

Faculty of Chemistry, University of Gdańsk

MChem Małgorzata Marzena Kogut-Günthel

**COMPUTATIONAL APPROACHES TO
CHARACTERISE BIOLOGICALLY ACTIVE
SYSTEMS CONTAINING PROTEINS,
CARBOHYDRATES, AND IONS**

Promotor:

dr hab. Sergey A. Samsonov, prof UG

Promotor pomocniczy:

dr Martyna Maszota-Zieleniak

*Department of Theoretical Chemistry
Laboratory of Molecular Modelling*

Gdańsk 2023

Foremost, I would like to express my sincere gratitude to my supervisors, dr hab. Sergey Samsonov, prof. UG and dr Martyna Maszota-Zieleniak for the continuous support of my PhD study and research and their patience, motivation, enthusiasm, and immense knowledge.

Their guidance helped me in all the time of research and writing of this Thesis.

I would also like to thank dr hab. Dariusz Wyrzykowski, prof. UG and his lab members for sharing their professional expertise and close collaboration.

Finally, I would like to thank all my colleagues from Department of Theoretical Chemistry for the friendly working atmosphere.

The results presented in this Thesis were obtained with the support of National Science Centre (project Sonata Bis “Modeling of glycosaminoglycan-induced formation of protein structure and enhancement of biologically relevant protein-ligand interactions”, UMO-2018/30/E/ST4/00037).

Przede wszystkim pragnę serdecznie podziękować moim Promotorom dr hab. Sergeyowi Samsonowowi, prof. UG i dr Martynie Maszocie-Zieleniak za nieustanne wsparcie podczas moich badań prowadzonych w ramach studiów doktoranckich, za okazaną cierpliwość, motywację, entuzjazm i ogromną wiedzę. Ich pomoc towarzyszyła mi przez cały okres trwania badań oraz podczas pisania niniejszej pracy. Dziękuję również dr hab. Dariuszowi Wyrzykowskiemu, prof. UG i jego Współpracownikom za podzielenie się swoją fachową wiedzą oraz pomyślną współpracę. Na koniec chciałbym podziękować wszystkim Koleżankom i Kolegom z Katedry Chemii Teoretycznej, którzy stworzyli przyjazną atmosferę w pracy.

Przedstawione w pracy wyniki uzyskano dzięki wsparciu projektu Narodowego Centrum Nauki pt. „Modelowanie molekularne tworzenia struktury białkowej przez glikozaminoglikany oraz ich wpływ na biologicznie istotne oddziaływania białko-ligand” (UMO-2018/30/E/ST4/00037).

Table of Contents

Abbreviations	5
Abstract	6
Streszczenie	8
1. Introduction	10
1.1. Macromolecules and their biological relevance	10
1.2. The role of ions in carbohydrate-containing systems	15
1.2.1. Glycosaminoglycans	15
1.2.2. The role of ions in glycosaminoglycan-containing systems	19
1.2.3. Cyclodextrins	22
1.2.4. The role of surfactants in cyclodextrin-containing systems	25
1.3. The role of ions in protein-containing systems	28
1.3.1. The role of surfactants in albumin-containing systems	28
1.3.2. Albumins	30
2. Methods	32
2.1. Experimental approaches to characterise interactions in receptor-ligand systems	34
2.1.1. X-ray crystallography	34
2.1.2. Nuclear magnetic resonance	36
2.1.3. Other experimental methods	38
2.2. Theoretical approaches to characterise interactions in receptor-ligand systems	42
2.2.1. Electrostatic potential calculations	42
2.2.2. Molecular docking	43
2.2.3. Molecular dynamics	49
2.2.4. Binding free energy calculations	60
2.2.5. Challenges of molecular modelling of GAG and metal ions-containing systems	65

3. Goals of the research	70
4. Summary of the publications included in the PhD Thesis	71
4.1. Computational insights into the calcium ions role in protein-glycosaminoglycan systems	71
4.2. Impact of calcium ions on the structural and dynamic properties of heparin oligosaccharides by computational analysis	73
4.3. Affinity and putative entrance mechanisms of alkyl sulfates into the β -CD cavity	75
4.4. Physicochemical nature of sodium dodecyl sulfate interactions with bovine serum albumin revealed by interdisciplinary approaches	77
4.5. Effect of tetraphenylborate on physicochemical properties of bovine serum albumin	79
4.6. Induced circular dichroism as a tool to monitor the displacement of ligands between albumins	81
5. Conclusions and outlook	83
6. Literature	85
7. Publications included in the PhD Thesis	115
Appendix	116

Abbreviations

AA – all-atom	LFMM – ligand field molecular dynamics model
APP – amyloid precursor protein	LIE – linear interaction energy
APRIL – a proliferation-inducing ligand	LR-CDs – large ring cyclodextrins
ATIII – antithrombin III	MALDI – matrix-assisted laser desorption ionisation
ATIII – antithrombin III	MD – molecular dynamics
BMP – bone morphogenic protein	MG – Mansonone G
BSA – bovine serum albumin	MM-GBSA – molecular mechanics Generalised-Born surface area
[B(Ph₄)]⁻ – tetraphenylborate ions	MM-PBSA – molecular mechanics Poisson-Boltzmann surface area
CAPE – caffeic acid phenethyl ester	MS – mass spectrometry
Ca²⁺ – calcium ions	NMR – nuclear magnetic resonance
CD – cyclodextrin	OPLS – optimised potentials for liquid simulations
CDAM – cationic dummy atom model	PBSA – Poisson-Boltzmann surface area
CG – coarse-grained	PCPE – procollagen c-proteinase enhancer
CNs – coordination numbers	PDB – protein data bank
COP – colloid osmotic pressure	PES – potential energy surface
CS – chondroitin sulfate	PG – proteoglycans
CT – charge transfer	pGAG – phosphorylated GAG
C4-S – chondroitin 4-sulfate	PrP – platelet-rich plasma
C6-S – chondroitin 6-sulfate	RDF – radial distribution function
DO – Drude oscillator	REMD – replica exchange molecular dynamics
DS – dermatan sulfate	RS-REMD – repulsive-scaling replica exchange molecular dynamics
DMD – dynamic molecular docking	RMSD – root-mean-square deviation
DVN – divanillin	SA – serum albumin
DVT – divanillate	SAXS – small-angle-X-ray scattering
ECM – extracellular matrix	SDS – sodium dodecyl sulfate
EI – electron ionisation	SPR – surface plasmon resonance
ELISA – enzyme-linked immunosorbent assay	IMP – tissue inhibitor of metalloproteinase
ESI – electrospray ionisation	TM – transition metals
ESP – electrostatic potential	TRP – tryptophan
FEP – free energy perturbation	US – umbrella sampling
FF – force field	VEGF – vascular endothelial growth factor
FGF – fibroblast growth factor	WHAM – weighted histogram analysis method
FGFR – fibroblast growth factor receptor	XRD – X-ray diffraction
GAGs – glycosaminoglycans	
GAG-DB – glycosaminoglycan database	
HA – hyaluronic acid	
HCI – HP cofactor II	
HFE – hydration-free energy	
HP – heparin	
HS – heparan sulfate	
HSA – human serum albumin	
IDM – induced dipole moment	
IL-7 – interleukin 7	
IOD – ion-oxygen distance	
ITC – isothermal titration calorimetry	
KS – keratan sulfate	

Abstract

Biomolecular recognition, including the binding of small molecules, peptides, carbohydrates, and proteins to their target receptors, is fundamental in biological processes such as immune responses, cellular signalling, and catalysis. Uncovering the physical mechanisms of biomolecular recognition and characterising the critical biomolecular interactions is vital to understanding their functions. Furthermore, these processes are implicated in developing various human diseases and serve as potential drug targets. One way to get a deeper understanding of these interactions at the atomic level is by deploying computational tools. Computational methods are constantly advancing to model biomolecular recognition and predict binding thermodynamics thanks to the increasing accessibility to the power of supercomputers. They are not only complementary to experiments but also recognised as powerful tools capable of providing experimentally inaccessible insights.

This PhD Thesis aimed to characterise biologically active systems consisting of proteins, carbohydrates, and ions using theoretical approaches. The molecular systems chosen as the research objects were glycosaminoglycans (GAGs), linear anionic periodic polysaccharides with metal ions, cyclodextrin (CD) with anionic surfactants, and bovine serum albumin (BSA) with small molecules. In the first part of the Thesis, I focused on investigating the role of ions in carbohydrate-containing systems. The initial goal was to understand the calcium ion (Ca^{2+}) role in annexin- Ca^{2+} -heparin (HP) and HP- Ca^{2+} systems at the atomic level. To do this, I: a) examined how the most commonly deployed standard molecular modelling tools are sensitive and accurate to investigate the protein-ion-GAG systems, and b) rigorously characterised how different Ca^{2+} parameters affect HP's structural and dynamic properties in the simulation. Next, β -CD systems with some alkyl sulfates (SXS) were analyzed (where X=8,10 and 12 denote the number of carbon atoms in the alkyl chain of the sulfate). In this study, I: a) investigated how the alkyl chain length affects the CD-SXS interactions; b) proposed a potential molecular mechanism for the entrance of the ion into the CD cavity; and c) determined how the initial SXS orientation influences the formed inclusion complex. I also reviewed and summarised four aspects of the currently deployed theoretical approaches for investigating protein-GAG complexes, including molecular docking, free binding energy calculations, modelling ion impacts, and addressing the phenomena of multipose binding of GAGs to proteins.

The second part of the Thesis was dedicated to the BSA-containing systems and the BSA interactions with small molecules, including ions. Firstly, I deployed computational techniques

to monitor the influence of pH and temperature on the interactions in the BSA-sodium dodecyl sulfate (SDS) complex and localised potential binding sites and poses. Then, I localised two binding sites in the BSA-[B(Ph)₄]⁻ complex. Finally, I applied the umbrella sampling (US) protocol to investigate the free energy profile of divanillate (DVT) and divanillin (DVN) orientation change in terms of the dihedral angle between the planes defined by the aromatic moieties of DVT/DVN on the BSA/HSA surfaces or in the absence of the protein.

The results of my PhD research contributed to a better understanding role of ions in biologically active systems, and an attempt to develop novel protocols to model these systems more efficiently was undertaken. Moreover, the limitations of computational methods are discussed in detail, and potential solutions to overcome those challenges are proposed.

Streszczenie

Rozpoznawanie biomolekularne, w tym wiązanie małych cząsteczek, peptydów, węglowodanów i białek z ich docelowymi receptorami, ma kluczowe znaczenie w procesach biologicznych, takich jak odpowiedzi immunologiczne, sygnalizacja komórkowa i kataliza. Poznanie mechanizmów rozpoznawania biomolekularnego i scharakteryzowanie krytycznych oddziaływań biomolekuł ma istotne znaczenie dla zrozumienia ich funkcji. Procesy te mogą przyczyniać się do rozwoju różnych chorób w organizmie człowieka, mogą więc służyć jako potencjalne cele dla leków. Jednym ze sposobów lepszego zbadania oddziaływań biomolekuł na poziomie atomowym jest wykorzystanie metod teoretycznych, takich jak modelowanie molekularne. Metody teoretyczne są stale rozwijane, aby bardziej efektywnie modelować coraz bardziej skomplikowane układy biomolekularne, a także przewidywać termodynamikę ich wiązania. Nie tylko uzupełniają one metody eksperymentalne, ale również są stanowią potężne narzędzia, które umożliwiają obserwację procesów, których nie można zbadać metodami eksperymentalnymi.

Badania prowadzone w ramach niniejszej rozprawy doktorskiej miały na celu scharakteryzowanie biologicznie aktywnych układów składających się z białek, węglowodanów i jonów przy użyciu teoretycznych metod obliczeniowych. Do badań wybrałam następujące układy molekularne: glikozaminoglikany (GAG), liniowe anionowe periodyczne polisacharydy z jonami metali, cyklodekstryna (CD) z anionowymi środkami powierzchniowo czynnymi oraz albumina surowicy bydlęcej (BSA) z małymi cząsteczkami. W pierwszej części badań skupiłam się na scharakteryzowaniu roli jonów w układach zawierających węglowodany. Początkowym celem było zrozumienie roli jonów wapnia (Ca^{2+}) w układach aneksyna- Ca^{2+} -heparyna (HP) i HP- Ca^{2+} na poziomie atomowym. Aby zrealizować ten cel: a) zbadalam, jak bardzo czułe i dokładne są najczęściej stosowane, standardowe narzędzia do modelowania molekularnego układów białko-jon-GAG, oraz b) scharakteryzowałam, w jaki sposób różne parametry Ca^{2+} wpływają na strukturalne i dynamiczne właściwości HP podczas symulacji dynamiki molekularnej. Następnie układy β -CD z niektórymi siarczanami alkilu (SXS) zostały przeanalizowane (gdzie X=8,10 i 12 oznaczają liczbę atomów węgla w łańcuch alkilowym siarczanu). W tej części badań: a) zbadalam, jak długość łańcucha alkilowego wpływa na oddziaływania w układach CD-SXS; b) zaproponowałam potencjalny mechanizm molekularny wejścia jonu do wnętrza pierścienia CD oraz c) ustaliłam, w jaki sposób początkowa orientacja SXS wpływa na powstający

kompleks inkluzyjny. Dokonałam również przeglądu literatury i podsumowałam cztery aspekty, które są obecnie stosowane w metodach teoretycznych do badań kompleksów białko-GAG między innymi dokowania molekularnego, obliczeń swobodnej energii wiązania, modelowania wpływów jonów i zjawiskami wielopozycyjnego wiązania się GAGów z białkami.

Druga część pracy poświęcona była układom zawierającym albuminę surowicy zwierzęcej (BSA) oraz oddziaływaniom BSA z małymi cząsteczkami, w tym z jonami. Na tym etapie badań zastosowałam metody obliczeniowe do monitorowania wpływu pH i temperatury na oddziaływanie w kompleksie BSA-dodecylosiarczan sodu (SDS) oraz zlokalizowania potencjalne miejsc wiązania się ligandu. Następnie ustaliłam dwa miejsca wiązania się liganda do receptora w kompleksie BSA-[B(Ph)₄]⁻. Na końcu zastosowałam protokół próbkowania parasolowego (US) w celu uzyskania profilu energii swobodnej zmiany orientacji diwanilanu (DVT) i diwaniliny (DVN) względem BSA/ albuminę surowicy ludzkiej (HSA) lub pod nieobecność białka.

Wyniki badań uzyskanych w ramach rozprawy doktorskiej przyczyniają się do lepszego zrozumienia roli jonów w aktywnie biologicznych układach molekularnych. Opracowałam także nowe protokoły obliczeniowe, które pozwalają na najefektywniejsze modelowanie tych układów. Ponadto szczegółowo omówiłam ograniczenia dostępnych metod obliczeniowych, a także zaproponowałam potencjalne rozwiązania opisanych problemów.

1. Introduction

1.1. Macromolecules and their biological relevance

Macromolecules are vital cellular components in biological systems responsible for carrying out numerous functions vital for the development and perseverance of living organisms. Proteins, carbohydrates, lipids and nucleic acids are the four major classes of those biological macromolecules. To investigate their function, structure, and behaviour, it is critical to understand the nature of interactions between any cluster of macromolecules.

Proteins play a pivotal role in countless processes in the cell, serving the following functions: hormonal (insulin), protection (immunoglobulin), transport (haemoglobin), storage (ferritin), contractile (myosin), enzymatic (sucrase), and structural (actin) [1-7]. Generally, proteins achieve their biological goals via direct physical interaction with other molecules (ligands). A ligand can be virtually anything: another protein, peptide, nucleic acid, membrane, organic substrate, or a small molecule (for example, metal ion, oxygen, or solvent). Three models were proposed to explain the protein-ligand binding mechanism: a) "lock and key" [8], b) induced fit [9], and c) conformational selection [10-11]. The "lock and key" model assumes that the protein and the ligand are rigid, and their binding interfaces should be perfectly matched. Therefore, only the correctly sized and shaped ligand (the key) can be inserted into the binding pocket (keyhole) of the protein (the lock) (Figure 1A). However, the "lock and key" model does not explain the experimental evidence that a protein binds its ligand when its initial shapes do not match. This resulted in the induced fit model, which assumes that the binding site in the protein is flexible and that the interacting ligand induces a conformational change at the binding site (Figure 1B). The conformational selection model postulates that the native state of a protein does not exist as a single, rigid conformation but rather as a vast ensemble of conformational states that coexist in equilibrium with different population distributions and that the ligand can bind selectively to the most suitable conformational state/substate, ultimately shifting the equilibrium towards this state (Figure 1C). In other words, the unbound protein can sample the same conformation as the ligand-bound state.

