information from the whole unreduced initial dataset.

The scree plot (Fig. 5) was generated in order to assess the optimal amount of PCs needed to explain as much of the variance in the descriptor dataset as possible while still offering a reduction in dimensionality. Since the goal of applying PCA was not to reduce the descriptor dataset for visualization in a 2D- nor 3D-space but rather to

find an overarching model describing the relationship between characteristics of the unbound HS hexasaccharides and their protein binding affinity, the amount of PCs for subsequent analysis was reduced to 4 PCs. It can be seen that the first 4 PCs combined explain 74% of the variance of the initial descriptor dataset, whereas adding the 5th PC would not increase that cumulative variance in a meaningful way.

To decipher the chemical nature of the constructed PCs, the contribution of each descriptor to the PCs was calculated and visualized by means of a heatmap (Fig. 6). The main contributors to the first PC (PC1) are H-bond descriptors (with the exception of Hbond3) and terms relating to enthalpy and entropy. PC2 is mostly influenced by R_gyration, Length, ring fluctuation (except fluct_r2), ESURF, enthalpy/entropy terms, and most of the glycosidic linkage dihedral angle descriptors. The dihedral angle descriptors are the main driving force behind PC3, whereas enthalpy, entropy and H-bond terms contribute less to this PC. The dominant HS properties governing PC4 are the ring pucker descriptors, the first minimum of the first glycosidic linkage dihedral angle and, to a lesser extent, Hbond3. The colors in Fig. 6 indicate the interplay between different descriptors and strength of contribution to each PC.

Some of the observed relationships are to be expected, e.g. the opposition of VDW to EEL – in cases where structural complementarity plays a greater role, the van der Waals interaction energy matters more than the electrostatic energy and vice versa. The antagonistic behavior of the low- and high-H-bond descriptors is also understandable and in accordance with previous observations. However, Hbond3 shows a surprising pattern: while for PC1 and PC2 it is not strongly tied to either of the two other groups of H-bond descriptors, in PC3 it follows the same trend as low-H-bond terms, while in PC4 it behaves more similar to high-H-bond descriptors. Another intriguing detail is the complete absence of fluct_r2 from PC2, despite the strong presence of all other fluctuation-related descriptors in that PC. Moreover, the two descriptors R_gyration and Length operate in the same direction in all descriptors except PC1, where they act in an opposing manner.

Additional information on the cooperation between different descriptors and PCs can be gained from inspecting the loading plots (Supplementary Fig. 4). In loading plots, the length of the vectors indicate the importance of the given descriptor for the PC. The orientation in space of the vectors depicts whether the descriptor has an influence on both, only one or none of the PCs considered in the plot.

Analysis of the interactions between descriptors in the PC-space

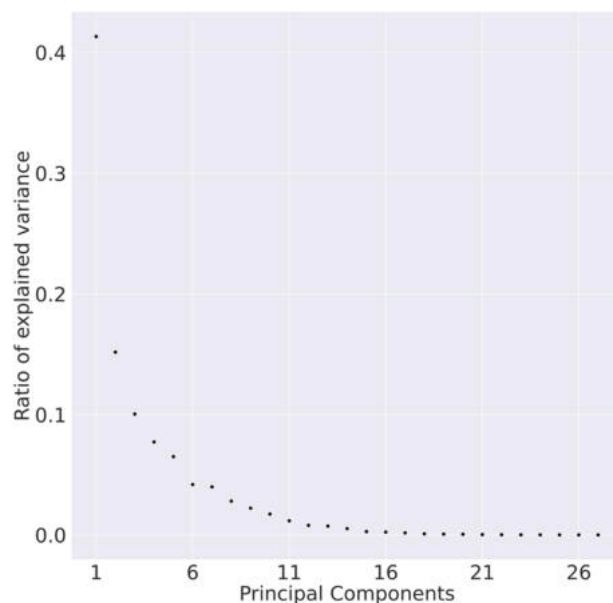


Fig. 5. Scree plot detailing the fraction of variance explained by the Principal Components.

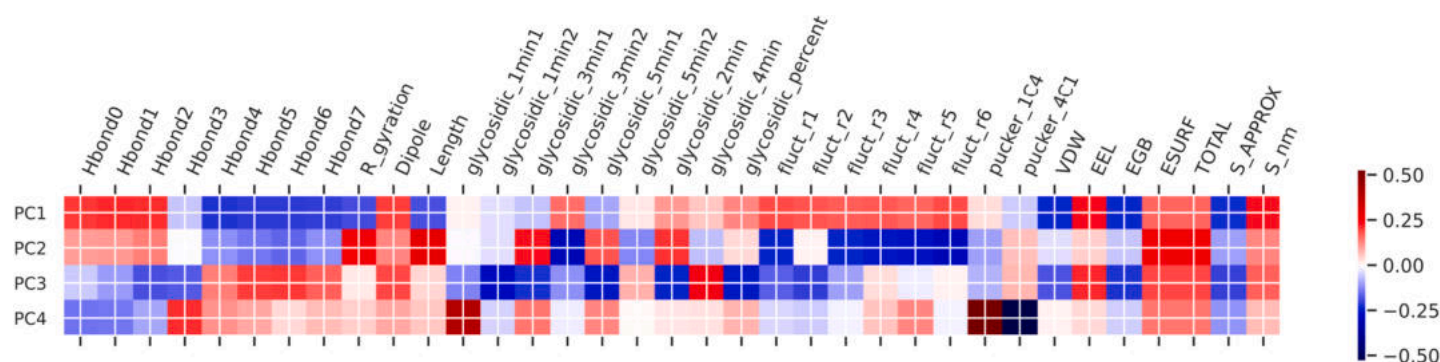


Fig. 6. Heatmap showing the contribution of descriptors of unbound HS molecules to the Principal Components.

illustrates a variety of coexisting conformations characterized by distinct values adopted by the analyzed descriptors. A low amount of intramolecular H-bonds formed, together with a less compact molecule (high radius of gyration and length of the HS molecule) coincides with high ring fluctuations, a larger dipole moment, high (unfavorable) electrostatic energy values, a greater entropy stemming from vibration and stretching of the molecule, as well as less entropy linked to the bending of the HS molecule. This corresponds to an extended, highly polar HS molecule; by analyzing the biplots of [Supplementary Fig. 4](#), it can be seen that oligosaccharides likely to be found in such conformations are e.g. 9A, 9C, 7A, 7C, 8A, 8B, 8C, 1A, 4A, 9B, 5A. A contrasting conformation would be characterized by a more compact structure, a higher amount of intramolecular H-bonds, less fluctuation of the sugar rings, low solvation energy, more entropy contributed by bending of the molecule and low electrostatic energy. The HS hexasaccharides that seem to correspond to this combination of characteristics are 1B, 1C, 2A, 2B, 2 C, 3B, 4B, 4C, 5B, 5C.

A different look on the conformation would consider the glycosidic linkages. A preference of the dihedral angles of the 4th glycosidic linkage for specific dihedral values (i.e. a high concentration of observations belonging to the dihedral angle minimum, described by glycosidic_4min) co-occurs with dihedral angle values spread out more for the 1st, 2nd, 3rd, 5th glycosidic linkages (less observations found in the minima during the MD simulations). This situation would be accompanied by very low fluctuations of the 1st and 2nd sugar rings, a high electrostatic energy, higher solvation energy, and a higher vibrational entropy (S_{nm}). HS hexasaccharides that could be linked to this conformation include 2A, 3A, 3B, 3C, 5A, 6A, 9A, 9B, 9C.

While ring pucker descriptors contribute mostly to PC4, which explains the least variation in the initial dataset compared to the other PCs, the information they carry is still valuable for describing the characteristics of some of the HS oligosaccharides. A strong preference for the 1C_4 sugar ring puckering is linked to a high concentration of dihedral bonds of the 1st glycosidic linkage that belong to the minimum dihedral angle values, as well as a high fraction of structures throughout the MD trajectory that display 3 intramolecular H-bonds. This arrangement of characteristics can be linked to oligosaccharides 1A, 1B, 6A, 7B. The opposite situation, corresponding to a preference for 4C_1 sugar ring puckering, can be seen to be characteristic especially for HS 9A.

When taking into account the HS-binding affinities determined by ([Chopra et al., 2021](#)), additional information can be gained about the preference of the HS-binding proteins regarding the characteristics of the hexasaccharides. As detailed in [Fig. 1](#) in ([Chopra et al., 2021](#)), ATIII shows a high affinity for 7A, 7B, a weaker affinity for 4A, 4B, and only minimal affinity for the other hexasaccharides. In the PC-space, unbound 7A and 7B exhibit low intramolecular H-bonding and high ring fluctuations, while differing in the ESURF and length of the molecules (7A is more extended than 7B) and ring puckering conformations (7A prefers 4C_1 , while 7B prefers 1C_4). Unbound 4B forms more intramolecular H-bonds than 7A and 7B. Both 4A and 4B are characterized by a less defined conformation of the glycosidic linkages (limited

preference of the minimum conformation) and less fluctuation of the sugar rings.

The other analyzed proteins show a clear preference for hexasaccharides 9A and 3A. According to our MD analysis, in its unbound form 9A is compact, forms little to no intramolecular H-bonds, exhibits high fluctuation of sugar ring atoms and shows a strong preference for 4C_1 ring puckering. At the same time, 3A is not characterized by considerable fluctuations of its sugar rings. Both unbound 9A and 3A show little preference of glycosidic linkage dihedral bond minima.

The obtained PCs provide a sort of model explaining the physico-chemical properties of the 27 unbound 3-O-sulfated HS hexasaccharides. Each of the HS molecules differs to some extent from the others in the PC-space which is significantly reduced in dimensionality relative to the original descriptor-space yet contains all relevant information on the chemical and structural characteristics of the studied GAGs. The observed differences may contribute to the understanding of HS-protein binding specificity and of the sulfation code.

3.5. Linear regression analysis

Linear regression was used to evaluate the extent to which descriptors of the unbound HS hexasaccharides can explain HS-protein binding affinity. The experimental binding affinity data between the 27 analyzed HS molecules and a set of protein targets known to specifically recognize 3-O-sulfation was taken from ([Chopra et al., 2021](#)). The independent variables of the linear regression model were the four PCs obtained from PCA, with the binding affinity value expressed in Relative Fluorescent Unit (RFU) as response variable. The results of this analysis are detailed in [Supplementary Table 3](#).

For all proteins except ATIII, PC1 and PC3 are consistently statistically significant ($p < \alpha = 0.05$). These PCs corresponded mostly to, in the case of PC1, H-bonds, enthalpy and entropy, as well as glycosidic linkage dihedral angles for PC3. Hence, it could be assumed that those sets of descriptors of unbound HS hexasaccharides are unlikely to have any substantial influence on the HS/ATIII binding affinity. While the other two PCs, PC2 and PC4, do not exhibit statistically significant links to the response variable, it cannot be ruled out that the descriptors these PCs summarize contribute to the binding affinity of the GAG-protein complexes. For all proteins except ATIII, the PC-based linear regression model offers a high degree of explanation of the target variable, with R^2 values ranging between 0.68 and 0.8 (adjusted R^2 between 0.62 and 0.77). Thus, the MD-derived descriptors of unbound 3-O-sulfated HS hexasaccharides, simulated by MD, are alone capable of explaining up to around 70% of the binding affinity data in 3-O-sulfated HS-protein interactions. In the case of ATIII, the R^2 equals 0.11 (adjusted $R^2 = -0.06$), which indicates that the PC-based linear regression models fails at explaining the variance in HS/ATIII binding affinity. Consequently, it can be deduced that the descriptors serving as main signals in the four PCs considered are not the driving force behind the binding affinity between ATIII and 3-O-sulfated HS.

3.6. Cluster analysis

In order to discover any patterns and subgroups in the binding affinity data in relation to the descriptor-based PC space, cluster analysis by means of Hierarchical Clustering (HC) with average linkage was conducted. The clusterization of the PC versus binding affinity correlation matrix is depicted as a heatmap in Fig. 7.

Three main groups of proteins could be identified by cluster analysis – “group 0” composed only of ATIII, “group 1” comprising BMP-2, FGFR-I and Stab-2, and “group 2” corresponding to RAGE, Nrp-1, HC-II, FGF-7, FGF-9. All proteins are positively-correlated with PC1 (the H-bond and free energy PC), however for ATIII this correlation is the weakest. In the case of PC3, ATIII shows negative correlation with all other proteins and a strong positive correlation with the values of PC3. The same pattern, although reversed in direction, can be observed for PC4, however here groups 1 and 2 differ visibly in the strength of correlation. In the case of PC2, only group 1 follows a clear trend as a whole, with all group 1 proteins negatively correlated to PC2. ATIII shows only a very weak positive correlation, while group 2 is more heterogeneous in its correlation to PC2 across its protein members, with some showing a weak positive and other a weak negative correlation.

Cluster analysis by HC revealed differences between the proteins in PC-space that may point to interesting information about the character of their binding with 3-*O*-sulfated HS. Among the descriptors considered, the binding between group 2-proteins and HS seems to be thus mostly linked to H-bond formation, enthalpy and entropy, as well as the glycosidic bond dihedral angle. Based on the HC analysis, HS/ATIII binding is influenced by, in order of importance, glycosidic bond dihedral angle descriptors, H-bonds and free energy components, ring puckering, and the least by sugar ring fluctuations. Meanwhile, the binding between group 1-proteins and 3-*O*-sulfated HS can be said to be

influenced by all four PCs considered.