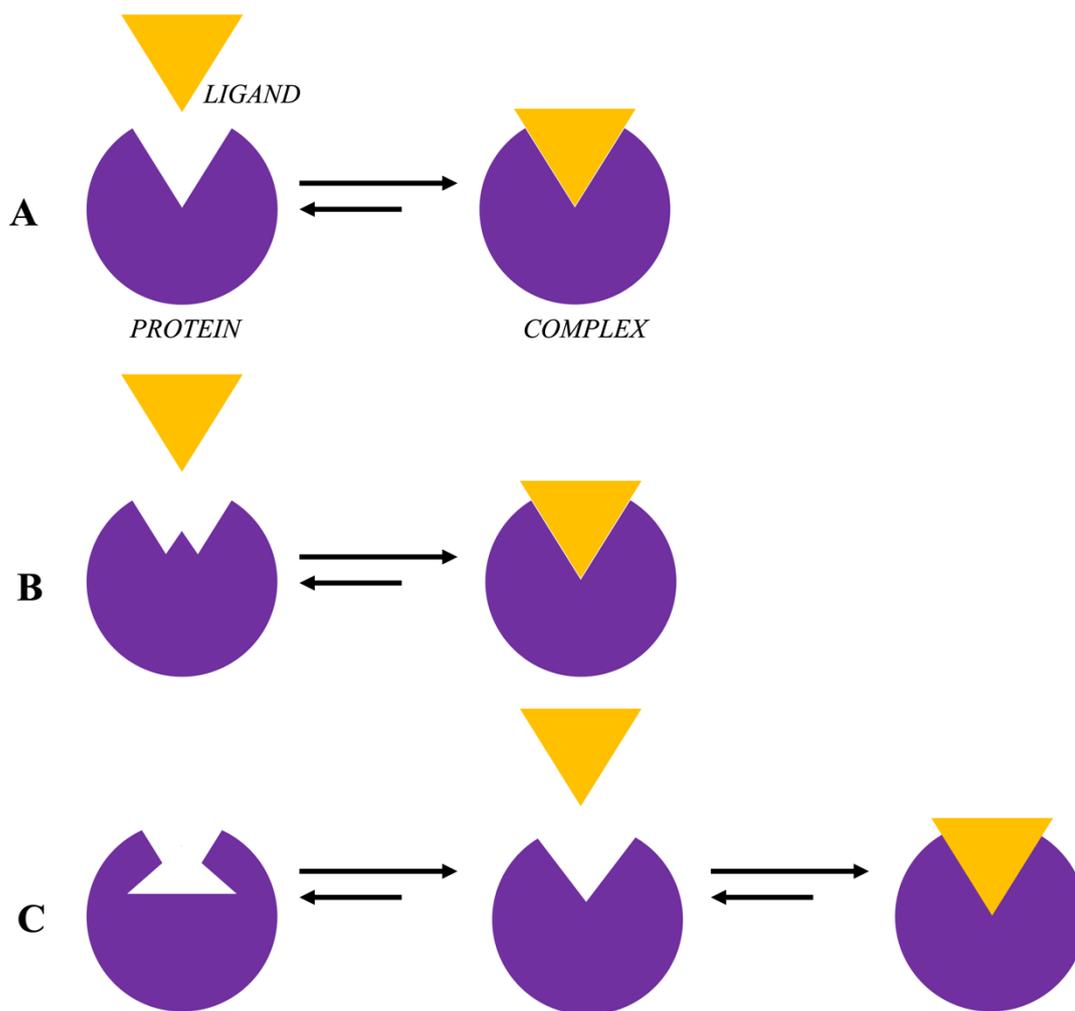
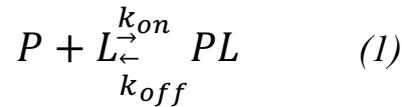


Figure 1. Schematic illustrations of the three protein-ligand binding models: “lock and key” (A), induced fit (B) and conformational selection (C) [10].

Molecular recognition is defined as a process in which biological macromolecules interact noncovalently with each other or small organic molecules, driven by van der Waals forces, electrostatic forces, hydrogen bonds and hydrophobic forces, to form a particular complex [12-14]. Understanding the process of biomolecular recognition is pivotal for molecular biology [9, 15-16] and represents the advancement of rational drug design [17-18]. Biomolecular recognition is described by binding specificity and binding affinity. Conventional specificity is a preference for a protein/ligand binding to a particular protein receptor over other competitive alternatives [19]. At the same time, affinity refers to the strength of the interaction between the ligand and its target protein. It is defined by either binding and dissociation constants (K_b and K_d) or Gibbs energy of binding (ΔG). In a simple, reversible protein-ligand (1:1) system, equilibrium exists between the free molecules protein (P), ligand (L), and their

complex (PL) that associate and dissociate at specific rates (described by rate constants k_{on} and k_{off} respectively) (1):



In the experiments, the binding affinity is described using equilibrium dissociation (K_d) or binding (K_b) constants (2):

$$K_b = \frac{[PL]}{[P][L]} = \frac{1}{K_d} = \frac{k_{on}}{k_{off}} \quad (2)$$

At equilibrium under standard conditions (298.15 K, 1 atmosphere), the Gibbs energy (free energy) of binding describes the energy difference between two states (3):

$$\Delta G_b = RT \ln([P][L]) - RT \ln[PL] = -RT \ln \frac{[PL]}{[P][L]} = -RT \ln K_b = RT \ln K_d \quad (3)$$

where R denotes the gas constant, and T is the absolute temperature at which binding occurs. ΔG_b can also be dissected into enthalpy (ΔH_b) and entropy (ΔS_b) (4):

$$\Delta G_b = \Delta H_b - T \Delta S_b \quad (4)$$

It must be stressed that molecular recognition is an element of a more complex, functionally crucial mechanism that consists of life's essential features, including information processing, metabolism, and self-replication. To better comprehend protein functions, one must thoroughly investigate and understand the detailed underlying mechanisms responsible for the protein-ligand interactions as much as possible. To do so, the more in-depth description, characterisation, and quantification of the thermodynamic patterns that govern/drive the formation of a given complex are pivotal [20]. Furthermore, since the rational drug design aims to use the knowledge of the structural data and protein-ligand binding mechanisms to optimise the process of discovery and development of new drugs, a profound understanding of the nature of the molecular recognition/interactions is of significant interest in facilitating this process [21].

Recently, natural biomacromolecules have attracted increased attention as carriers in biomedicine [22]. Biomacromolecules used as carriers include proteins (albumin, transferrin, lipoproteins, silk fibroin, collagen, keratin) and polysaccharides (chitosan, cyclodextrin, glycosaminoglycan, and pectin). They can be naturally obtained from animals and plants in abundant amounts and are renewable resources. They have an excellent affinity to organisms and weak immune rejection. They can also be degraded *in vivo* by enzymes, with metabolites of low toxicity to organisms [23-25]. Researchers reported multiple biomacromolecule-based carriers in the form of prodrugs, drug conjugates, nanoparticles, microcapsules, hydrogels, and tissue engineering scaffolds [26-28]. Furthermore, their use as drug carriers has been shown to reduce systemic toxicity and immunogenicity [29-31]. Moreover, the amine, carboxyl, and hydroxyl groups present within the structures of these biomacromolecules can be used for chemical modifications, making them of great significance in the biomedical field. From the biomacromolecules mentioned above, the following systems were chosen to be the research subject in this Thesis: glycosaminoglycans (GAGs) with metal ions, cyclodextrin (CD) with anionic surfactants (SXS) and bovine serum albumin (BSA) with small molecules (including ions).

Metalloglycomics studies the interactions of metal ions and coordination compounds with biologically relevant oligosaccharides, particularly GAGs, glycoproteins and proteoglycans. Glycomics is a broad scientific discipline aiming to explain glycans' diverse structures and functions in biological systems [32-34]. GAGs belong to long periodic negatively charged, often sulfated polysaccharides. Glycoproteins are essential macromolecules in the living organism and are involved in cellular processes. These molecules are found in the extracellular matrix (ECM) and are built from polysaccharide chains covalently bound to the amino acid side chains of the proteinic part. Proteoglycans (PGs), also present in the ECM, contain protein cores covalently bound to polysaccharides containing amino sugars, including GAGs [35-36]. In PGs, the bigger component is a carbohydrate, whereas, in glycoprotein, it is a protein. Moreover, PGs can form a gel-like substance by restraining the water molecules, making it possible for ions, hormones, and nutrients to move through them freely. PGs also interact with fibroblast growth factors (FGFs) via its polysaccharide chain, which supports FGFs binding to their specific cell-surface receptors [37-38]. Furthermore, the inflammatory response is prolonged when PGs interact with chemokines via their GAG parts on endothelial cells' surface [39]. Finally, PGs and GAGs are reported to be widely used as carriers for targeted therapy delivery [40-43].

Another group of highly valued biomacromolecules as drug carriers are cyclodextrins (CDs). CDs are cyclic polysaccharides with an amphiphilic structure composed of a hydrophilic exterior and hydrophobic cavity frequently used in drug design. The poorly water-soluble molecules (guests) can be encapsulated into CDs thanks to their ability to form inclusion complexes. Additionally, CDs reduce the immune stimulation of nucleic acid by avoiding nonspecific interactions in the physiological environment. It was also reported that CDs enhance the permeability of oligonucleotides across biological membranes [44].

The last family of drug-carrier described in this PhD Thesis are albumins. They belong to the main proteins in the human body (and many animal species) and are instrumental in the transport of various ions and in maintaining the colloidal osmotic pressure of the blood [45]. Albumins bind different endogenous and exogenous ligands, including water and metal cations, bilirubin, hormones, metalloporphyrin, nitric oxide, aspirin, warfarin, ibuprofen and phenylbutazone [46]. Almost all known drugs and toxic substances can bind to albumin [47]. Therefore, albumin not only determines their pharmaco- and toxicokinetics to a large extent but also transports them to target tissues or sites for further biotransformation. Since albumin is one of the most widely researched biological carriers, it has already been successful in clinical applications. In this context, the albumin-drug conjugates, albumin-binding prodrugs, albumin nanoparticles and albumin fusion proteins or peptides used in clinical studies were summarised by Zhang *et al.* [22].

1.2. The role of ions in carbohydrates-containing systems

This chapter reviews two groups of carbohydrates: glycosaminoglycans (GAGs) and cyclodextrins (CDs), and their interactions with various interacting partners, including ions.

1.2.1. Glycosaminoglycans

GAGs are long linear negatively charged periodic polysaccharide compounds. They are composed of repeating disaccharide units in every mammalian tissue [48]. The “glyco-“ prefix refers to galactose or a uronic sugar (glucuronic acid or iduronic acid) attached to amino sugar (N-acetylglucosamine or N-acetylgalactosamine). Alterations in the type of monosaccharides and various sulfation patterns (also known as sulfation code) result in the different major categories of GAGs, including hyaluronic acid (HA), chondroitin sulfate/dermatan sulfate (CS/DS) heparin/heparan sulfate (HP/HS) and keratan sulfate (KS). The molecular structure, charge and natural size of each GAG are summarised in Table 1 and Figure 1.

Table 1. Composition, charge, and size of GAGs.

Common name	Disaccharide unit composition	Charge per disaccharide unit	Size, disaccharide units
Hyaluronic acid (HA)	GlcNAc(β 1 \rightarrow 4)GlcA(β 1 \rightarrow 3)	-1	<30 000 [49]
Chondroitin sulfate (CS)	GalNAc(β 1 \rightarrow 4)GlcA(β 1 \rightarrow 3)	CS-A: -2 CS-C: -2 CS-D: -3 CS-E: -3	40-100 [50]
Dermatan sulfate (DS)	GalNAc(β 1 \rightarrow 4)IdoA(α 1 \rightarrow 3)	-2	40-100 [50]
Heparan sulfate (HS)	GlcNS(α 1 \rightarrow 4)IdoA(α 1 \rightarrow 4)	From -1 to -4	25-200 [51]
Heparin (HP)	GlcNS(α 1 \rightarrow 4)IdoA(α 1 \rightarrow 4)	-4	10-50 [52]
Keratan sulfate (KS)	GlcNAc(β 1 \rightarrow 3)Gal(β 1 \rightarrow 4)	-2	5-34 [53]

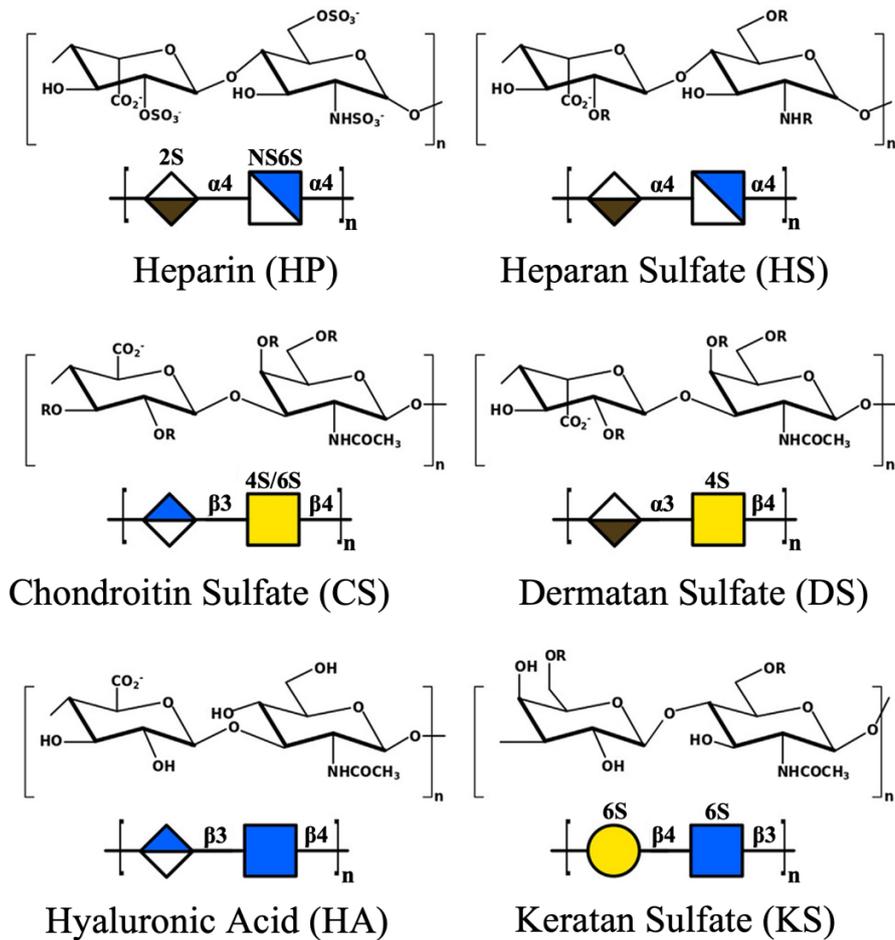


Figure 2: Repeating disaccharide units of GAGs with their SNFG representation. *R* denotes the position that can be sulfated [54].