Taken together with the information about contribution of each descriptor to the PCs, it is possible to further speculate on the exact character of the HS-protein binding. PC3 is strongly linked to descriptors of glycosidic linkage dihedral angle minima and for all those descriptors except glycosidic_4min the correlation to PC3 is negative. A positive correlation between the HS-binding affinity of a protein and PC3 consequently also means a negative correlation to the glycosidic linkage descriptors and a more spread out distribution of dihedral angle values sampled from the MD trajectories. Conversely, a negative correlation with PC3, as is the case for ATIII, means a positive correlation with the glycosidic bond descriptors, thus pointing towards a preference of a high fraction of data points in the minima and a specificity of dihedral angle values of the HS molecules sampled from the MD simulations.

Similar conclusions can be drawn about proteins of group 1 and sugar ring fluctuation and puckering. Both sugar ring puckering and fluctuation describe the flexibility of the HS molecule. The strong correlation of group 1-binding affinity with PC2 and PC4 indicates a clear preference of certain conformational states. A negative correlation with PC2 corresponds to a positive correlation with fluctuation descriptors, thus a greater link between binding affinity and high fluctuation and consequently a greater entropy of the HS molecule. At the same time, the negative correlation with PC4 entails a link between binding affinity and a tendency for the 4C_1 ring pucker conformation, thus less flexibility in terms of ring puckering. Conversely, no strong correlation with PC4, as is the case e.g. for some proteins of group 2, would indicate no strong tendency towards either pucker conformation and thus perhaps an increased flexibility of the HS molecules bound with high affinity. Interestingly, ATIII is known to have a preference of certain ring pucker conformations of its ligands (Demeter et al., 2018), while in our analysis only a weak correlation with PC4 could be observed (Fig. 7). However,

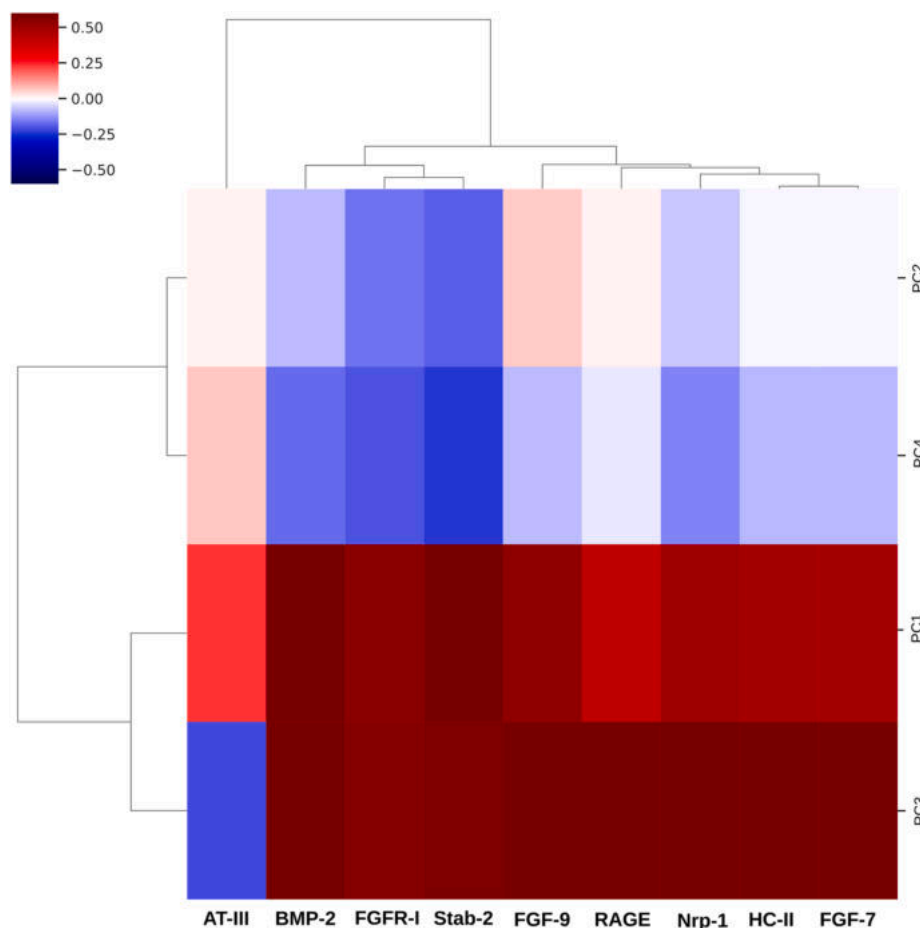


Fig. 7. Heatmap and hierarchical clustering of the correlation coefficients between PCs and HS-protein binding affinity values.

(Stancanelli et al., 2018) found that the ring pucker conformation of unbound HP is not obligatory for high-affinity binding, as the correct conformation is adopted by HP upon binding to ATIII. Therefore, a weak correlation with PC4 does not necessarily correspond to a lack of preference of sugar ring conformation.

3.7. Assessment of binding affinity based on the MD-derived Principal Components

In order to use the obtained PCs as a scoring function, adding descriptors of either the protein or specifically the HS-binding site would be necessary. However, knowledge of structure or function similarity, evolutionary relationships between proteins or co-localization of proteins in different sub-cellular locations, tissues or organs can also be used to form hypotheses on the interaction of proteins with ligands. We used the experimental binding affinity between Fibroblast Growth Factor 2 (FGF-2) and the 27 3-O-sulfated HS hexasaccharides (Chopra et al., 2021) to examine the usefulness of our PCA-based approach in predicting the approximate physico-chemical nature of the HS/FGF-2 binding. Due to being a member of the FGF family and interacting with Nrp-1 (Uniewicz and Fernig, 2008), we hypothesized that the binding of FGF-2 to the 27 HS molecules will be similar in character to the binding affinity of FGF-7, FGF-9 and Nrp-1. Therefore, FGF-2 would be assigned to group 2'' of HS-binding proteins by HC.

We calculated the mean binding affinity for each sub-group identified by cluster analysis (i.e. group 0, group 1, group 2) and calculated the Pearson product-moment correlation between the FGF-2 binding affinity and the mean binding affinities. The highest correlation was between FGF-2 and the mean binding affinity of group 2 (correlation coefficient equal to 0.82), followed by group 1 (correlation coefficient 0.78) and ATIII (group 0, correlation coefficient = -0.006). After including the FGF-2 data into the HC analysis, the clustering algorithm thus correctly sorted FGF-2 to group 2. As can be seen in Supplementary Fig. 5, the binding profile of FGF-2 is indeed similar to the binding profile of the other proteins of group 2: RAGE, Nrp-1, HC-II, FGF-7, FGF-9. Proteins of this group seem to lack the slightly higher selectivity of HS binding that can be observed for proteins of group 1. Therefore, the PCA approach, trained on data not including the FGF-2 binding affinity, was able to distill the information contained in the unbound HS molecules in such a way that enabled us to correctly link the experimental binding affinity with principal components containing the most relevant physico-chemical information on the unbound HS structures.