GAGs are pivotal in numerous cellular processes by forming complexes with many proteins, mainly by electrostatic interactions [55]. The GAG biosynthesis process occurs in the Golgi apparatus, where an unsulfated polysaccharide chain is first synthesised, and then the chain becomes sulfated [56]. GAGs' functions within the body are widespread and strictly determined by their molecular structure. The structure of HA is the simplest one of all GAGs and does not require additional sulfation of functional groups in the Golgi apparatus as the other GAGs. Instead, the molecule consists of sequentially bound glucuronic acid and N-acetylglucosamine residues [57]. It is the only GAG not bound covalently within any proteoglycan [58]. HA can form linear chains with a broad spectrum of polysaccharide chain lengths in a living organism, starting from 100 to even 30 000 disaccharide units per chain [49]. The highly polar structure of HA makes it capable of binding water 10000 times its weight. Therefore, it plays a vital role in the lubrication of synovial joints and wound healing processes [59]. Clinicians use HA exogenously to promote tissue regeneration and skin repair, and HA has demonstrated safety and efficacy for this purpose [60]. HA is used in various cosmetic

products and shows promising effects in promoting skin tightness and elasticity and improving aesthetic scores [61]. Besides its water-binding capabilities, HA is involved in promoting and inhibiting angiogenesis and, therefore, is engaged in carcinogenesis [59]. The presence of HA breakdown products initiates an inflammatory response and, therefore, serves as a signal of injury [62].

CS comprises N-acetyl-D-galactosamine (GalNAc) and D-glucuronic acid (GlcA) per disaccharide unit [63]. The sulfation of hydroxyl groups can occur either in the position of the 4th carbon atom of GalNAc (resulting in chondroitin 4-sulfate; C4-S, CS-A) or of the 6th carbon atom of the same amino sugar (chondroitin 6-sulfate; C6-S; CS-C), resulting in a net charge -2 per a disaccharide unit. In some cases, sulfation can occur in both positions of the 4th and 6th carbon atoms of GalNAc, resulting in a „highly sulfated” chondroitin 4,6-sulfate (C4,6-S; CS-E) with a net charge of -3 per a disaccharide unit. This instance occurs in animals other than mammals. Compared to HA, CS chains are much shorter; they are usually composed of 40 to 100 recurring disaccharide units [50]. Considering the biological role of CS, it can be found in the cartilage, which is involved in the bone resorption process [64]. DS differs from CS in that instead of GlcA acid, and it is predominantly made up of L-iduronic acid (IdoA). Like CS, the net charge of DS is -2 per disaccharide unit [65]. It is possible that in its polysaccharide chain, DS can contain GlcA acid, making it a hybrid DS/CS molecule. Depending on its location, DS can be predominantly sulfated at the 4th carbon atom of GalNAc when present in the skin or at the 6th carbon atom of amino sugar when present in the umbilical cord [66]. Despite being mostly present in the skin, DS can also be found in blood vessels and lungs and has antithrombotic activity [67].

HS and HP are GAGs composed of D-glucosamine (GlcN) and IdoA disaccharide units [68]. Whereas the HS sulfation pattern may vary between the disaccharide unit (giving 48 variants in total [69]), HP (which is one of the 48 variants) is always sulfated at the 2nd (N-sulfation) and 6th (O-sulfation) carbon atom of GlcN and 2nd carbon atom of IdoA. Therefore, the net charge for HS alters between -1 and -4, and for HP, it is -4. Functions of HS include ECM organisation and modulation of cellular growth factor signalling by acting as a bridge between receptors and ligands. In the ECM, HS interacts with collagen, laminin, and fibronectin, supporting cell-to-cell and cell-to-extracellular matrix adhesion [48]. HS also has a role in cellular growth factor signalling. Moreover, HS facilitates the formation of FGF-FGFR complexes, resulting in a signalling cascade that leads to cellular proliferation. The construction of these complexes depends on the degree of HS sulfation [48]. HP represents the

earliest recognised biological role of GAGs for its use as an anticoagulant. The mechanism for this role involves its interaction with the protein antithrombin III (ATIII). The ATIII-HP interactions cause a conformational change in ATIII, enhancing its ability to function as a serine protease inhibitor of coagulation factors. To exhibit various clinical anticoagulation activities, different molecular weights of HP were investigated [59]. Interactions between HP/HS and proteins mediate diverse pathophysiological processes, including blood coagulation, cell growth and differentiation, host defence and viral infection, lipid transport and metabolism, cell-to-cell and cell-to-matrix signalling, inflammation, angiogenesis, and cancer [70-73]. Therefore, understanding HP/HS-protein interactions at the molecular level is of fundamental importance for biology and should aid in developing precise glycan-based therapeutics [70, 74].

KS does not have any uronic acid in its structure. KS comprises the basic repeating disaccharide D-galactose β 1–4 linked to GlcNAc-6-sulfate [75]. Besides, in its polysaccharide chain, KS can contain variable numbers of mannose, fucose, sialic acid and N-acetylgalactosamine that account for its heterogeneity. KS has been well-researched for its functional role in the cornea and the nervous system. The cornea comprises the body's richest known source of KS, followed by brain tissue [76]. The degree of sulfation of KS is a determining factor of its functional status. Abnormal sulfation patterns of KS because of specific genetic mutations result in increased cornea opacity and lead to visual disturbances [76]. Also, KS plays a central regulatory role in neural tissue development.

Given the biological roles of GAGs, they are significant players in the medical treatment of disorders associated with disruptions of the processes, as mentioned earlier, including cancer [77], autoimmune diseases [78], Alzheimer's disease [79], Parkinson's disease [80] and arthritis [81]. Recently, it has been reported that HP binds to the spike glycoprotein of severe acute respiratory syndrome coronavirus (SARS-CoV), Middle East respiratory syndrome coronavirus (MERS-CoV) and to both monomeric and trimeric spike of SARS-CoV-2, responsible for the COVID-19 pandemic [82]. GAGs also represent one of the critical targets for regenerative medicine [83-85]. Additionally, GAGs might be subjected to potential chemical modifications, which could expand their properties. Such changes might involve sulfation [86-88] or introducing other negatively charged groups (for example, phosphate groups [89]) in place of sulfates. These modifications are performed to attenuate the strength and specificity of protein-GAG interactions and, in turn, enhance the inhibition or triggering of a protein's activity.

1.2.2. The role of metal ions in GAG-containing systems

Metals are omnipresent in nature and account for prodigious chemical diversity. From element 1 (H) to element 109 (Mt), there are 84 metals, seven metalloids and only 18 nonmetals in the periodic table. Al, Fe, Ca, Na, K, Mg, and Ti account for approximately 25 % of the earth's crust. Due to their low electronegativities, metals are easily ionised and highly reactive, enabling them to participate in many catalytic processes. Metals and their ions are widely distributed and play critical roles in biochemistry, chemistry, geochemistry, material sciences and countless biological processes. Approximately 30-40 % of biomolecules require metal ions to perform their functions [90]. Metals in biomolecules can play (i) a structural role when they contribute to the proper folding and confer stability to proteins or RNA structures [91], (ii) a functional role when they promote catalysis [92], a regulatory role when they trigger a metal induced-structural/dynamical response of the hosting protein to simulate/initiate a specific cellular process [93] or maintain metal availability [94]. Roughly one-third of the structures in the PDB contain metal ions [95].

The divalent calcium (Ca^{2+}) and magnesium (Mg^{2+}) ions have essential regulatory roles in cells. Lack of iron is frequent in cancer patients and is linked to surgery complications and animal experiments [96]. Moreover, metal ions participate in about one-third of enzyme interactions [97]. These ions can modify electron flow in either enzyme or substrate and effectively control an enzyme-catalysed reaction. They can assist in binding and orientating substrate concerning functional groups in the enzyme's active site [98]. Copper is identified as an indispensable metal element and is predominantly associated with copper-dependent cellular enzymes. Metal ions serve in countless metalloenzymes, are integrated into pharmaceuticals, and are utilised as inorganic drugs for many diseases [99-100]. Additionally, industrial catalysis depends on metal chemistry since more than 80% of currently used large-scale chemical processes lean on solid catalysts, mainly based on transition metal (TM) chemistry [101].

GAGs interact with several metal ions in the ECM, mediating countless biological processes in which they are involved. They display ion-specific interactions with monovalent and divalent ions [102], such as Na^+ , K^+ , Ca^{2+} and Mg^{2+} [103-104]. Multiple studies reported that HA changes its molecular conformations when interacting with Ca^{2+} [105-108]. Interactions between HP and a wide range of metal ions were inspected experimentally using atomic absorption and spectrophotometry, and the overall trend for the HP-metal affinity was reported to be: $\text{Mn}^{2+} > \text{Cu}^{2+} > \text{Ca}^{2+} > \text{Zn}^{2+} > \text{Co}^{2+} > \text{Na}^+ > \text{Mg}^{2+} > \text{Fe}^{3+} > \text{Ni}^{2+} > \text{Al}^{3+} > \text{Sr}^{2+}$ [109]. Given

that HP is a strongly anionic polyelectrolyte, it has long been recognised that HP, and other GAGs, are ideal candidates for a broader scope of functions, including the binding and releasing cations [110] regardless of if these are the micro-ions (e.g., metal cations and cationic dyes) or the macro-cations (e.g., the basic polypeptides or proteins). Additionally, the conformation of sulfated iduronic acid IdoA(2S) residues is sensitive to the identity of adjacent residues and their substitution patterns in the fragment of HP and HS [111-114]. The glycosidic linkage geometry is also affected by an altered substitution pattern [115], affecting the electrostatic properties of the polysaccharides and, therefore, their interactions with ions. Especially Ca^{2+} , which can bind somewhat reversibly to multi-sulfate and -carboxylate centres of heparinoids, may trigger macromolecular conformation changes, leading to biological activity effects [116]. Subsequently, the crucial role of metal ions in protein-GAG interaction was shown experimentally for a series of protein-GAG systems. The Alzheimer's disease-related A4 amyloid precursor protein (APP) regulates cell growth, neurite outgrowth and adhesiveness through binding to HS proteoglycans. The binding of both Zn^{2+} and Cu^{2+} to APP was previously localised and suggested to control APP formation and stability [117-118]. Multhaup *et al.* showed that APP binds in a time-dependent and saturable manner to the GAGs-chains of proteoglycans but not CS alone [119]. Furthermore, they demonstrated an interaction between the high-affinity HP binding site within the carbohydrate domain of APP and the APP's Zn^{2+} binding site. Moreover, the affinity toward HP was increased two- to four-fold in the presence of micromolar Zn^{2+} . Micromolar concentrations of Zn^{2+} modulated the binding of APP to the HP-binding site of proteoglycans and, as shown before by Bush *et al.*, induced the aggregation of soluble amyloid βA4 protein [120]. The effects of divalent cations on HP binding, structure, and thrombin inhibition rates of HP cofactor II (HCII) were examined by Eckhert and Raag [121]. Experiments with affinity chromatography and surface plasmon resonance (SPR) showed that Zn^{2+} , to a lesser extent Cu^{2+} and Ni^{2+} , enhanced the interaction between HCII and HP. Moreover, metal chelate chromatography and increased intrinsic protein fluorescence in the presence of Zn^{2+} indicated that HCII has metal ion-binding propensities suggesting that protein-GAG-ion systems are especially challenging to understand since ions can bind to both protein and GAG parts of the corresponding complexes. Altogether the findings supported the hypothesis that Zn^{2+} induces a conformational change in HCII that favours its interaction with HP. Ricard-Blum and co-workers aimed to determine the kinetics and affinity of endostatin-HP/HS interactions and investigate the effects of divalent cations on these interactions and the biological activities of endostatin [122]. The binding of endostatin to

HP and HS indeed required the presence of divalent cations. Adding Zn^{2+} to endostatin enhanced its binding to HP and HS by roughly 40%. Zhang *et al.* used SPR to investigate the effects of metal ions (under physiological and non-physiological concentrations) on protein-HP/HS interactions [123]. The results revealed that under non-physiological metal ions concentrations, various metal ions presented different effects on HP binding to fibroblast growth factor-1 (FGF1) and interleukin-7 (IL7). While individual metal ions at physiological concentrations had little impact on protein binding, the mixed metal ions reduced the FGF1-HP or IL7-HP binding affinity, changing its binding profile. Divalent cations are critical in regulating HP's anti-Factor Xa activity. There is compelling evidence that Ca^{2+} , Cu^{2+} , and Zn^{2+} are essential in several protein-HP interactions, influencing these complexes' affinity, specificity, and stability [124-125]. A previous study by Shao *et al.* showed that the conformational changes induced by Ca^{2+} are necessary for the interaction between HP and Annexin II [126]. Ca^{2+} are experimentally detected in the annexin-HP interface (PDB ID: 1G5N [127], 2HYV [126] and 2HYU [126]). In the study of Weiss *et al.*, the addition of EDTA and BAPTA strongly inhibited the binding of PCPE-1 and its NTR domain to immobilised GAGs, indicating that the binding to PCPE-1 and the NTR domain to HP/HS depends on the divalent cations [128]. BAPTA inhibited the binding to a greater extent than EDTA, suggesting that Ca^{2+} takes part in the binding. The binding of PCPE-1/NTR to HP/HS requires divalent cations and most likely Ca^{2+} , as reported previously for other extracellular proteins such as endostatin, a fragment of collagen V, the $\alpha 5\beta 1$ integrin and certain annexins. However, the critical role of Ca^{2+} in binding the NTR domain of PCPE-1 to HP remains unclear.

1.2.3. Cyclodextrins

Cyclodextrins (CDs) belong to the class of cyclic oligosaccharides, and the most common ones are α -, β - and γ -CD formed by 6, 7 and 8 α -(1,4) D-glucopyranose units, respectively (Figure 3).

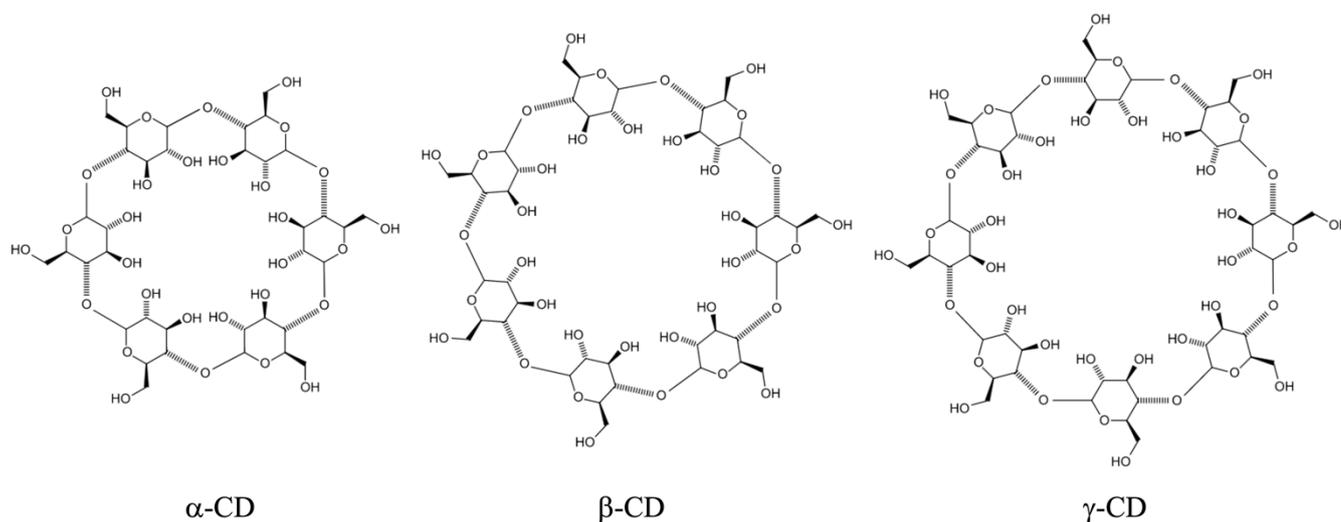


Figure 3. Chemical structures of α -, β -, and γ -cyclodextrins [129].

CDs are obtained from the biodegradation of starch using glucanotransferase enzymes. Since the glucopyranose unit possesses chair conformation, the CD is not cylindrical but exhibits a torus-like shape (truncated cone). The primary hydroxyl groups are present on the narrow edge, and secondary hydroxyl groups are on the broader edge of the cone on the outer surface. Skeletal carbons with hydrogen atoms and oxygen bridges are present in the interior of the CD molecules accounting for the hydrophilic outer surface and lipophilic central cavity of CD [130-132]. The internal cavity's size varies between 4 Å and 8 Å, depending on the number of sugar units forming the macrocycle. The most relevant physicochemical properties of native α -, β -, and γ -CD are summarised in Table 2. The exceptional features of CDs are their tendency to form inclusion complexes (also known as host-guest complexes) with a wide variety of solid, liquid, and gaseous compounds through molecular complexation. Such type involves a guest molecule (drug) which is correctly fitted into the lipophilic cavity of a host (CD molecule) [133]. During the formation of an inclusion complex, there is neither breakage nor formation of covalent bonds [134-135]. When the CD is added to an aqueous solution, the polar water molecules enter the lipophilic cavity of the CD. Still, they are immediately replaced by the more favoured guest molecule [136]. The formation of the inclusion complex is exothermic

and entropy-controlled but not entropy-driven. The main driving force of this process is the release of enthalpy-rich water molecules from the cavity. Water molecules are displaced by more hydrophobic guest molecules present in the solution to attain an apolar-apolar association and decrease CD ring strain resulting in a more stable lower energy state [137].

The inclusion complex is also stabilised by hydrogen bond formation, electrostatic and van der Waals interactions, hydrophobic effect, and in the case of α -CD, changes in conformational tensions [138]. Self-assembly is the phenomenon where simple molecules spontaneously organise into a complex system. The self-organisation happens via several interaction mechanisms: unspecific electrostatic π - π interactions, dispersion, hydrophobic interactions, or more specific bindings such as the host-guest complexation or ‘lock-and-key’ binding [139-141].

Table 2. Basic physicochemical properties of native α -, β -, and γ -CD. Data reported from references [142-143].

	α -CD	β -CD	γ -CD
Glucopyranose units	6	7	8
Chemical formula	C ₃₆ H ₆₀ O ₃₀	C ₄₂ H ₇₀ O ₃₅	C ₄₈ H ₈₀ O ₄₀
Molar mass [g/mol]	972	1134	1296
Solubility in water	129.5	18.4	249.2
Outer diameter [nm]	1.52	1.66	1.77
Cavity height [nm]	0.78	0.78	0.78
Cavity diameter [nm]	0.45-0.53	0.60-0.65	0.75-0.83
Cavity volume [nm ³]	0.174	0.262	0.427
Number of bound water molecules	6.4	9.6	14.2
Number of water molecules inside the cavity	3.6	6.3	8.9
Density [g/cm ³]	1.48	1.44	1.52

Multiple studies present that CD encapsulation can improve the following guests' properties.

- *Water solubility.* Many potential drugs have poor water solubility and stability and, therefore, cannot be transported in the body, significantly limiting their applications. Because CDs are hydrophobic inside, low-polarity guests easily bind them forming an inclusion complex. Hence, the hydrophilic outer surface confers new physicochemical properties to the guest molecules. The most remarkable is the improvement of the water solubility of guests described in [144].
- *Transportation of active substances.* The hydrophilic groups of the surface of the CD-guest inclusion complex can interact with the phospholipid bilayer of the cell membrane. The complex (CD-guest) can enter the cell and release the guest under specific conditions [145].
- *Biological activity.* Improved antioxidant, anticancer and antibacterial activities of some natural compounds due to the presence of CD were reported by Zhong *et al.* [146-148].
- *Bitterness masking.* Many natural compounds are known for their bitter taste, which significantly limits their applications in food. CDs are slightly sweet substances [149] and can form a barrier in the taste buds. Therefore, guests' undesirable flavours can be partially or entirely reduced or removed by inclusion.
- *Facilitate compound extraction.* Organic solvents are frequently used extractants for compound extraction, which may cause environmental pollution and harm human health. Using green extracting agents such as CDs is one solution to this issue [150].

Additionally, CDs are biocompatible, do not elicit an immune response, and have low toxicity in animals and humans. Therefore, they are increasingly used in the pharmaceutical, food and material industries [129, 151-153].

It should be mentioned that CDs can have more than 8 glucose units, and those species are referred to as large-ring CDs (LR-CDs). LR-CDs comprising up to 31 glucose units have been purified and characterised [154]; the existence of even larger CDs with a degree of polymerization (dp) of up to several hundreds of glycosyl units was reported [155]. In 2002, an extensive review of LR-CDs was published by Larsen [155].

1.2.4. The role of surfactants in CD-containing systems

The formation of CD-surfactant inclusion complexes is a very active area of research. Not only due to their high relevance to the cosmetic, food, and medical industry but also because the mixtures tend to form highly ordered supramolecular aggregates, resulting from the delicate balance of many interaction types.

Surfactants are amphiphilic molecules meaning they consist of both hydrophilic and hydrophobic moieties. Therefore, they can reduce the surface tension between two immiscible phases [156], and they can be obtained either from biological materials (biosurfactants) or chemically (synthetic surfactants). The hydrophilic part includes heteroatoms such as oxygen, nitrogen, sulfur, and phosphorus, which occur in the following functional groups: amine, amide, alcohol, thiol, ester, ether, acid, sulfate, sulfonate, phosphate, etc., whereas the hydrophobic part is usually paraffin, cycloparaffin or aromatic hydrocarbon, which may have halogens. Because of their dual affinity, surfactants are unstable in organic or polar solvents. Based on their chemical composition, surfactants can be divided into ionic and non-ionic, and the former can be either cationic or anionic, depending on their charge [157-158]. The endless diversity of available surfactants and their responsive properties to different stimuli provide countless possibilities for colloidal scientists aiming to spontaneously form complex materials from simple building blocks. Host-guest supramolecular complexes formation involving CDs and surfactant molecules is predominantly driven by non-covalent interactions [159]. The formation of the covalent bond between the host and the guest is avoided to enable the guest to escape freely from the CD cavity. CD-surfactant inclusion complexes are well-studied [160-164]. Although the size-match compatibility between the host and guest is a simple concept and sometimes sufficient to predict potential interactions, the CD-surfactant complex's exact molecular stoichiometry and conformation should be assessed individually for each case. The most frequently observed stoichiometries (CD: surfactant) are 1:1, 2:1 and 1:2 and depend on the alkyl chain length of the surfactant and cavity size [165-166]. Generally, a 1:1 stoichiometry is seen for surfactants with a shorter alkyl chain (eight or fewer carbon units). The 2:1 stoichiometry is most pronounced for surfactants with longer alkyl chains (twelve or more carbon units). The stability of these complexes is dictated by the interactions between the rim of the CD and the surfactant head group [167-168]. In the case of gemini surfactants (composed of two hydrophilic head groups and two hydrophobic tails linked by a spacer at the head groups or closed to them), two CDs interact with one alkyl chain giving 4:1 stoichiometry

[169]. In the case of α - or β - CD, it is unlikely to host two surfactants simultaneously. However, this event may occur for the γ -CD because its large cavity can accommodate two surfactants. The 1:2 stoichiometry was also observed when the excess surfactant concerning the CD concentration was present [170]. Furthermore, the accommodation of large-sized guest molecules in α -CD is difficult due to its small cavity comprising six glucose molecules, decreasing its application [171]. On the other hand, γ -CD, composed of 8 glucose units, has a larger cavity area and cannot be mass-produced by industry based on its enormous production cost; hence, it is not broadly used either [172]. Finally, the β -CD, which consists of 7 glucose units and has an intermediate cavity area and a decreased production cost, is the most commonly used CD in the industry [173]. Nonetheless, the cavity of β -CD is spatially limited, and so is its application. The β -CD skeleton is kept unchanged to overcome this issue, but the various functional groups are added to change its chemical and physical properties [174]. In these lines, several different chemically modified CDs were derived, including dimethyl- β -CD (DM- β -CD), hydroxypropyl- β -CD (HP- β -CD) and sulfobutylether- β -CD (SBE- β -CD). The ionic character of the surfactant head group, the alkyl chain length, and the host's conformation influences the CD-surfactant complexes' thermodynamic binding parameters. The series of n-alkyl trimethylammonium bromide and n-alkyl sulfates are the most studied in terms of the surfactant ionic character [161, 175]. The study by Ondo showed the small effect of the head group in the 1:1 stoichiometry complex, and the slightly reduced stability of those complexes was observed when the SO_4^- group was substituted with cationic groups. The impact of the head group was also predicted for the 2:1 stoichiometry complex. Therefore, if energetically favourable, a stable inclusion complex is formed [163]. The formation of inclusion complexes between CDs and non-ionic surfactants has also been investigated [176-179]. The effect of the alkyl chain length on the thermodynamic parameters of surfactants has been explored in-depth [163, 180-183]. The longer the alkyl chain, the stronger the hydrophobic character of the guest and, therefore, the stronger the interaction, which was observed by an increase in the binding constant. The chain length of six carbons for ionic and eight for non-ionic surfactants were described as the critical chain length for the CD-surfactant complex formation [182]. The stronger influence of the head group in shorter chain lengths is observed. Furthermore, for hydrocarbon chains longer than fourteen, the increase in the hydrophobic character is less pronounced. The effect of counterions (ions to maintain electric neutrality in the system) in the case of ionic surfactants cannot be ignored. The ions in solutions were demonstrated to interact with native CDs [152]. Studies conducted by Valente and Söderman showed that for a set of

alkyltrimethylammonium halogenides (C_nTAC)-CD complexes, the interactions depended little on the counterion type (either Cl^- or Br^-), based on the stability constants (K) values. Similarly, Junquera *et al.* reported that Br^- , from $C_{12}TAB$ participated in the association process by binding to β -CD and hydroxypropyl- β -CD (HP [184]. Macpherson and Palepu determined Na^+ activities in sodium dodecylsulfate and sodium dodecanoate solutions containing β -CD [185]. However, the effect of counterions on binding constants and thermodynamic parameters of ionic surfactant complexes with CD is relatively small [162, 186].

1.3. The role of ions in albumin-containing systems

In this chapter, the nature of interactions between albumins, anionic surfactants and other small molecules are reviewed.

1.3.1. The role of surfactants in protein-containing systems

The interactions of proteins with surfactants were thoroughly examined due to their applications in the cosmetic, food, and pharmaceutical industries [187-190]. Ionic surfactants interact strongly with protein and form protein-surfactant complexes in the aqueous solution, resulting in protein denaturation at lower surfactant concentrations relative to other denaturants such as guanidinium chloride or urea. The protein-ionic surfactant interactions are predominantly driven by either hydrophilic (between the charged head group and oppositely charged patches of the protein chain) or hydrophobic (between hydrophobic patches on the protein and surfactant hydrophobic tails) or van der Waals interactions [190-191]. However, the relative importance of these types of interactions is still being investigated [192-193]. An increase in the surfactant molecule tail length increases the tendency of protein denaturation, suggesting the strong impact of the hydrophobic interaction on the surfactant-induced unfolding of the protein [194]. Simultaneously, non-ionic surfactants either do not or weakly interact with globular proteins, demonstrating the crucial role of electrostatic interactions [195]. Lately, MD simulations indicated that the electrostatic and hydrophobic interactions are equally vital in protein unfolding regarding the tail insertion into protein structure [196]. Investigation of protein-surfactants (both synthetic and microbial) interactions is of considerable interest in numerous areas of biotechnology and industry, including biomedicine, cosmetics, food, pharmaceutical, medical, and environmental sectors [157, 197-199]. Previously listed three main driving forces (hydrophilic, hydrophobic, and van der Waals interactions) in protein-surfactant systems are determined by the nature of both molecules and their concentrations [191, 200]. Moreover, these molecular interactions affect the native structure of proteins by either promoting or preventing processes of aggregation, denaturation, or enzymatic activity [200]. Interestingly, biosurfactants can halt the denaturation of proteins or reduce their aggregation [156-157]. Most studied protein-surfactant systems consist of globular proteins such as BSA, α -lactoglobulin and β -glucosidase. On the other hand, little research focused on the fibrous protein-surfactant systems. Some of them, keratin, silk fibroin and type I collagen, were investigated in the presence of ionic and non-ionic surfactants [201-

207]. Some researchers found that the molecular interactions between fibrous proteins and ionic/non-ionic surfactants resemble those of globular proteins-surfactant systems [158, 197, 204-206]. Parameters that have a determinant effect on the protein-surfactant interactions are 1) the chemical nature of a surfactant (ionic/non-ionic); 2) the surfactant's concentration; 3) the secondary structure of the protein (α -helix and β -sheets) [208-209]. The hydrophilic group of the surfactant impacts the stability of the protein since it can bind tightly to the protein, leading to its denaturation and contributing to the solubilisation of the membrane proteins [200]. Anionic surfactants are usually protein-denaturing agents [158]. One of the best-known anionic surfactants for having strong electrostatic interactions with proteins is sodium dodecyl sulfate (SDS) [157, 195, 210-211]. These interactions are established between the positively charged amino acids (LYS, ARG) present in the protein's primary structure with the polar heads of surfactants and between their aliphatic regions with the hydrocarbon chains of the surfactant [195].

1.3.2. Albumins

Serum albumin (SA) belongs to the albumin superfamily together with alpha-albumin, alpha-fetoprotein, and D-binding protein (VDP) [212]. This family is present exclusively in vertebrates, and SA can be found in mammals, birds, some species of frogs, lampreys, and salamanders (the complete list is available on albumin.org [213]). Albumin is the major circulating protein in healthy adults (with an average physiological concentration of 0.6 Mm). It is synthesised in the liver and responsible for 80 % plasma colloid osmotic pressure (COP). Depending on the species, albumin is a single-chain polypeptide chain of 580 to 585 amino acid residues. It is a non-glycosylated polypeptide with hydrophobic cavities and patches, lacking prosthetic groups. In X-ray crystallography, albumin is depicted as a heart-shaped tertiary structure and possesses 17 pairs of disulfide bridges. The protein comprises three homologous domains in structural features and has only one free cysteine. Albumin can bind to the majority of known drugs and many nutraceuticals and toxic substances, mainly determining their pharmaco- and toxicokinetics [45]. Human albumin and its respective representatives in cattle and rodents have structural features that distinguish species differences in functional properties. Albumin is both passive and active in pharmacokinetic and toxicokinetic processes, possessing several enzymatic activities. When a new drug is being developed, testing for its binding to albumin is a standard procedure. Routinely, the bovine serum albumin (BSA) is used in toxicological and pharmacological experiments *in vitro* as a human serum albumin (HSA) model because of its 88% resemblance in structure, availability, and low cost. This heart-shaped protein consists of 583 amino acid residues, and its molecular weight is approximately 66.8 kDa. The whole system is divided into three homologous domains (I, II and III), and each domain is subdivided into A and B subdomains with unique binding properties [214]. In the subdomain IB and subdomain IIA, BSA has two tryptophan (TRP) amino acid residues, TRP 134 and TRP 213, respectively (Figure 4). BSA shows discrete binding sites with different specificities, the most vital being site-I and site II, located in hydrophobic cavities of subdomains IIA and IIIA, respectively [215]. Site markers are small molecules that have specific binding locations in albumin structure and are often used in investigating the interactions of various ligands with the protein. To site-I markers belong warfarin, phenylbutazone, dansylamide and iodipamide, whereas ibuprofen, flufenamic acid and diazepam are site-II markers [217].

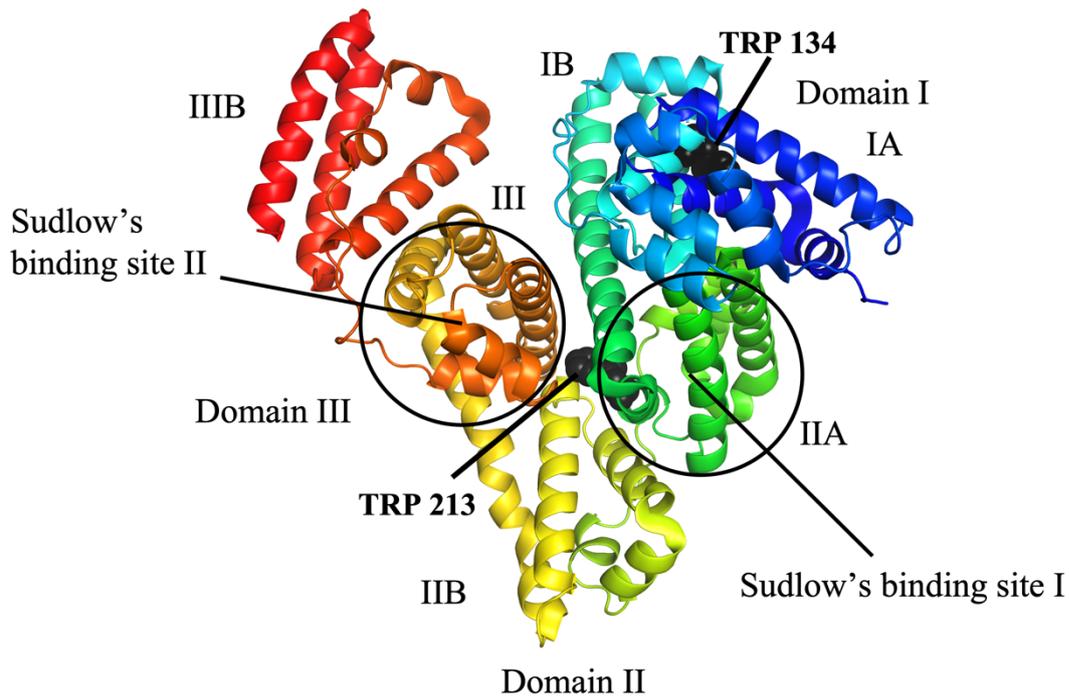


Figure 4. Structure and binding sites of bovine serum albumin (BSA, PDB ID: 4F5S) [217]. Residues TRP 134 and TRP 213 are depicted as spheres in black.