4. Conclusions

The analysis of the physico-chemical descriptors of unbound HS hexasaccharides using computational approaches gave insight into the specificity and nature of binding between HS and a set of diverse proteins as well as into understanding of the sulfation code. Differences in sequence dictated preferences of unbound HS molecules regarding shape and flexibility. The conformational landscape of unbound HS that arose from the cooperation of physico-chemical descriptors is likely an important factor in the determination of binding affinity and specificity in HS/protein binding.

The elucidation of the links between properties of the HS molecules to the HS/protein binding affinity determined by (Chopra et al., 2021) as well as removing redundancy from the available data was achieved by applying PCA to the initial descriptor dataset and identifying four principal components corresponding to the main groups of characteristics that explained most of the variance of the HS descriptors. The calculation of correlation between the obtained PCs and the HS-protein binding affinity values followed by hierarchical clustering of the correlation coefficients, as well as analysis of the association between PCs and binding affinity data clearly showed that information obtained from MD simulations of unbound HS molecules can offer insight into HS/protein binding. All analyses underlined the difference in

HS-binding specificity between ATIII and the other proteins, which is in accordance with multiple experimental and theoretical studies (e.g. Mosier et al., 2012; Raghuraman et al., 2010). Additionally, we were able to further divide the remaining proteins into two subsets depending on the correlation of their HS-binding affinity with PCs describing the unbound HS GAGs.

The linear regression analysis on the PCA and HS-binding affinity data showed that, except for ATIII, the binding of HS by the proteins could be explained to a high degree by the information summarized in the PCs. Therefore, PCA not only reduced the dimension of the initial descriptor dataset but also allowed us to identify the main characteristics of HS that can putatively contribute to protein binding affinity. However, depending on the HS/protein complex considered, a part of variance in binding affinity was left unaccounted for by the HS descriptors and PCs. This unexplained variance may be caused by the lack of descriptors of either the proteins or the GAG-binding sites located on those proteins in our study. While GAG-binding sites on proteins are usually composed of clusters of positively-charged amino-acid residues, (Sarkar and Desai, 2015) describe the importance of other types of residues in the binding site and its vicinity for the specificity of protein/GAG binding. (Joseph et al., 2015) identified residues of IL-8 outside of the GAG-binding site that do not directly bind the GAG but contribute to GAG-binding by enhancing the plasticity of the binding site in order to accommodate GAG molecules. Therefore, the differences between the proteins that can be reflected by other descriptors than those included in our study, not only restricted to the binding site size and amino-acid sequence, is likely to further improve our model. Additional factors known to influence GAG/protein binding affinity and specificity, which could contribute to our analysis in future research, is the presence of ions (Kogut et al., 2021; Multhaupt, 1994), the change in protein conformation upon GAG-binding (Rosenberg et al., 1997; Sage et al., 2013) or considering the presence of water molecules in the GAG-binding site of proteins (Mosier et al., 2012).

The use of *in silico* methods allowed us to study in detail the properties of 3-O-sulfated HS hexasaccharides in relation to their protein binding specificity. This shed light on relationships between structural properties of the HS molecules and protein binding affinity. A deeper understanding of HS/protein-binding specificity attributed to the sulfation code and of the physico-chemical characteristics of HS molecules that assist in the identification and design of specific GAG molecules and their mimetics with applications in medicine and pharmacology.

CRedit authorship contribution statement

Annemarie Danielsson: Conceptualization, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Małgorzata M. Kogut:** Formal analysis, Investigation, Visualization. **Martyna Maszota-Zieleniak:** Formal analysis, Investigation, Visualization. **Pradeep Chopra:** Data curation, Investigation. **Geert-Jan Boons:** Conceptualization, Supervision. **Sergey A. Samsonov:** Conceptualization, Investigation, Funding acquisition, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

All software used in this study is available for download free of charge from the websites of the respective developers: AMBER (<https://ambermd.org/>), Python (<https://www.python.org/>), R (<https://www.r-project.org/>). The trajectories obtained using MD

simulations as well as the results of PCA are available upon request. The binding affinity data was obtained from (Chopra et al., 2021).

Acknowledgments

This study was funded by the National Science Centre of Poland (Narodowe Centrum Nauki, grant number UMO-2018/31/G/ST4/00246) and the National Institute of Health (grant HLBI R01HL151617 for G.-J. B.) We thank dr. Margrethe Gaardl s for technical support.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.compbiolchem.2022.107716](https://doi.org/10.1016/j.compbiolchem.2022.107716).