2. Methods

The following two chapters, *2.1.* and *2.2.* describe experimental and computational approaches applied to characterise receptor-ligand interactions, respectively.

Many experimental techniques are deployed to examine various aspects of receptor-ligand interactions. X-ray crystallography, nuclear magnetic resonance (NMR), small-angle-X-ray scattering (SAXS), and cryo-electron microscopy deliver atomic-resolution or near-atomic-resolution structures of unbound proteins and the protein-ligand complexes, which are used to investigate the changes in structure and dynamics between the free and bound forms as well as relevant binding events. Determining the structures of the protein targets and, in particular, their complexes with lead compounds serves as the starting point for most of the computational tools outlined in the next chapter. They are also crucial components for rational drug design, enabling one to see what part of the ligand interacts with which residues of the target protein and how the ligand could be expanded or modified. Moreover, the deployment of structural biology techniques in drug design increased significantly over the last decade due to significant technological advances and the popularity of computational approaches that rely on the implementation of the available experimental structural data. Other experimental methods applied to study receptor dynamics involved in binding, including mass spectroscopy (MS), circular dichroism (CD), fluorescence spectroscopy, surface plasmon resonance (SPR), enzyme-linked immunosorbent assay (ELISA), and isothermal titration calorimetry (ITC), are also described.

Experiments can provide comprehensive information regarding receptor-ligand structures and interactions and accurate data on thermodynamics and kinetics. However, they are laborious, time-consuming, and expensive. Therefore, in many cases, additional computational studies are conducted to look at the system from a different perspective and complement experimental findings by providing new insights at the atomic level. Moreover, the theoretical analysis might improve the quality of understanding of the results obtained by the experimental methods as described below.

- Structure-based computational approaches are helpful in all aspects of investigating receptor-ligand binding events. For instance, if the experimental structure of a protein is not available, theoretical tools like homology modelling, threading or *ab initio* prediction allow for the construction of structural models that can be used to predict receptor-ligand binding.

- Whereas molecular docking methods can quickly predict the most favourable structure of the complex and assess the binding affinity, molecular dynamics (MD) simulations, allow the description of the dynamic behaviour of a molecular system. In this way, MD acts as a “computational microscope” that offers a “real-time visualisation” of events, including protein folding, protein conformational changes and protein-protein interactions regarding the flexibility of the molecules and the possible conformational changes caused by mutations in different environments (pH, temperature, or salt concentration). Moreover, MD simulations can be performed in an all-atom (AA) or coarse-grained (CG) model, depending on the size of the system. The AA model is recommended for small molecules such as ligands, peptides, protein subdomains and sometimes whole proteins. AA model is also a choice for glycans, lipids and amphiphilic polymers. Larger molecules, such as relatively big proteins or networks of multiple proteins, are handled by CG modelling. The number of degrees of freedom of biomolecules is reduced in CG modelling; therefore, it usually requires a much lower computational cost than AA.
- Free energy calculations offer more accurate binding affinity prediction compared to the binding strength estimations provided by molecular docking. They consider all thermodynamically relevant phenomena, such as the receptor dynamics/flexibility, explicit inclusion of the solvent, and the difference between receptor-ligand interactions in the complex and their interactions with water and counterions in their unbound forms.

Lastly, the challenges related to modelling GAG- and metal ion-containing systems are also reviewed at the end of this chapter in section **2.2.5**.

However, these sections do not represent an exhaustive overview of existing experimental and computational methodologies in the field of receptor-ligand interactions. Still, they provide a general introduction to the methods used in this Thesis and the theoretical concepts behind them.

2.1. Experimental approaches to characterise interaction in receptor-ligand systems

2.1.1. X-ray crystallography

X-ray crystallography is one of the most used experimental methods to study crystal structures through X-ray diffraction (XRD) techniques and is a powerful way of obtaining detailed information on structure and bonding within a crystal. The X-ray beam is diffracted upon its collision with atoms into smaller beams, and then the angles and intensities of these diffracted beams are measured. Consequently, a three-dimensional map of the density of electrons within the crystal is obtained to which individual atoms are assigned to acquire the complete structure of the compound [218]. The main culprits of this method are the high cost of a single experiment and the requirement of having the compound in a crystal form. Furthermore, due to the crystallised form of the studied macromolecule, the resulting structure might not always correspond to its native conformation as the influence of the natural living organism-environment is not considered. Moreover, the crystallographic resolution, defined by the crystal's distribution of the studied molecule, determines the structure's quality.

In addition, studying GAG-containing systems with X-ray crystallography remains an arduous task. Due to the very flexible nature of GAGs, it might not be possible to crystallise a desired protein-GAG complex of required purity. Although enzymatic cleavage of GAGs allows obtaining a specific chain length, the composition of the polysaccharide chain is rarely well-defined [219-221]. At the same time, the chemical synthesis of GAGs involves a complex series of reactions using specific protecting groups to obtain the intended product [222]. Despite all that, identifying and characterising the GAG binding sites in some protein-GAG systems using X-ray crystallography was successful [223]. In addition, Baron *et al.* demonstrated that X-ray crystallography could be a valuable tool to explain the biological role of protein-GAG interactions [224]. According to Glycosaminoglycan Database (GAG-DB) created by Perez *et al.* [225], there are 113 X-ray structures of GAG-containing structures available in PDB. The included structures' highest and lowest resolutions are 1.25 Å (PDB IDs: 1RWH, 2JCQ) and 3.45 Å (PDB ID: 4NDZ), respectively. Of 113 known X-ray structures, 106 are protein-GAG complexes, with 92 containing non-modified GAGs. Among these 92 structures, the length of oligosaccharides spans from dimers to decamers. However, only three X-ray structures (PDB ID: 1G5N, 2HYU, 2HYV) corresponding to three various protein-Ca²⁺-

HP systems (including a GAG bound to and containing ions in the GAG binding interface) are available [225].

Furthermore, X-ray structures of several CD crystalline complexes have revealed the presence of guest molecules in the host macrocyclic cavity [226]. Crystal structures of the following inclusion complexes were reported: β -CD-cholesterol [227], two crystal structures of a hydrated 2:1 β -CD-fluconazole complex [228], β -CD-bithiophene [229], β -CD-7-hydroxycoumarin and β -CD-4-hydroxycoumarin [230]. Additionally, there are several CD-drug complexes for which XRD provided structural details. Some of these cases include the following: inclusion of S-ibuprofen in β -CD [231], inclusion complexes of essential oil in α -CD and β -CD, and ibuprofen combinations with substituted β -CD [232], trimethoprim- β -CD [233], omeprazole- β -CD [234], and p-aminobenzoic acid- β -CD [235].

X-ray crystallography is a leading technique in investigating albumin structures and albumin-ligand complexes reflected in the number of readily available PDB structures. According to the PDB database, there are 193 X-ray structures for albumin (alone or with various ligands), including 140 and 6 for HSA and BSA, respectively.

2.1.2. Nuclear Magnetic Resonance

Nuclear magnetic resonance (NMR) spectroscopy is the only technique that provides atomic-resolution structures in solution. It provides information on chemical compounds' structural and dynamic characteristics, reaction state, and chemical environment. During an NMR experiment, a magnetic field is applied to the sample, affecting the nuclei spin. The energy released by these nuclear spins returning to their original states is detected and analysed. The energy differs for the same atomic nucleus in different chemical environments, either between other compounds or between the bound and unbound states. The intramolecular magnetic field around an atom in a molecule changes the resonance frequency depending on its chemical environment, offering access to details such as the electronic structure of a molecule and its functional groups [236]. NMR is broadly used in organic chemistry to confirm the identity of a substance. Since the measurement of a sample is conducted in solution, it is possible to study proteins in an environment close to the biological one.

However, the diversity in molecular arrangements, periodicity and dynamics displayed by glycans renders traditional NMR strategies employed for proteins and nucleic acids insufficient [237]. Due to the unique properties of GAGs, structural studies often require the adoption of a different repertoire of tailor-made experiments and protocols. Experiments using isotopic labelling may tackle challenges associated with spectral overlap and increase sensitivity. Multinuclear NMR studies also proved valuable in studying the interactions of the HP-metal ions. ^1H and ^{23}Na NMR spectroscopy demonstrated that Na^+ , Ca^{2+} and Mg^{2+} interacted at low pH with the carboxylic acid form of HP by long-range electrostatic interactions [238]. At higher pH and subsequent deprotonation of the carboxylic acid, a site-specific contribution to the binding of Ca^{2+} , Zn^{2+} and La^{3+} was observed. Furthermore, the release of HP-associated Na^+ in the presence of competing cations can be studied by ^{23}Na NMR spectroscopy. The results also suggested site-specific binding for Ca^{2+} and Zn^{2+} but not Mg^{2+} . Lerner *et al.* measured NMR relaxation rates of ^{23}Na , ^{29}K , ^{25}Mg and ^{43}Ca ions with bovine nasal cartilage proteoglycans and hog mucosal HP. They deduced that relaxation rates were determined mainly by polymer concentrations and charge density. HP binds monovalent and divalent cations to a much greater extent compared to proteoglycans [239]. The binding preferences in solution for Ca^{2+} have been delineated and displayed similarities to those reported in the crystallographic studies for protein- Ca^{2+} -HP systems [125-126, 240]. Chevalier *et al.* investigated the specific binding of Ca^{2+} to synthetic hexasaccharide models of modified HP.

They found that the carboxylate groups of the iduronate residue and the N-sulfate and 6-O-sulfate of GlcNS are crucial for Ca^{2+} binding. Sulfate selectivity and conformation are affected – the sulfate at position 2 of IdoA(2S) in a synthetic hexasaccharide is not essential for binding. Still, specific binding is similar to the regular HP dp6, whereas Ca^{2+} binds rather weakly when the substrates lack the 6-O-sulfate of glucosamine [125, 240].

Unquestionably, NMR is also a powerful technique to elucidate the binding constant and stoichiometry of the complexation reaction in surfactant-CD systems. The method is, however, limited to systems where the signal from the surfactant can be separated from those of CD [162, 241]. A different approach to determining the binding constant offered by NMR diffusometry is to exploit the free compounds' and complex's different diffusion coefficients [167, 242-243]. NMR has also been reported as an essential tool for the structural characterisation of more complex inclusion systems. 2D NMR spectra offer information on the particular vicinity of the different functional groups, enabling the precise determination of the relative position of the host and guest molecules [244-246].

2.1.3. Other experimental techniques

Mass spectrometry (MS)

Another powerful technique used to characterise protein-ligand interactions is mass spectrometry (MS). In this method, the tested sample is ionised either with the beam of electrons (electron ionisation EI), the application of high voltage to liquefy the sample to create an aerosol (electrospray ionisation, ESI) or with the previously ionised by EI reagent gas that subsequently reacts with the tested sample. MS offers several key advantages over NMR spectroscopy and other experimental techniques. These are superior sensitivity, faster speed and the ability to monitor the exchange in molecular complexes, enabling the detection of multimolecular complexes' existence.

The defined sequence of HP oligomers offers a level of simplification suitable for electrophoresis in combination with electrospray ionisation mass spectrometry (ESI-MS), and tandem MS is frequently utilised in sequencing studies [247]. The formation of glycosidic bond cleavages from GAG anions is balanced compared to competing processes for both low and high-charge states. In the case of low charge states, losses of sulfate groups result in the most abundant product ions in the tandem mass spectra [247] for binding. Recently Pepi *et al.* published a detailed review where the developments in MS for GAGs analysis are reported. MS techniques, especially ESI and matrix-assisted laser desorption ionisation (MALDI), are powerful tools for studying host-guest complexes [248]. ESI-MS permits the transfer of complex ions from the liquid or solid phase into the gas phase. This enables establishing the stoichiometry of an inclusion complex, estimating the energy of the host-guest interactions [249-250] and investigating the gas-phase reactions of inclusion complexes with various ligands [251-252].

Circular dichroism

Biomolecules, including proteins, are made of chiral subunits that produce signals upon illumination by circularly-polarised light in the near and far ultraviolet wavelengths ranges where the amide and carbonyl groups of the polypeptide backbone absorb. Circular dichroism spectroscopy is an optical spectroscopic tool which exploits the differential absorption of left- and right-circularly polarised light by such chromophores. It can extract structural information

about protein conformation [253]. This method is routinely used to discern the secondary structure of a protein (far UV region) and to monitor the local tertiary structure environment of aromatic amino acid residues (near UV region). Furthermore, Matsuo *et al.* used circular dichroism to characterise the structure of GAGs in solution [254]. In contrast, Zsila proposed that HP and C6-S can form stable complexes with small molecules such as berenil and pentamidine [255]. Synchrotron radiation circular dichroism is a sensitive tool to characterise the conformation of uronic acid. Therefore, it was deployed to investigate the effect of metal ions on the conformation of modified disaccharides consisting of one IdoA(2S) unit and a linked modified monosaccharide. Rudd *et al.* reported that virtually all disaccharide/cation combinations resulted in unique spectra implying that the contribution of metal ions must be considered while looking at the conformation and flexibility of disaccharides [256]. The applications of circular dichroism were extensively reviewed in investigating CD complexes [257] and drug discovery, including albumins [258].

Fluorescence spectroscopy

Another experimental method to study receptor-ligand interactions is fluorescence spectroscopy. One of the possible applications of this method is to measure enzymatic activity in the absence and presence of GAGs. It enables us to get information about enzyme kinetics and enzyme inhibition. The substrate and the product have different adsorption spectra. Therefore, adding substrate allows for observing the changes in specific signals. This type of assay is believed to be more sensitive than spectrophotometric assays. However, it can suffer from interference caused by impurities and the instability of many fluorescent compounds when exposed to light. Fluorescence spectroscopy might provide crucial data about how the GAG binding influences a protein's enzymatic activity [259]. Fluorescence spectroscopy can provide information about the inclusion complex formation between a host and a guest by recording the change in fluorescence and intensity. It was widely deployed to investigate various CD-guest systems [260-262]. This method is a sensitive and selective tool to study the interaction between proteins and drugs/ligands [263]. The aromatic amino acids, mainly tryptophan (TRP) and tyrosine (TYR) are responsible for the intrinsic fluorescence of the proteins [264]. Due to the presence of two TRP residues in BSA, fluorescence spectroscopy is the most exploited and straightforward technique for this molecular system [265].