References

- Almond, A., 2018. Multiscale modeling of glycosaminoglycan structure and dynamics: current methods and challenges. *Curr. Opin. Struct. Biol.* 50, 58–64. <https://doi.org/10.1016/j.sbi.2017.11.008>.
- Ashikari-Hada, S., Habuchi, H., Sugaya, N., Kobayashi, T., Kimata, K., 2009. Specific inhibition of FGF-2 signaling with 2-O-sulfated octasaccharides of heparan sulfate. *Glycobiology* 19, 644–654. <https://doi.org/10.1093/glycob/cwp031>.
- Brickman, Y.G., Ford, M.D., Gallagher, J.T., Nurcombe, V., Bartlett, P.F., Turnbull, J.E., 1998. Structural modification of fibroblast growth factor-binding heparan sulfate at a determinative stage of neural development. *J. Biol. Chem.* 273, 4350–4359. <https://doi.org/10.1074/jbc.273.8.4350>.
- Case, D.A., Betz, R.M., Cerutti, D.S., Cheatham III, T.E., Darden, T.A., Duke, R.E., Giese, T.J., Gohlke, H., Goetz, A.W., Homeyer, N., 2016. AMBER. University of California, San Francisco, p. 2016.
- Chopra, P., Joshi, A., Wu, J., Lu, W., Yadavalli, T., Wolfert, M.A., Shukla, D., Zaia, J., Boons, G.-J., 2021. The 3-O-sulfation of heparan sulfate modulates protein binding and lyase degradation. *Proc. Natl. Acad. Sci. U. S. A.* 118, e2012935118. <https://doi.org/10.1073/pnas.2012935118>.
- Clausen, T.M., Sandoval, D.R., Spliid, C.B., Pihl, J., Perrett, H.R., Painter, C.D., Narayanan, A., Majowicz, S.A., Kwong, E.M., McVicar, R.N., Thacker, B.E., Glass, C. A., Yang, Z., Torres, J.L., Golden, G.J., Bartels, P.L., Porell, R.N., Garretson, A.F., Laubach, L., Feldman, J., Yin, X., Pu, Y., Hauser, B.M., Caradonna, T.M., Kellman, B. P., Martino, C., Gordts, P.L.S.M., Chanda, S.K., Schmidt, A.G., Godula, K., Leibel, S. L., Jose, J., Corbett, K.D., Ward, A.B., Carlin, A.F., Esko, J.D., 2020. SARS-CoV-2 Infection Depends on Cellular Heparan Sulfate and ACE2. *e15 Cell* 183, 1043–1057. <https://doi.org/10.1016/j.cell.2020.09.033>.
- Darden, T., York, D., Pedersen, L., 1993. Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98, 10089–10092. <https://doi.org/10.1063/1.464397>.
- Demeter, F., Gy ngy si, T., Bereczky, Z., K v r, K.E., Herczeg, M., Borb s, A., 2018. Replacement of the L-iduronic acid unit of the anticoagulant pentasaccharide idraparinux by a 6-deoxy-L-talopyranose - Synthesis and conformational analysis. *Sci. Rep.* 8, 13736. <https://doi.org/10.1038/s41598-018-31854-z>.
- Denys, A., Allain, F., 2019. The Emerging Roles of Heparan Sulfate 3-O-Sulfotransferases in Cancer. *Front. Oncol.* 9, 507. <https://doi.org/10.3389/fonc.2019.00507>.
- Esko, J.D., Kimata, K., Lindahl, U., 2009. Proteoglycans and Sulfated Glycosaminoglycans. In: Varki, A., Cummings, R.D., Esko, J.D., Freeze, H.H., Stanley, P., Bertozzi, C.R., Hart, G.W., Etzler, M.E. (Eds.), *Essentials of Glycobiology*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY).
- Ferreras, L., Moles, A., Situmorang, G.R., El Masri, R., Wilson, I.L., Cooke, K., Thompson, E., Kusche-Gullberg, M., Vives, R.R., Sheerin, N.S., Ali, S., 2019. Heparan sulfate in chronic kidney diseases: Exploring the role of 3-O-sulfation. *Biochim. Biophys. Acta Gen. Subj.* 1863, 839–848. <https://doi.org/10.1016/j.bbagen.2019.02.009>.
- Gama, C.I., Hsieh-Wilson, L.C., 2005. Chemical approaches to deciphering the glycosaminoglycan code. *Curr. Opin. Chem. Biol.* 9, 609–619. <https://doi.org/10.1016/j.cbpa.2005.10.003>.
- Gama, C.I., Tully, S.E., Sotogaku, N., Clark, P.M., Rawat, M., Vaidehi, N., Goddard, W.A., Nishi, A., Hsieh-Wilson, L.C., 2006. Sulfation patterns of glycosaminoglycans encode molecular recognition and activity. *Nat. Chem. Biol.* 2, 467–473. <https://doi.org/10.1038/nchembio810>.
- G tz, A.W., Williamson, M.J., Xu, D., Poole, D., Le Grand, S., Walker, R.C., 2012. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* 8, 1542–1555. <https://doi.org/10.1021/ct200909j>.
- Guerrini, M., Guglieri, S., Casu, B., Torri, G., Mourier, P., Boudier, C., Viskov, C., 2008. Antithrombin-binding octasaccharides and role of extensions of the active pentasaccharide sequence in the specificity and strength of interaction. Evidence for very high affinity induced by an unusual glucuronic acid residue. *J. Biol. Chem.* 283, 26662–26675. <https://doi.org/10.1074/jbc.M801102200>.
- Guglieri, S., Hricovini, M., Raman, R., Polito, L., Torri, G., Casu, B., Sasisekharan, R., Guerrini, M., 2008. Minimum FGF2 binding structural requirements of heparin and heparan sulfate oligosaccharides as determined by NMR spectroscopy. *Biochemistry* 47, 13862–13869. <https://doi.org/10.1021/bi801007p>.
- Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M.H., Brett, M., Haldane, A., Del R  , J.F., Wiebe, M., Peterson, P., G rard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., Oliphant, T.E., 2020. Array programming with NumPy. *Nature* 585, 357–362. <https://doi.org/10.1038/s41586-020-2649-2>.
- Hendriks, J., Planelles, L., de Jong-Odding, J., Hardenberg, G., Pals, S.T., Hahne, M., Spaargaren, M., Medema, J.P., 2005. Heparan sulfate proteoglycan binding promotes APRIL-induced tumor cell proliferation. *Cell Death Differ.* 12, 637–648. <https://doi.org/10.1038/sj.cdd.4401647>.
- Hintze, V., Samsonov, S.A., Anselmi, M., Moeller, S., Becher, J., Schnabelrauch, M., Scharnweber, D., Pisabarro, M.T., 2014. Sulfated glycosaminoglycans exploit the conformational plasticity of bone morphogenetic protein-2 (BMP-2) and alter the interaction profile with its receptor. *Biomacromolecules* 15, 3083–3092. <https://doi.org/10.1021/bm5006855>.