Surface Plasmon Resonance (SPR)

SPR is a label-free technique and does not require additional assays, reagents, or laborious sample preparation. The essential advantage of this method is that it responds to changes in the refractive index induced by molecular binding events. In SPR biosensors, the biological recognition element is immobilised at the sensor surface, and the analyte solution usually passes via a microfluidic channel across the sensor analyte interface. The biological recognition event between the analyte and biological recognition element results in a change in the refractive index close to the sensor surface, which is determined as a change in the resonance angle of the surface plasmon (excited free electrons near the surface) [266]. Compatible light energy photons provide this resonant excitation. The amplitude of the resulting plasmon electromagnetic or evanescent wave is maximal at the interface between the plasmon generating (sensor surface, usually a noble metal) and the emergent medium [267]. SPR uses changes in the refractive index adjacent to the sensor due to the binding of the analyte. SPR biosensors can detect small changes in the refractive index where biomolecules bind with the receptors immobilised on the optical transducer surface [268-269]. SPR was also used to examine the role of metal ions on HP binding to fibroblast growth factor-1 (FGF-1) and interleukin-7 (IL-7) [116]. This method deploys resonant oscillation of conduction electrons at the interface between negative and positive permittivity materials stimulated by incident light [270]. When investigating protein-HP systems by SPR, HP is preferably immobilised onto the sensor chip compared to protein since this mimics more closely natural systems where HS is found at the cell surface as a part of proteoglycan and binds to the target protein [271-272].

Enzyme-linked immunosorbent assay (ELISA)

A broad spectrum of protein-ligand systems could also be inspected with ELISA. This method uses a solid-phase enzyme immunoassay to detect whether the ligand is present in a liquid sample using antibodies directed against the measured protein. ELISA is frequently complemented with SPR measurements. With either the GAG or proteins immobilised, Hintze and co-authors found that HA derivatives interact stronger than the corresponding CS derivatives with the same sulfation degree [273-275]. Pan *et al.* used this technique to examine the strength of GAG binding to platelet-rich plasma (PrP) by obtaining optical density values [276].

Isothermal Titration Calorimetry (ITC)

This technique directly measures the heat generated or absorbed by ligand-macromolecule interactions under constant temperature. ITC can measure thermodynamic parameters of binding directly, and there is no need for labelling, immobilisation, or any other modification of the investigated interacting partners. Furthermore, ITC is highly versatile, as any binding results in either consumed or released heat. Therefore, there are no limitations based on size, optical or other properties of binders. ITC is the most sensitive tool available for the identification of a) the affinity constant (K); b) stoichiometry of the interaction (N); and c) the enthalpy change (ΔH), which represents the heat taken up or released during the reaction. Subsequently, the entropy (ΔS) and the Gibbs free energy (ΔG) can be calculated using the standard thermodynamic relationships (5):

$$\Delta G = -RT \ln K = \Delta H - T\Delta S \quad (5)$$

ITC proved to be a great tool to provide data in many areas, from cellular biology to chemistry [277-279]. It was deployed to determine the thermodynamics and stoichiometry of GAG binding [280-283] and to characterise the interaction mechanism of CDs with various guests [284-289]. Moreover, ITC was used to characterise adsorption on vesicles [290], micellar-based systems [284, 291-292], self-associating systems [293], polyelectrolyte complexes [294], nucleic acid interactions with cations [291].

2.2. Theoretical approaches used to characterise interactions in receptor-ligand systems

2.2.1. Electrostatic potential calculations

Electrostatic potential calculations have proved helpful in understanding and predicting molecular properties [295-296]. Notably, electrostatic properties analysis approaches can assist as a predictive tool of a molecule's chemical reactivity and ability to form certain types of interactions. A common way to visualise the electrostatic properties of molecules is via electrostatic potential (ESP) isosurfaces [297]. One popular method to calculate such surfaces is PBSA (Poisson-Boltzmann surface area), where the solvent is treated implicitly (potential of mean force is applied to approximate the averaged behaviour of many highly dynamic solvent molecules), and the Poisson-Boltzmann equation is implemented [298-299] (6):

$$\nabla \cdot [\epsilon(x)\nabla\phi(x)] - \kappa(x)^2\sinh(\phi(x)) = -4\pi\rho(x) \quad (6)$$

where $\phi(x)$ denotes the electrostatic potential, $\rho(x)$ is the charge distribution function, $\epsilon(x)$ is the spatial dielectric function, and $\kappa(x)$ is the modified Debye-Hückel parameter. Samsonov *et al.* demonstrated that the PBSA method was successful in 68 out of 73 cases in predicting GAG binding regions for protein-GAG complexes in a non-redundant dataset [300] due to the dominating role of electrostatics in these systems. PBSA successfully predicted the GAG binding region in TIMP-3 protein [301].

2.2.2. Molecular docking

Basic concepts in molecular docking

In molecular modelling, molecular docking is a computational approach that predicts the preferred orientation of one molecule to a second one when a ligand and a target are bound to form a stable complex [302]. In turn, knowledge of the preferred orientation (pose) may help predict the strength of association or binding affinity between two molecules. Furthermore, this technique became vital in developing various rational drug design protocols comprising structure-based virtual screening for identifying novel candidates and comprehending the essential chemical elements that guide protein-ligand interactions in relevant biological targets [303-304].

A docking procedure consists of three interrelated components:

- prediction of the binding site (defined as a region on the macromolecule that binds to another molecule);
- prediction of the binding pose (the ligand's conformation, its position and orientation within this site);
- assessment of the binding affinity corresponding to the binding poses with a scoring function.

Most docking algorithms rely on the predefined binding site, so the search space is limited to a comparatively small protein region, defined as local docking. The information about the sites can also be obtained by comparing the target protein with a family of proteins sharing a similar function or with proteins co-crystallised with other ligands. In the absence of knowledge about the binding sites, cavity detection programs and online servers, for example, GRID [305-306], POCKET [307], SurfNet [308-309], PASS [310] and MMC [311] can be used to identify putative binding sites on the protein surfaces. Docking without any assumption about the binding site is called blind or global docking.

The search algorithm thoroughly searches the potential energy landscape to find the global energy minimum. Docking applications are classified by their search algorithms, defined by rules and parameters applied to predict both the receptor and ligand conformations. When one considers the flexibility of the ligand and/or the receptor, docking algorithms are categorised into two large groups: rigid-body and flexible docking. In rigid docking, the search algorithm

explores different positions for the ligand on the receptor's surface using only the translational and rotational degrees of freedom for the ligand. Flexible ligand docking adds exploration of torsional degrees of freedom of the ligand to this process.

Docking applications usually deploy one or more of the search algorithms listed below to place the ligand within the binding site (also referred to as ligand's placement):

- *Fast shape matching*. It considers the geometrical shape complementarity between two molecules and generates more favourable putative conformations of predicted binding sites [312].
- *Incremental construction*. In this approach, ligands are split into fragments docked separately in the receptor site and combined eventually [313-314].
- *Monte Carlo simulation-based*. Here, the docked ligand is fitted inside the receptor site through many random positions and rotations to decrease the possibility of getting trapped in a local free energy minimum [315].
- *Simulated annealing*. It uses every docking pose for a simulation, with the temperature decreasing gradually in regular intervals in each simulation cycle [316-317].
- *Distance geometry-based algorithm*. This search method uses information expressed through intra- and intermolecular distances and assembles them, allowing the calculations of structures or conformations that are in agreement with them [318].
- *Evolutionary programming*. Here, computational models of natural evolutionary processes are used to explore the conformational space of flexible ligands and, simultaneously, sample available binding modes in the protein binding site [319-320].

Generally, docking algorithms predict several poses for the ligand within the binding site. Subsequently, scoring functions are deployed to assess the steric complementarity between the ligand and the receptor and their chemical complementarity. Scoring functions can be divided into three main classes:

- *Force-field*. This scoring function deploys non-bonded terms of classical molecular mechanics force fields, some of which are physics-based [321-323].
- *Knowledge-based*. Here, the scoring functions are based on statistical observations of intermolecular contacts identified from structure databases [324-329].
- *Empirical*. These functions use several intermolecular interaction terms calibrated through a regression procedure, where theoretical values are fitted as close as possible to particular experimental data [330-334].

Ideally, combining the search algorithm and the scoring function should result in a solution close to the natural ligand position. In practice, docking algorithms are tested to reproduce known ligand conformations in an X-ray crystal structure within a given margin of accuracy and to recognise one of the conformations closest to the experimental structure as the best solution.

Molecular docking of GAGs

Despite plenty of available conventional docking software, originally developed and optimised for other types of ligands, most of it does not perform at the required quality level for docking GAGs [335-336]. Although the tested docking programs predicted the ligand binding poses correctly in many cases (placement performance is acceptable), the ranks of the docking poses were often poorly assigned (scoring is profoundly challenging). Uciechowska *et al.* conducted a molecular docking benchmarking study for a non-redundant dataset of 28 protein-GAG structures, comparing 14 docking programs and their performance for these systems. Despite AutoDock3 being the most successful one [336-337], it still experiences several serious limitations. The most fundamental one is the limit of 32 torsional degrees of freedom for the ligand, which renders longer GAGs (>dp8) unfeasible to be docked fully flexibly. Based on this, most researchers focus on short GAGs [337]. Another set of docking programs dedicated to GAGs (based on the AutoDock program) included Vina-Carb [338] and its advanced derivative GlycoTorch Vina [339], which both outperformed AutoDock Vina [340] and Glide (not designed for GAGs, works well for some protein-GAG complexes though) [341]. There are also online docking software servers such as ClusPro [342, 343], HADDOCK (High Ambiguity Driven biomolecular DOCKing) [344], or SwissDock [345] that can be used to dock GAGs. Some of them proved successful for several systems: 1) ClusPro was used recently to study interactions between SARS-CoV-2 (severe acute respiratory syndrome coronavirus and HS [346]; 2) HADDOCK demonstrated its efficiency in a GAG-related study where CXCL-8 (interleukin 8) interactions with HP were examined [347]. Alternatively, by using molecular docking software, one may manually place a GAG near the predicted binding site and run the molecular dynamics (MD) simulations to refine the binding pose [348-349].

To overcome the common difficulties in GAG docking, specific approaches were proposed to dock GAGs. For example, Dynamic Molecular Docking (DMD), a combination of molecular docking and MD, was suggested [300]. This steered-MD technique applies the additional

potential to move a GAG (or any ligand) from a distant position toward the *a priori* known binding site on the receptor's surface. The major drawback of this method is the demand for prior knowledge on a binding site, which is not always available. In addition, it may be computationally expensive, owing to the required size of the periodic boundary box and the use of the explicit solvent model. Another technique to dock GAGs is a fragment-based approach and is also suitable for longer molecules made up of repeating units [350]. Here, the protein's surface is sampled by docking trimeric fragments of a GAG which are further assembled in a long chain based on their overlaps. This rather basic idea enables docking longer GAGs without any length-associated limitations. Still, if the GAG docking site is close to the negatively charged amino acid residues, this method could fail to dock trimeric fragments near such residues. As a consequence, docking solutions for longer GAG fragments would not be obtained. A novel approach named repulsive scaling replica exchange molecular dynamics (RS-REMD) appears to address the mentioned challenges [351]. In RS-REMD, van der Waals' radii are increased in different replicas, whereas other types of interactions are unaffected. It enables a robust and extensive search for the proper binding sites and poses on the protein surface simultaneously, leaving the docked molecule and the receptor sidechains flexible. Furthermore, it was adapted for protein-GAG systems and tested on a set of 21 protein-GAG complexes, which included GAGs of various lengths (pentamers, hexamers and heptamers) [352]. In 19 out of 21 cases, the method found the binding site correctly. Even though this approach might not be effective in predicting binding sites within a protein's pocket (e.g., an enzymatic active site) or distinguishing GAG binding poses of the opposite polarity, it is possible to obtain satisfactory results in the MD simulations of the nanosecond scale while considering the effect of solvent.

The explicit solvent model is expected to be superior to the implicit one in terms of docking quality due to a more realistic description of the interactions and solutes [353]. Marcisz *et al.* proposed an improved RS-REMD protocol that introduced the explicit solvent water model [354]. They observed an improvement in the performance of docking GAG molecules. They also found that despite using a more accurate explicit water model, they did not observe any increase in computational expenses in comparison to the corresponding implicit solvent approach.

Finally, all the above-specific GAG docking approaches bring an extra complexity compared to standard docking methods for a user. They may be too complicated to handle for nonexperts in molecular modelling. For this reason, Marcisz *et al.* proposed a more straightforward

approach to dock longer GAGs that consists of four steps: (1) docking a short (hexameric) GAG in all-atom representation; (2) elongating the docked GAG using periodic units in a coarse-grained (CG) representation; (3) running an MD simulation to find an ensemble of GAG conformations for the entire GAG molecule; (4) calculating binding free energy [355]. The authors believe that such a combination of molecular docking, MD, and MD-based free energy calculation schemes will be more effective than classical docking approaches yet not cause many technical difficulties for a user.

Molecular docking in CD systems

The formation of host-guest supramolecular complexes was demonstrated with fluorescence spectroscopy, UV-vis spectroscopy, atomic force microscopy and infrared spectroscopy. These traditional methods used to explore the mechanism of host-guest action of CD, however, have significant limitations and are unable to keep up with the rapid development of supramolecular intelligent nano-systems of CD [356]. Therefore, molecular docking has been used in the study of CD-guest interactions, involving predictions of the optimal binding conformation, to explain the experimental results of CD-guest interactions and explore the functional groups, specific sites, and directions of how the guest molecule enters CDs. Self-aggregation and the binding ratio of CDs and guests could also be responsible for the biased results. Illapakurthy *et al.* compared three docking software programs (SYBL DOCK, FlexiDOCK and DOCK 4.0.1) for artemisinin, its derivatives and 2-hydroxypropyl- β -CD [357]. They concluded that the DOCK 4.0.1 results were the best in the HP- β -CD system. An extensive assessment of the advantages and limitations of various docking software was performed and summarised by Wang and colleagues [358]. They indicated that GOLD and LeDock have the best sampling quality to identify the correct ligand binding poses, whereas AutoDock Vina has the best scoring approach. Nevertheless, the correlation between experimental binding affinities and docking scores was weak for most receptor–ligand docking, implying that the current scoring function is still to be improved for this system. It was highlighted that the scoring efficiency for different receptor targets varied even when using the same protocols within the same software. Therefore, the docking software and the appropriate protocols should be selected and tuned carefully for each particular molecular system.

Despite the successes in molecular docking, persisting challenges must be addressed. For example, there is a need to account for explicit water molecules in the protein binding region,

which is crucial for properly describing the interactions between drug/ligand molecules in the docking process. The conformational flexibility of protein molecules in the binding process may be limited during the docking process. Therefore, combining molecular docking with other tools such as MD, network pharmacology, and machine learning is recommended to improve theoretical accuracy and provide more valuable guidance from the experiments. Consequently, new molecular docking software with advanced performance is constantly being developed and appearing on the market.

2.2.3. Molecular Dynamics

MD simulations are a valuable tool for investigating the behaviour of biomolecules and biomolecular systems and soft matter in general at a level of detail that is still inaccessible in experiments [359-363]. In this computational method, molecules are studied in terms of their evolution in time, yielding trajectories of the system as an output [364]. These trajectories are calculated numerically by solving Newton's equations of motion for a system of interacting particles. The forces between the particles and their potential energies are calculated using specific interatomic potentials defined in specific force fields (FFs) [365].

In all-atom (AA) MD, a molecule is described as a series of charged points (atoms) linked by springs (bonds). An FF is applied to tell the time evolution of bond lengths, bond angles and torsions, the non-bonding Lennard-Jones, and electrostatic interactions between atoms (*equation 7*). The FF is a collection of equations and associated constants designed to reproduce molecular geometry and selected properties of tested structures.

$$U(R) = \sum_{bonds} k_r (r - r_{eq})^2 + \sum_{angles} k_\theta (\theta - \theta_{eq})^2 + \sum_{dihedrals} k_\phi (1 + \cos[n\phi - \gamma]) \quad (7)$$

$$+ \sum_{impropers} k_\omega (\omega - \omega_{eq})^2 + \sum_{i < j}^{atoms} \epsilon_{ij} \left[\left(\frac{r_m}{r_{ij}} \right)^{12} - 2 \left(\frac{r_m}{r_{ij}} \right)^6 \right] + \sum_{i < j}^{atoms} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}$$

where r is the bond length, θ is the atomic angle, ϕ is the dihedral angle, ω is the improper dihedral angle, r_{ij} is the distance in between atom i and j ; k_r , k_θ , k_ϕ , and k_ω are force constants; r_{eq} , θ_{eq} , and ω_{eq} are equilibrium values; the dihedral term is a periodic term characterised by k_ϕ multiplicity (n) and phase shift (γ); ϵ_{ij} is related to the Lennard-Jones well depth; r_m is the distance at which the potential reaches its minimum, q_i and q_j are the charges on the respective atoms and ϵ_0 is the dielectric constant. The polarisation component is usually expressed as:

$$E_{pol} = -\frac{1}{2} \sum_i^{atom} \vec{\mu}_i \overrightarrow{E_i^{(0)}} \quad (8)$$

where μ_i is an induced atomic dipole and $E_i^{(0)}$ is an initial electrostatic field causing this polarisation. Additionally, charges not central to atoms but off-centre (as for lone pairs) can be

included in the FF. AA FF provides parameters for every atom in a system, including hydrogens. Most routinely used FFs families in AA MD of proteins include AMBER [366-367], CHARMM [368] and GROMOS [369]. The accuracy and predictive power of MD simulations based on AA FF are constantly improving due to the simultaneous improvements in high-performance computing hardware, more accurate methods for calculating the potential energy of conformational ensembles, and more efficient methods for conformational sampling. Nowadays, MD simulations at the microsecond scale of systems containing hundreds of thousands of atoms are performed routinely. With specialised supercomputers, it has been possible to perform milliseconds-length simulations [370], and the simulations of entire cellular structures have been attempted [371].

CG approaches

The AA MD approach, however, has its limitations that could be significant in analysing specific aspects of receptor-ligand systems. Describing the large-scale process might require very long MD simulation timescales and, in general, is computationally expensive due to the high number of elements for which potential functions need to be calculated. These limitations can be overcome by a coarse-grained (CG) simulation approach. In the CG, atoms are unified into groups of atoms constituting one centre of interaction (a pseudo atom). By decreasing the degrees of freedom, much longer simulation times can be examined for more extensive systems at the expense of molecular details [372]. In the case of proteins, CG MD simulations can be performed using various force fields, MARTINI [373-375], UNRES (from UNited RESidue) [376], PRIMO [377], AWSEM [378], SCORPION [379] and SIRAH [380-381]. The MARTINI FF, which has an empirical nature, is the most common CG FF [382]. It is applied to lipid systems, proteins, nucleic acids, glycans, polysaccharides [383], and in material sciences. It is also suitable for simulating glycolipid membranes and monotopic and transmembrane proteins. Due to its intrinsically consistent CG approach, MARTINI provides the opportunity to extend the simulation to an extensive range of molecules, such as different lipid types, sterols, sugars, peptides, and polymers. Such flexibility is needed to examine complex carbohydrate-based systems (glycoconjugates, functionalised glycomaterials) or investigate protein-carbohydrate interactions [384]. The UNified Coarse gRaiNed model (UNICORN) is physics-based [385-386] and was created by merging models corresponding to

proteins (UNRES) [387-388], nucleic acids (NARES-2P) [389] and for polysaccharides (SUGRES-1P) [385, 390-391].

MD simulations of carbohydrates

Because of the crucial role of protein-carbohydrates in human biology, there is considerable interest in deploying MD simulations to help characterise these systems and aid in the rational drug design of new therapeutics [392-395].

The routinely used AA FF for carbohydrates are CHARMM36, GLYCAM06, GROMOS, and OPLS-AA-SEI [396-401]. CHARMM36 FF, derived from the CHARMM AA biomolecular FF to carbohydrates, offers a way to treat monosaccharides in their pyranose [402] and furanose forms [403], including sulfate and phosphate derivatives [404], all types of glycosidic linkages, glycans, and glycoproteins, with full provision for dynamic simulation in aqueous media [405]. Based on CHARMM Drude, there are advances in a polarisable empirical FF for hexopyranose. They bring significant corrections in treating a series of monosaccharides and their glycosidic linkages [406-411]. Through a compatible parametrisation with the AMBER family of FFs, the GLYCAM06 FF handles many monosaccharides, including those present in GAGs [412]. It is possible to build carbohydrates of all sizes and conformations [413] and to examine their structures. AMBER offers parameter extensions for glycoproteins, glycolipids [414-415], lipopolysaccharides [416], lipids [417], proteins, and nucleic acids that are needed to simulate the assemblies of all these components. GROMOS, attributed to a united atom framework, illustrates an extensive family of carbohydrate FFs for hexopyranose-based saccharides. GROMOS engine enabled improved parametrisations to account for essential carbohydrate features. Among others, they include parameters for hexopyranose in an explicit solvent (GROMOS 45A4 and GROMOS 53A6GLYC) [418], ring conformational equilibria in hexopyranose (GROMOS 56ACARBO) [419-421], chitosan and its derivatives (GROMOS 56ACARBO_CHT) [422], furanose-based carbohydrates (GROMOS 56ACARBO/CARBO_R) [423], and glycan structure simulation in glycoproteins (GROMOS96 43A1) [424-425]. The ongoing developments of the all-atom optimised potentials for liquid simulations (OPLS) [426-427] FF offers an improvement to correct the performance in the estimation of the conformational changes around the glycosidic torsion angles and treatment of the carbohydrate-carbohydrate interactions in solution and explicit-water simulations [428].

Recently, Plazinska *et al.* compared CHARMM, GLYCAM06 and GROMOS in the context of their abilities to describe the structural and dynamic features of protein-carbohydrate interactions [429]. All the tested FFs seemed to be able to reproduce the crucial CH- π interactions, while the magnitude of the predicted total protein-carbohydrate unbinding energies varied. The unbinding energy values were either lower (CHARMM) or higher (GLYCAM) than the experimental values. However, both FFs displayed a good correlation between the experimental and theoretical structural data. GROMOS, on the other hand, had the lowest deviation from experimental data in terms of predicted unbinding free energy. Still, the correlation between experimental and GROMOS-derived theoretical data was lower compared to CHARMM and GLYCAM06. The authors concluded that further improvements to the parameters describing the protein-carbohydrate interactions are required.

Lazar *et al.* observed that conformational collapse of oligosaccharides in the MD simulations with GLYCAM06j force field could occur for saccharide chains containing the deoxy sugar α -L-rhamnose after 400 ns. They concluded that the irreversible collapse involved forming a hairpin structure, with a hinge in the middle of the saccharide followed by the stabilisation of the hairpin via multiple hydrogen bonds between opposing arms of the chain. At the same time, GLYCAM06j recently proved capable of yielding good agreement with the NMR data for the conformational ensemble of Arixta, a heparan sulfate-specific pentasaccharide (GlcNS6D–GlcA–GlcNS3S6S–IdoA2S–GlcNS6S) [430].

In terms of CG models for carbohydrates, there have been some developments in the last few years. In 2004 Molinaro and Goddard proposed M3B – a pioneer CG model for malto-oligosaccharide in an aqueous solution [431]. The hexapyranose ring was represented as three beads corresponding to C1, C4 and C6 atoms of an atomistic model, while the water molecule was represented as a single particle. A 2-body Morse function defined the long-range forces. This model was beneficial in studying the water-glucan system. Liu *et al.* developed a similar CG model for α -D-glucopyranose, with each glucopyranose represented as three beads [432]. The MARTINI FF was also extended for carbohydrates [383]. Here, the monomeric saccharide unit is represented using three beads. The model was applied to amylose and curdlan, which resulted in the reproduction of structural properties. Lopez *et al.* also used this model to investigate the mechanical and physicochemical properties of cellulose I β , including the bending resistance of cellulose nanofibers [383]. Yu and Lau developed a CG α -chitin model based on the MARTINI FF to examine the longitudinal dimension of chitin fiber, which is more than hundreds of nanometers long [433]. Benner *et al.* developed a CG model for chitosan

to explore the effect of the degree of acetylation on its self-assembly in solution [434]. For chitosan polysaccharides, another CG model was developed based on a free-energy landscape for glycosidic bonds [435]. The model was used to study the equilibrium properties of chitosan in solution regarding the degree of deacetylation and polymerisation, ionic strength, and pH. Moreover, the MARTINI FF has been applied to investigate the polysaccharides' second virial coefficient of osmotic pressure, a thermodynamic solution attribute [436]. In contrast to the multi-bead per monomer models, as mentioned above, Srinivas suggested one bead per monomer CG model to study the conformations of I β cellulose [437]. Subsequently, the authors presented a CG model for natural cellulose in an explicit water system [438]. Glass *et al.* proposed a residue scale CG model termed REACH (Realistic Extension Algorithm via Covariance Hessian) [439]. A single bead represented each monomer, and the associated FF was developed through the atomistic simulation trajectory of cellulose fibril in an aqueous solution at various temperatures. This model was developed for I β cellulose and used to characterise its elastic attributes and degradation as a function of length and temperature. In the work of Poma *et al.*, a unified CG model for polysaccharides and protein systems was described [440]. Here, the glucose unit was represented by a single pseudoatom, similarly to the representation of each amino acid residue by one bead at the C α atom.

In the case of carbohydrates, the CG scaling is customised according to the nature of the concerned question. The CG multi-bead model is suitable for studying various conformational states of polysaccharides and other chemical specificity issues. On the other hand, the CG models representing a glycosyl residue as a single bead are more effective in investigating the structure and dynamics of polysaccharide assembly. The unified CG model for a protein-polysaccharide system can help explore the process of enzymatic hydrolysis of polysaccharides which has essential implications in the biofuel industry.

AA MD simulations of GAGs

The development of sophisticated Molecular Mechanics FFs, especially for GAGs, has been the subject of research over the past 50 years, revealing increasingly detailed properties of GAGs [441]. GLYCAM FF has an advantage due to the availability of parameters for unsaturated uronic acids [412]. However, several studies show that the three commonly used FFs (GLYCAM, CHARMM, and GROMOS) yield pretty similar results [412, 418, 442].

AA MD is broadly used in the analysis of protein-GAG interactions. Vuorio *et al.* described the formation of the CD-44-HA complex [348]. Microsecond timescale MD simulations of this system enabled a proposal of three different binding poses, one corresponding to the crystallographic structure. Additionally, it was noticed that, in all three poses, there was one particular CD-44 amino acid residue involved in HA binding, and its mutation abolished GAG binding. This work provided valuable information about the mechanism of HA binding, which might be essential in explaining the role of HA in CD44 activity. AA MD was also deployed by Uciechowska-Kaczmarzyk *et al.* to investigate the effect of HP binding on the conformation of the Vascular Endothelial Growth Factor (VEGF) [443]. The authors suggested a stable structure of the VEGF-HP complex whose conformation would allow interaction with the VEGF receptor, which could be crucial in cell signalling. Another example of deploying AA MD is the work of Potthoff *et al.* Here, the study of GAG binding by PCPE-1 was performed, and also, the potential role of Ca^{2+} in this complex stability was investigated [444]. The findings enabled the identification of the GAG-binding domain of PCPE-1. It was also observed that long GAG chains might bind stronger than shorter oligosaccharides. Sepuru *et al.* characterised the CXCL5-HS interactions with AA MD simulations [445]. The study indicated that the roles of the individual lysines are not equivalent and that helical lysine plays a critical role in determining binding geometry and affinity in the CXCL5-HS complex. Moreover, binding interactions and GAG geometry in CXCL5 were reported to be novel and distinctly different compared to the related chemokines CXCL1 and CXCL8. Krieger *et al.* applied a comparable methodology to study IL-8-HP interactions [446]. The authors proposed a “horseshoe” model in which a dimer of IL-8 is bound by HP 24-mer, which interacts with C-helices of IL-8 and estimates its stability by analysing root-mean-square deviation (RMSD) for IL-8 and HP. While IL-8 dimer was observed to be stable, the RMSD for HP suggested high flexibility of the polysaccharide chain and raised the question of whether it might be possible to bind dimeric HP, which was further investigated experimentally. These are examples of how AA MD was deployed to study protein-GAG systems. It is worth noticing that numerous proteins have been analysed with MD in terms of their interactions with GAGs, including: TIMP-3 [301], sclerostin [447], matrix metalloproteinase [447], APRIL [448-449], FGFs [450-452], endostatin [121], IL-8 [454-457], RANTES [458], IL-10 [459], CXCL14 [460], CXCL12 [461-462], langerin [452], BMPs [275], LOX [463], cathepsin proteases [464]. Whereas some of these studies are purely theoretical, others combine AA MD with experimental methods such as NMR, SPR, MS and diverse biochemical and cellular assays.

Lately, MD simulations have also been successfully used to provide valuable insight into the structure and dynamics of various GAG-metal ions systems. For example, Guevench *et al.* explained the consequences of the sulfation pattern and the binding of Na⁺, Ca²⁺, or Mg²⁺ on CS and other GAGs [465-468]. By combining two-dimensional infrared (IR) spectroscopy experiments with MD simulations, Giubertoni *et al.* showed that Ca²⁺ binding to HA increases the flexibility of this GAG [469]. Furthermore, Samantray *et al.* deployed MD simulations to describe the influences of sulfation, salt type, and salt concentration on the structural heterogeneity of GAGs [470].

Some recently published papers report multiple computational studies performed on large libraries of GAGs or GAG-mimetics, supporting the high-throughput potential of computational approaches to study GAGs and their interactions with proteins. Joshi *et al.* performed a set of MD simulations for a library of HA chains of various lengths complexed to hyaluronan lyase [471]. Torrent and colleagues studied an extensive array of HP chains of various lengths *in silico* for their capacity to bind up to 20 different viral, animal and human proteins, including sulfotransferase, heparinase, immune system-related proteins, proteins inhibitors, cell adhesion proteins, blood clotting components, growth factors and their receptors [472]. An extensive computational study of HP and CS of various lengths in complex with different angiogenic growth factors, cell surface receptors, chemokines and glycosidases was performed [87]. Sankaranarayanan *et al.* conducted an *in silico* combinatorial library screening consisting of the automated construction of viral GAGs to generate a library of HPs from 2- to 8-mers examined for their binding to antithrombin and thrombin [473].

CG MD simulations of GAGs

While AA approaches are sufficient to model short GAGs (6-8 units) and their interactions [474], CG models are needed to model very long GAGs, which consists of hundreds and thousands of monosaccharides block. This is crucial since such long GAGs are present in the ECM [68] and guide processes such as the creation of protein gradients [36] or the establishment of collagen networks [475]. The first CG model of GAGs, along with the respective FF, for use in Monte Carlo simulations was proposed by Bathe *et al.* [476]. The Almond group developed the first model applicable in the MD simulations [477-478], which covered HS, HA, CS and DS. This CG model uses empirical energy functions for glycosidic linkage and the ring pucker. It has provided previously unseen details of the structure-dynamics

relationships of GAGs in the context of PGs [478]. Samsonov *et al.* developed an alternating approach of coining 28 pseudo atoms corresponding to functional groups of 17 different GAG residues and a CG parameter set compatible with the AMBER FF. The set consisted of naturally occurring GAGs (HP, desulfated HS, CS, DS, and HA) and their artificially sulfated derivatives [479]. SUGRES-1P is also being extended to include GAGs [350].

As an alternative approach to using the CG model, Whitmore *et al.* proposed using glycosidic linkage and monosaccharide ring conformations from unbiased AA explicit-solvent MD simulations of short GAG polymers to construct conformational ensembles for GAGs of an arbitrary length rapidly. The group developed a dedicated algorithm to generate a library of non-sulfated chondroitin from 10- to 200-mer, compared to MD-generated ensembles for internal validation [466]. Subsequently, the same approach was deployed to libraries of HA and non-sulfated DS, KS, and HS [467].

AA MD for CD

MD simulations are a widely used tool to assess the steric interaction of complexes and predict the characteristic of CD-drug interactions either *in vacuo* or in the solution. The most popular FF used to simulate CD-guest complexes are GROMOS96 and CHARMM [480]. Since CDs are frequently modified with functional groups, it is essential to use the appropriate FF that can compatibly and appropriately account for these modifications in the simulations. Therefore, a new strategy called dedicated FF treats the guest molecules with the standard FF and modified CDs with the FF containing delocalised atoms since the substituted hydroxyl groups are considered [481]. He *et al.* showed that amphotericin B (AmB) makes a steady complex with γ -CD through the inclusion of the macrolide ring into the CD hole, whereas the guest ring cannot enter the β -CD because of spatial restrictions [482].