- Ho, G., Broze, G.J., Schwartz, A.L., 1997. Role of heparan sulfate proteoglycans in the uptake and degradation of tissue factor pathway inhibitor-coagulation factor Xa complexes. *J. Biol. Chem.* 272, 16838–16844. <https://doi.org/10.1074/jbc.272.27.16838>.
- Huang, Y., Mao, Y., Zong, C., Lin, C., Boons, G.-J., Zaia, J., 2015. Discovery of a Heparan Sulfate 3-O-Sulfation Specific Peeling Reaction. *Anal. Chem.* 87, 592–600. <https://doi.org/10.1021/ac503248k>.
- Hunter, J.D., 2007. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* 9, 90–95.
- Irie, A., Yates, E.A., Turnbull, J.E., Holt, C.E., 2002. Specific heparan sulfate structures involved in retinal axon targeting. *Dev. Camb. Engl.* 129, 61–70.
- Johnson, C.E., Crawford, B.E., Stavridis, M., Ten Dam, G., Wat, A.L., Rushton, G., Ward, C.M., Wilson, V., van Kuppevelt, T.H., Esko, J.D., Smith, A., Gallagher, J.T., Merry, C.L.R., 2007. Essential alterations of heparan sulfate during the differentiation of embryonic stem cells to Sox1-enhanced green fluorescent protein-expressing neural progenitor cells. *Stem Cells Dayt. Ohio* 25, 1913–1923. <https://doi.org/10.1634/stemcells.2006-0445>.
- Jokiranta, T.S., Cheng, Z.-Z., Seeberger, H., J  si, M., Heinen, S., Noris, M., Remuzzi, G., Ormsby, R., Gordon, D.L., Meri, S., Hellwage, J., Zipfel, P.F., 2005. Binding of complement factor H to endothelial cells is mediated by the carboxy-terminal glycosaminoglycan binding site. *Am. J. Pathol.* 167, 1173–1181. [https://doi.org/10.1016/S0002-9440\(10\)61205-9](https://doi.org/10.1016/S0002-9440(10)61205-9).
- Joseph, P.R.B., Mosier, P.D., Desai, U.R., Rajarathnam, K., 2015. Solution NMR characterization of chemokine CXCL8/IL-8 monomer and dimer binding to glycosaminoglycans: structural plasticity mediates differential binding interactions. *Biochem. J.* 472, 121–133. <https://doi.org/10.1042/BJ20150059>.
- Kinnunen, T., Raulo, E., Nolo, R., Maccarana, M., Lindahl, U., Rauvala, H., 1996. Neurite outgrowth in brain neurons induced by heparin-binding growth-associated molecule (HB-GAM) depends on the specific interaction of HB-GAM with heparan sulfate at the cell surface. *J. Biol. Chem.* 271, 2243–2248. <https://doi.org/10.1074/jbc.271.4.2243>.
- Kirschner, K.N., Yongye, A.B., Tschampel, S.M., Gonz  lez-Outeiri  , J., Daniels, C.R., Foley, B.L., Woods, R.J., 2008. GLYCAM06: a generalizable biomolecular force field. *Carbohydr. J. Comput. Chem.* 29, 622–655. <https://doi.org/10.1002/jcc.20820>.
- Koehler, L., Samsonov, S., Rother, S., Vogel, S., K  hling, S., Moeller, S., Schnabelrauch, M., Rademann, J., Hempel, U., Pisabarro, M.T., Scharnweber, D., Hintze, V., 2017. Sulfated Hyaluronan Derivatives Modulate TGF-  1 Receptor Complex Formation: Possible Consequences for TGF-  1 Signaling. *Sci. Rep.* 7, 1210. <https://doi.org/10.1038/s41598-017-01264-8>.
- Kogut, M.M., Maszota-Zieleniak, M., Marcisz, M., Samsonov, S.A., 2021. Computational insights into the role of calcium ions in protein-glycosaminoglycan systems. *Phys. Chem. Chem. Phys.* PCCP 23, 3519–3530. <https://doi.org/10.1039/d0cp05438k>.
- Kraushaar, D.C., Rai, S., Condac, E., Nairn, A., Zhang, S., Yamaguchi, Y., Moremen, K., Dalton, S., Wang, L., 2012. Heparan sulfate facilitates FGF and BMP signaling to drive mesoderm differentiation of mouse embryonic stem cells. *J. Biol. Chem.* 287, 22691–22700. <https://doi.org/10.1074/jbc.M112.368241>.
- K  n  , G., Huster, D., Samsonov, S.A., 2021. Investigation of the structure of regulatory proteins interacting with glycosaminoglycans by combining NMR spectroscopy and molecular modeling - the beginning of a wonderful friendship. *Biol. Chem.* <https://doi.org/10.1515/hsz-2021-0119>.
- Li, Y., Sun, C., Yates, E.A., Jiang, C., Wilkinson, M.C., Fernig, D.G., 2016. Heparin binding preference and structures in the fibroblast growth factor family parallel their evolutionary diversification. *Open Biol.* 6, 150275. <https://doi.org/10.1098/rsob.150275>.
- Liu, J., Shriver, Z., Pope, R.M., Thorp, S.C., Duncan, M.B., Copeland, R.J., Raska, C.S., Yoshida, K., Eisenberg, R.J., Cohen, G., Linhardt, R.J., Sasisekharan, R., 2002. Characterization of a heparan sulfate octasaccharide that binds to herpes simplex virus type 1 glycoprotein D. *J. Biol. Chem.* 277, 33456–33467. <https://doi.org/10.1074/jbc.M202034200>.
- Luo, Y., Ye, S., Kan, M., McKeehan, W.L., 2006. Structural specificity in a FGF7-affinity purified heparin octasaccharide required for formation of a complex with FGF7 and FGFR2IIIb. *J. Cell. Biochem.* 97, 1241–1258. <https://doi.org/10.1002/jcb.20724>.
- McKinney, W., others, 2010. Data structures for statistical computing in python. In: *Proceedings of the 9th Python in Science Conference*. p. 51–6.
- Morgan, A., Sepuru, K.M., Feng, W., Rajarathnam, K., Wang, X., 2015. Flexible Linker Modulates Glycosaminoglycan Affinity of Decorin Binding Protein A. *Biochemistry* 54, 5113–5119. <https://doi.org/10.1021/acs.biochem.5b00253>.
- Mosier, P.D., Krishnasamy, C., Kellogg, G.E., Desai, U.R., 2012. On the specificity of heparin/heparan sulfate binding to proteins. Anion-binding sites on antithrombin and thrombin are fundamentally different. *PLoS One* 7, e48632. <https://doi.org/10.1371/journal.pone.0048632>.
- Multhaup, G., 1994. Identification and regulation of the high affinity binding site of the Alzheimer's disease amyloid protein precursor (APP) to glycosaminoglycans. *Biochimie* 76, 304–311. [https://doi.org/10.1016/0300-9084\(94\)90163-5](https://doi.org/10.1016/0300-9084(94)90163-5).

- Nakato, H., Kimata, K., 2002. Heparan sulfate fine structure and specificity of proteoglycan functions. *Biochim. Biophys. Acta* 1573, 312–318. [https://doi.org/10.1016/S0304-4165\(02\)00398-7](https://doi.org/10.1016/S0304-4165(02)00398-7).
- O'Donnell, C.D., Shukla, D., 2008. The Importance of Heparan Sulfate in Herpesvirus Infection. *Virology* 323, 383–393. <https://doi.org/10.1007/s12250-008-2992-1>.
- Panitz, N., Theissen, S., Samsonov, S.A., Gehrcke, J.-P., Baumann, L., Bellmann-Sickert, K., Köhling, S., Pisabarro, M.T., Rademann, J., Huster, D., Beck-Sickinger, A. G., 2016. The structural investigation of glycosaminoglycan binding to CXCL12 displays distinct interaction sites. *Glycobiology* 26, 1209–1221. <https://doi.org/10.1093/glycob/cwv059>.
- Patel, V.N., Likar, K.M., Zisman-Rozen, S., Cowherd, S.N., Lassiter, K.S., Sher, I., Yates, E. A., Turnbull, J.E., Ron, D., Hoffman, M.P., 2008. Specific heparan sulfate structures modulate FGF10-mediated submandibular gland epithelial morphogenesis and differentiation. *J. Biol. Chem.* 283, 9308–9317. <https://doi.org/10.1074/jbc.M70995200>.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Pejler, G., Danielsson, A., Björk, I., Lindahl, U., Nader, H.B., Dietrich, C.P., 1987. Structure and antithrombin-binding properties of heparin isolated from the clams *Anomalocardia brasiliana* and *Tivela mactroides*. *J. Biol. Chem.* 262, 11413–11421. [https://doi.org/10.1016/S0021-9258\(18\)60822-1](https://doi.org/10.1016/S0021-9258(18)60822-1).
- Pempe, E.H., Xu, Y., Gopalakrishnan, S., Liu, J., Harris, E.N., 2012. Probing structural selectivity of synthetic heparin binding to Stabilin protein receptors. *J. Biol. Chem.* 287, 20774–20783. <https://doi.org/10.1074/jbc.M111.320069>.
- Pichert, A., Samsonov, S.A., Theissen, S., Thomas, L., Baumann, L., Schiller, J., Beck-Sickinger, A.G., Huster, D., Pisabarro, M.T., 2012. Characterization of the interaction of interleukin-8 with hyaluronan, chondroitin sulfate, dermatan sulfate and their sulfated derivatives by spectroscopy and molecular modeling. *Glycobiology* 22, 134–145. <https://doi.org/10.1093/glycob/cwr120>.
- Pomin, V.H., Mulloy, B., 2015. Current structural biology of the heparin interactome. *Curr. Opin. Struct. Biol.* 34, 17–25. <https://doi.org/10.1016/j.sbi.2015.05.007>.
- Pratt, C.W., Church, F.C., 1992. Heparin binding to protein C inhibitor. *J. Biol. Chem.* 267, 8789–8794.
- Pye, D.A., Vives, R.R., Hyde, P., Gallagher, J.T., 2000. Regulation of FGF-1 mitogenic activity by heparan sulfate oligosaccharides is dependent on specific structural features: differential requirements for the modulation of FGF-1 and FGF-2. *Glycobiology* 10, 1183–1192. <https://doi.org/10.1093/glycob/10.11.1183>.
- Qazi, H., Shi, Z.-D., Song, J.W., Cancel, L.M., Huang, P., Zeng, Y., Roberge, S., Munn, L. L., Tarbell, J.M., 2016. Heparan sulfate proteoglycans mediate renal carcinoma metastasis. *Int. J. Cancer* 139, 2791–2801. <https://doi.org/10.1002/ijc.30397>.
- R Core Team, 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Raghuraman, A., Mosier, P.D., Desai, U.R., 2010. Understanding Dermatan Sulfate-Heparin Cofactor II Interaction through Virtual Library Screening. *ACS Med. Chem. Lett.* 1, 281–285. <https://doi.org/10.1021/ml100048y>.
- Raghuraman, A., Mosier, P.D., Desai, U.R., 2006. Finding a needle in a haystack: development of a combinatorial virtual screening approach for identifying high specificity heparin/heparan sulfate sequence(s). *J. Med. Chem.* 49, 3553–3562. <https://doi.org/10.1021/jm060092o>.
- Revelle, William, 2021. psych: Procedures for Personality and Psychological Research [WWW Document]. URL (<https://CRAN.R-project.org/package=psych>).
- Roe, D.R., Cheatham, T.E., 2013. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* 9, 3084–3095. <https://doi.org/10.1021/ct400341p>.
- Rogers, C.J., Clark, P.M., Tully, S.E., Abrol, R., Garcia, K.C., Goddard, W.A., Hsieh-Wilson, L.C., 2011. Elucidating glycosaminoglycan-protein interactions using carbohydrate microarray and computational approaches. *Proc. Natl. Acad. Sci. U. S. A.* 108, 9747–9752. <https://doi.org/10.1073/pnas.1102962108>.
- Rosenberg, R.D., Shworak, N.W., Liu, J., Schwartz, J.J., Zhang, L., 1997. Heparan sulfate proteoglycans of the cardiovascular system. Specific structures emerge but how is synthesis regulated. *J. Clin. Invest.* 99, 2062–2070. <https://doi.org/10.1172/JCI119377>.
- Rother, S., Samsonov, S.A., Hempel, U., Vogel, S., Moeller, S., Blaszkiewicz, J., Köhling, S., Schnabelrauch, M., Rademann, J., Pisabarro, M.T., Hintze, V., Scharnweber, D., 2016. Sulfated Hyaluronan Alters the Interaction Profile of TIMP-3 with the Endocytic Receptor LRP-1 Clusters II and IV and Increases the Extracellular TIMP-3 Level of Human Bone Marrow Stromal Cells. *Biomacromolecules* 17, 3252–3261. <https://doi.org/10.1021/acs.biomac.6b00980>.
- Ryckaert, J.-P., Cicciotti, G., Berendsen, H.J.C., 1977. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* 23, 327–341. [https://doi.org/10.1016/0021-9991\(77\)90098-5](https://doi.org/10.1016/0021-9991(77)90098-5).
- Sage, J., Mallèvre, F., Barbarin-Costes, F., Samsonov, S.A., Gehrcke, J.-P., Pisabarro, M. T., Perrier, E., Schnebert, S., Roget, A., Livache, T., Nizard, C., Lalmanach, G., Lecaille, F., 2013. Binding of chondroitin 4-sulfate to cathepsin S regulates its enzymatic activity. *Biochemistry* 52, 6487–6498. <https://doi.org/10.1021/bi400925g>.
- Sankaranarayanan, N.V., Desai, U.R., 2014. Toward a robust computational screening strategy for identifying glycosaminoglycan sequences that display high specificity for target proteins. *Glycobiology* 24, 1323–1333. <https://doi.org/10.1093/glycob/cwu077>.
- Sankaranarayanan, N.V., Sarkar, A., Desai, U.R., Mosier, P.D., 2015. Designing “high-affinity, high-specificity” glycosaminoglycan sequences through computerized modeling. *Methods Mol. Biol. Clifton NJ* 1229, 289–314. https://doi.org/10.1007/978-1-4939-1714-3_24.
- Sankaranarayanan, N.V., Strebel, T.R., Boothello, R.S., Sheerin, K., Raghuraman, A., Sallas, F., Mosier, P.D., Watermeyer, N.D., Oscarson, S., Desai, U.R., 2017. A Hexasaccharide Containing Rare 2-O-Sulfate-Glucuronic Acid Residues Selectively Activates Heparin Cofactor II. *Angew. Chem. Int. Ed. Engl.* 56, 2312–2317. <https://doi.org/10.1002/anie.201609541>.
- Sarkar, A., Desai, U.R., 2015. A Simple Method for Discovering Druggable, Specific Glycosaminoglycan-Protein Systems. Elucidation of Key Principles from Heparin/Heparan Sulfate-Binding Proteins. *PLoS One* 10, e0141127. <https://doi.org/10.1371/journal.pone.0141127>.
- Sattelle, B.M., Shakeri, J., Cliff, M.J., Almond, A., 2015. Proteoglycans and Their Heterogeneous Glycosaminoglycans at the Atomic Scale. *Biomacromolecules* 16, 951–961. <https://doi.org/10.1021/bm5018386>.
- Schlorke, D., Thomas, L., Samsonov, S.A., Huster, D., Arnhold, J., Pichert, A., 2012. The influence of glycosaminoglycans on IL-8-mediated functions of neutrophils. *Carbohydr. Res.* 356, 196–203. <https://doi.org/10.1016/j.carres.2012.02.025>.
- Shworak, N.W., Shirakawa, M., Collic-Jouault, S., Liu, J., Mulligan, R.C., Birinyi, L.K., Rosenberg, R.D., 1994. Pathway-specific regulation of the synthesis of anticoagulant active heparan sulfate. *J. Biol. Chem.* 269, 24941–24952.
- Spencer, J.L., Bernanke, J.A., Buczek-Thomas, J.A., Nugent, M.A., 2010. A Computational Approach for Deciphering the Organization of Glycosaminoglycans. *PLOS ONE* 5, e9389. <https://doi.org/10.1371/journal.pone.0009389>.
- Stancanelli, E., Elli, S., Hsieh, P.-H., Liu, J., Guerrini, M., 2018. Recognition and Conformational Properties of an Alternative Antithrombin Binding Sequence Obtained by Chemoenzymatic Synthesis. *Chembiochem Eur. J. Chem. Biol.* <https://doi.org/10.1002/cbic.201800095>.
- Stringer, S.E., Gallagher, J.T., 1997. Specific binding of the chemokine platelet factor 4 to heparan sulfate. *J. Biol. Chem.* 272, 20508–20514. <https://doi.org/10.1074/jbc.272.33.20508>.
- Sun, Y.J., Chang, N.C., Hung, S.I., Chang, A.C., Chou, C.C., Hsiao, C.D., 2001. The crystal structure of a novel mammalian lectin, Ym1, suggests a saccharide binding site. *J. Biol. Chem.* 276, 17507–17514. <https://doi.org/10.1074/jbc.M010416200>.
- Taylor, G.J., Yorke, S.C., Harding, D.R., 1995. Glycosaminoglycan specificity of a heparin-binding peptide. *Pept. Res.* 8, 286–293.
- Thacker, B.E., Seamen, E., Lawrence, R., Parker, M.W., Xu, Y., Liu, J., Vander Kooi, C.W., Esko, J.D., 2016. Expanding the 3-O-Sulfate Proteome-Enhanced Binding of Neuropilin-1 to 3-O-Sulfated Heparan Sulfate Modulates Its Activity. *ACS Chem. Biol.* 11, 971–980. <https://doi.org/10.1021/acscchembio.5b00897>.
- Thacker, B.E., Xu, D., Lawrence, R., Esko, J.D., 2014. Heparan sulfate 3-O-sulfation: A rare modification in search of a function. *Matrix Biol., Prote Biol.* 35, 60–72. <https://doi.org/10.1016/j.matbio.2013.12.001>.
- Ugby, B., Schmidthoffer, I., Szilak, L., 2019. Heparan sulfate proteoglycan (HSPG) can take part in cell division: inside and outside. *Cell. Mol. Life Sci. CMLS* 76, 865–871. <https://doi.org/10.1007/s00018-018-2964-z>.
- Uniewicz, K.A., Fernig, D.G., 2008. Neuropilins: a versatile partner of extracellular molecules that regulate development and disease. *Front. Biosci. J. Virtual Libr.* 13, 4339–4360. <https://doi.org/10.2741/3008>.
- Viviano, B.L., Paine-Saunders, S., Gasunas, N., Gallagher, J., Saunders, S., 2004. Domain-specific modification of heparan sulfate by Qsulf1 modulates the binding of the bone morphogenetic protein antagonist Noggin. *J. Biol. Chem.* 279, 5604–5611. <https://doi.org/10.1074/jbc.M310691200>.
- Waskom, M.L., 2021. seaborn: statistical data visualization. *J. Open Source Softw.* 6, 3021. <https://doi.org/10.21105/joss.03021>.
- Witt, D.P., Lander, A.D., 1994. Differential binding of chemokines to glycosaminoglycan subpopulations. *Curr. Biol. CB* 4, 394–400. [https://doi.org/10.1016/S0960-9822\(00\)00088-9](https://doi.org/10.1016/S0960-9822(00)00088-9).
- Yokoyama, M., Matsuzawa, T., Yoshikawa, T., Nunomiya, A., Yamaguchi, Y., Yanai, K., 2020. Heparan sulfate controls skeletal muscle differentiation and motor functions. *Biochim. Biophys. Acta Gen. Subj.* 1864, 129707 <https://doi.org/10.1016/j.bbagen.2020.129707>.
- Yue, J., Jin, W., Yang, H., Faulkner, J., Song, X., Qiu, H., Teng, M., Azadi, P., Zhang, F., Linhardt, R.J., Wang, L., 2021. Heparan Sulfate Facilitates Spike Protein-Mediated SARS-CoV-2 Host Cell Invasion and Contributes to Increased Infection of SARS-CoV-2 G614 Mutant and in Lung Cancer. *Front. Mol. Biosci.* 8, 649575 <https://doi.org/10.3389/fmolb.2021.649575>.
- Zhang, B., Xiao, W., Qiu, H., Zhang, F., Moniz, H.A., Jaworski, A., Condac, E., Gutierrez-Sanchez, G., Heiss, C., Clugston, R.D., Azadi, P., Greer, J.J., Bergmann, C., Moremen, K.W., Li, D., Linhardt, R.J., Esko, J.D., Wang, L., 2014. Heparan sulfate deficiency disrupts developmental angiogenesis and causes congenital diaphragmatic hernia. *J. Clin. Invest.* 124, 209–221. <https://doi.org/10.1172/JCI71090>.
- Zhao, J., Zhu, Y., Song, X., Yuanyuan, Xiao, Su, G., Liu, X., Wang, Z., Xu, Y., Liu, J., Eliezer, D., Ramlall, T.F., Lippens, G., Gibson, J., Zhang, F., Linhardt, R.J., Wang, L., Wang, C., 2020. 3-O-Sulfation of Heparan Sulfate Enhances Tau Interaction and Cellular Uptake. *Angew. Chem. Int. Ed. Engl.* 59, 1818–1827. <https://doi.org/10.1002/anie.201913029>.