AA MD simulations have been deployed, mainly combined together with other computational and experimental techniques, to inspect the interaction mechanism of CD-guest molecules from different sites. Here are some examples of how previously mentioned (**Section 1.2.3**) characteristics of CDs and inclusion complexes can be monitored by AA MD simulations.

- *Water solubility.* The inclusion complex formed by different kinds of CDs possess different water solubility. The application of computational tools may rationalise the selection of CDs. β -CD and its derivatives are used extensively due to their low cost and effectiveness. Since γ -CD has a larger cavity and higher solubility, its inclusion

effect is better compared to β -CD when the volume of the guest is relatively large. Through molecular docking and MD simulations, da Silva and dos Santos found that the binding effect of cannabidiol with γ -CD was significantly higher than that with β -CD [483].

- *Transportation of active substances.* MD simulations can assist in modelling the inclusion penetration process of the membrane and analyse the main driving forces. Hotarat *et al.* simulated a process of the inclusion compounds formed by α -mangostin, β -CD and DM- β -CD. The authors identified the relationship between time and penetration distance of β -CDs, α -mangostin, and inclusion compounds via triplicated MD simulations. The changes in hydrogen bonds, van der Waals, and electrostatic interactions also revealed the α -mangostin penetration process and the release profile from β -CDs [484]. Another study by Zhao *et al.* applied molecular docking and MD simulations to examine CD-rutin inclusion complexes [485]. They reported that the α -CD cavity was too small to encapsulate lutein. In contrast, the large diameter of γ -CD allowed for the free rotation of lutein, leading to weak interactions in the γ -CD-lutein complex. The hydroxypropyl group in HP- β -CD spatially hinders lutein's free rotation, so the γ -CD-lutein inclusion complex had the highest binding energy. The CD-lutein transportation system effectively promotes lutein absorption and improves its biological activities.
- *Anticancer activity.* Caffeic acid phenethyl ester (CAPE) and its derivatives have specific cytotoxic effects on cancer cells. With molecular docking and MD simulations, Wadhwa *et al.* discovered that CAPE binds to the glucose regulatory protein (mortalin) and occupies most of the binding site of p53 [486]. Furthermore, the data revealed that the generation of an inclusion complex with γ -CD could compensate for the instability of CAPE. Mansonone G (MG) has a potential antitumor effect on malignancy. The inclusion process of β -CD with MG was examined by molecular docking and MD simulations [487]. The docking results revealed that the β -CD-MG complex could have two configurations with the aromatic ring (A-ring) or quinone ring (Q-ring) of MG entering the β -CD cavity. MD simulations verified that only the configuration with the A-ring entering the CD cavity was stable.
- *Antibacterial activity.* Cinnamaldehyde indicated inhibitory effects on the growth of *Aspergillus niger*, *Penicillium* and *Rhizopus*. Using molecular docking, Sun *et al.* established that the benzene ring of cinnamaldehyde was partially embedded in the

cavity [488]. At the same time, the carbonyl group from cinnamaldehyde formed a hydrogen bond with the hydroxyl group of CD, stabilising the complex structure. MD simulations revealed that modifying the hydroxyl group of β -CD could improve CD's inclusion ability and cinnamaldehyde's antibacterial activity. Among them, the DM- β -CD-cinnamaldehyde complex proved to be the most successful one.

- *Bitterness masking.* The phenols in olives and olive oil effectively repair the damage caused by reactive oxygen species, but they are known for their undesirable bitterness. Both β -CD and caffeine can form inclusion complexes with these phenolic compounds. Rescifina *et al.* conducted MD simulations of these two inclusion complexes using the CHARMM27 FF [489]. They established that β -CD and phenolics formed inclusion complexes in a 1:1 ratio. However, the binding constants of the CD-phenol inclusion complexes were 10–40 times higher than those of the caffeine-phenol complex.
- *Facilitating compound extraction.* A high-performance liquid chromatography study confirmed that α -CD, β -CD, and γ -CD improve the extraction of ephedrine and berberine from traditional Chinese medicine, with β - and γ -CD being the most effective [490]. MD simulations were used to investigate the inclusion conformations of ephedrine with α -CD, β -CD, and γ -CD using MD. The authors reported that ephedrine penetrated the CD cavity and that the benzene ring was oriented toward CD's primary or secondary hydroxyl group. Compared with α - and γ -CD, the β -CD cavity had a moderate volume, and the ephedrine was tightly packed, thus effectively increasing the dissolution rate of the ephedrine.

CG MD for CD

There are a few examples of CG MD simulations used to study CD-containing systems. These studies consisted either of classic host-guest interactions or more complex membrane-including systems. Wang *et al.* studied β -CD and adamantane. The CD structure was represented by two bead types: hydrophilic and hydrophobic. The reason for using CG MD was to investigate a multiblock copolymer. Lopez analysed the inclusion of cholesterol molecules into the CD combined with the simultaneous CD-dimers adsorption on the membrane/water interface [491]. They implemented the CG MD approach to validate whether cholesterol can be extracted more readily from a liquid-disordered phase than a liquid-ordered one. The CG systems contained the planar lipid bilayer and a small liposome. Cieplak described the binding of small,

functionalised dendrimer molecules to β -CD-terminated self-assembly monolayers [492]. The formerly performed all-atom MD simulations of the same system were deployed as a basis for using the longer timescales in the CG MD approach. Interestingly, Škvára and Nezbeda used a similar concept a decade later [493]. They performed all-atom MD simulations of methanol systems that consisted of racemic ibuprofen and β -CD to use those results for the later development of CG models.

2.2.4. Binding free energy calculations

This section discusses the recent developments and applications of MD to calculate binding free energy that is fundamental for the analysis of receptor-ligand complexes. Through the deployment of absolute alchemical methods, linear interaction energy (LIE), umbrella sampling (US), Molecular Mechanics-Poisson Boltzmann Surface Area (MM-PBSA), Molecular Mechanics-Generalized Born Surface Area (MM-GBSA), researchers can evaluate biomolecular interactions that drive molecular recognition at atomic resolution in terms of their binding free energies.

Free energy perturbation (FEP) was introduced by Robert W. Zwanzig in 1954 [494]. The energy difference between states A and B is calculated according to the following equation (9):

$$\Delta F(A \rightarrow B) = F_B - F_A = -k_B T \ln \left\langle \exp \left(\frac{-E_B - E_A}{k_B T} \right) \right\rangle_A \quad (9)$$

where k_B is Boltzmann's constant and T is the temperature. The triangular brackets denote an average over a simulation run for state A. In the FEP method, the free energy difference between states A and B is evaluated with the Zwanzig equation (10):

$$\Delta G^{FEP} = G_B - G_A = -k_B T \ln \left\langle e^{\frac{-(V_B - V_A)}{k_B T}} \right\rangle_A = +k_B T \left\langle e^{\frac{-(V_A - V_B)}{k_B T}} \right\rangle_A \quad (10)$$

where V_A and V_B are potential functions of the states A and B, triangular brackets denote the Boltzmann-weighted ensemble average generated according to the potential role of the corresponding state. Equation (9) indicates that the potential energy differences can be averaged over an ensemble generated using MD simulations that start from either state A or state B, thus allowing us to estimate the convergence by comparing free energy results between the forward and backward transformations. This approach is expected to be numerically accurate only when the energies of the initial and final states of the system differ only slightly so that one state might be regarded as a perturbation of the other. In a practical analysis, the FEP is broken down into multiple small steps (or windows) by simulating the transition from A to B via intermediate states and integrating along this pathway. FEP is a reliable method for predicting pKa constants [495], describing solvent effect [496] and is used in the virtual screening of ligands in drug design [497]. In GAG-containing systems, FEP was utilised to

estimate the GAG binding affinity of the antithrombin-HP pentasaccharide complex [498] and the FGF1-HP pentasaccharide complex [499]. However, this method might not be appropriate when considerable structural alterations occur, so better sampling should be deployed [498]. Linear interaction energy (LIE) was proposed by Aqvist *et al.* in 2001 [500]. Here, the binding free energy is estimated as a linear function of electrostatic and van der Waals interaction energies as described in the equation below (11):

$$\Delta E^{LIE} = \beta(E_{bound}^{ele} - E_{free}^{ele}) + \alpha(E_{bound}^{vdW} - E_{free}^{vdW}) + \gamma \quad (11)$$

Where the α and β parameters are van der Waals and electrostatic scaling factors, respectively, that may have different values in water and protein environments, while γ is a constant free energy term. E_{bound}^{ele} and E_{free}^{ele} denote electrostatic interactions of bound and free (unbound) state, and E_{bound}^{vdW} and E_{free}^{vdW} denote van der Waals interactions in bound and free (unbound) states. In addition, the difference in solvent-accessible surface area between the bound and free state of the ligand might be introduced to the LIE scheme [501-502] (12):

$$\Delta E^{LIE} = \beta(E_{bound}^{ele} - E_{free}^{ele}) + \alpha(E_{bound}^{vdW} - E_{free}^{vdW}) + \gamma(SASA_{bound} - SASA_{free}) \quad (12)$$

where SASA is a solvent-accessible surface area term. While performing a classical (unbiased) MD, the analysed systems might get trapped in a local minimum. Therefore, no other local minima in the potential energy surface (PES) are explored, even if an extensively long MD was performed [503-504]. Consequently, no transition state could have been observed and described in terms of its energetic characteristics. An external force must be applied to drag the system over the energy barrier to another local minimum to tackle this problem. For this reason, Torrie *et al.* proposed US in 1977 [505]. This method simulates the biochemical processes (e.g., ligand binding) by choosing a reaction coordinate(s) of interest (e.g., distance or angle). The range of selected coordinate(s) is split into windows, in which MD simulations are performed with biasing potential enforced on the reaction coordinate(s). The windows cover the analysed range of reaction coordinates and overlap with adjacent windows, assuring proper sampling over this coordinate. As the last step, the outputs from all windows are processed simultaneously by Weighted Histogram Analysis Method (WHAM) [506-507] or umbrella

integration to remove biasing potential. Consequently, the potential of mean force is constructed, which is the free energy profile along the chosen coordinate.

Another popular technique to estimate free energy binding is MM-PBSA (Molecular Mechanics-Poisson Boltzmann Surface Area) [508] and its approximation MM-GBSA (Molecular Mechanics-Generalized Born Surface Area). Here, the energy difference between the unbound and bound state is computed while considering the presence of a solvent. These calculations are performed with the implicit solvent model to avoid the direct impact of explicitly described solvent-solvent interactions that could lead to high energy fluctuations. In MM-PBSA, the solvent is described by the Poisson-Boltzmann equation (13):

$$\Delta G_{el} = \frac{1}{2} \sum_i q_i (\phi_{sol}(r_i) - \phi_{vac}(r_i)) \quad (13)$$

where ϕ_{sol} is the electrostatic potential of the solute atomic partial charges q_i in an environment corresponding to the region inside the molecular surface (as determined by atomic coordinates, radii, and probe sphere radius), and the outer part has a dielectric constant of 1.0 and ϵ_w , respectively. ϕ_{vac} denotes the electrostatic potential of the exact solute charges *in vacuo*.

MM-GBSA method deploys the approximation of the Poisson-Boltzmann equation (14):

$$\Delta G_{el} \approx \Delta G_{GB} = - \sum_i \frac{q_i^2}{2R_i} \left(1 - \frac{1}{\epsilon_w}\right) - \frac{1}{2} \sum_{i,j} i,j \frac{q_i q_j}{f^{GB}(r_{ij}, R_i, R_j)} \left(1 - \frac{1}{\epsilon_w}\right) \quad (14)$$

where ΔG_{el} is the electrostatic component, ΔG_{GB} is the reaction field component, R_i and q_i correspond to the effective Born radii and charges of the atoms i , respectively, ϵ_w is the dielectric water constant and the Generalized Born (GB) function, f^{GB} is (15):

$$f^{GB} = [r_{ij}^2 + R_i R_j \exp(-r_{ij}^2 / 4 R_i R_j)]^{1/2} \quad (15)$$

as the Born radii of atoms i and j decrease, the effective distance between the atoms f_{GB} increases. The electrostatic contribution to solvation free energy by either Poisson-Boltzmann or Generalized-Born is summed with the hydrophobic contribution term to get the complete solvation-free energy. The great advantage of LIE and MM-PBSA/GBSA is that they sample

only the configurational space of the initial and final states of the binding reaction, which drastically increases the efficiency of the calculations relative to the more rigorous approaches. However, the quality of their prediction has been questioned since the accuracy of LIE depends on a set of empirical and typically non-transferable parameters, while that of MM-PBSA is limited by evaluating the solvent contribution by continuum electrostatics.

As a next step, binding free energy (ΔG_{vacuum}^0) is computed to include interaction energy between the receptor and ligand (ΔE^0) and the entropy change ($T\Delta S^0$) upon binding if required (16).

$$\Delta G_{vacuum}^0 = \Delta E_{molecular\ mechanics}^0 - T\Delta S^0 \quad (16)$$

Assuming biological systems behave like a rigid rotor model, the translational and rotational entropies can be computed with standard statistical mechanical formulae. Subsequently, vibrational entropy contribution can be approximated with one of two methods: normal mode (NM) or quasi-harmonic (QH) [509-510]. In the NM approach, the vibrational frequencies of normal modes can be verified at various local minima of the potential energy surface [511]. Generally, for large systems, NM calculations are computationally demanding since they require the minimisation of each frame, building the Hessian matrix, and diagonalising it to achieve the vibrational frequencies (eigenvalues). Due to the Hessian diagonalisation, NM calculations scale roughly $(3N)^3$, where N is the number of atoms in the system. This approach allows for more accurate results but is computationally expensive and prone to a large margin of error, leading to significant uncertainty in the results [508, 512].

On the other hand, the QH approach can be used, where the eigenvalues of the mass-weighted covariance matrix constructed from every ensemble member can be approximated as frequencies of global, orthogonal motions. Although the QH technique is less computationally costly, a considerable number of either frame or ensemble members are required to extrapolate the asymptotic limit of the total energy for each ensemble, which in turn, increases the computational cost of the original simulation [513]. Relative free energy calculations between related systems (for example, binding similar ligands to the same protein) often assume the same solute energy for each system, thereby eliminating the requirement to calculate them explicitly [508]. In addition, the entropy contributions are frequently disregarded due to their high computational cost, and the obtained entropy values are much higher than expected,

particularly for protein-GAG complexes [509]. Finally, the binding free energy is obtained upon calculating binding free energy *in vacuo* and solvation energy (17):

$$\Delta G_{bind,solv}^0 = \Delta G_{vacuum}^0 + \Delta G_{solv,complex}^0 - (\Delta G_{solv,ligand}^0 + \Delta G_{solv,receptor}^0) \quad (17)$$

MM-PBSA and MM-GBSA were used to study protein-GAG, drug-GAG, drug-CD, and BSA-drug systems. Particularly for calculating binding free energies in PCPE-1 [444], VEGF-A [443], TIMP-3 [301], matrix metalloproteinases [448], FGF-2 [514], IL-8 [457], sclerostin [447], BMP-2 [275], ellipticine [515], and thioflavin T [516]. MM-PBSA and MM-GBSA were also used to investigate chiral discrimination of ibuprofen isomers in β -CD [517], chelate effect in CD dimers [518] or encapsulation mechanism of α -mangostin by β -CD [519]. The method was also deployed to characterise BSA-(oxidovanadium(IV)-salphen) [520], BSA-naphthalamide-polyamine derivatives [521], and different flavonoids and BSA [522].

Recently, Marcisz *et al.* investigated the methodological aspects of binding free energy analysis by MM-GBSA and LIE when applied to protein-GAG complexes [449]. To verify their results and evaluate both approaches, they extensively investigated the statistical relevance of the data from free energy calculations for APRIL protein complexes with tetrameric and hexameric GAGs. The authors concluded that 10 ns simulation in the case of APRIL-GAG or a similar system in size and interaction patterns is sufficient for analysing the MD trajectories, especially when working on a bigger data set (50 trajectories per complex). Additionally, it is recommended to rather increase the number of MD runs than elongate them, as was previously advised by Genheden and Ryde for other molecular systems [523]. Finally, it was reported that the MM-GBSA yielded more statistically reliable results than LIE, whereas the LIE method is substantially less computationally expensive. However, MM-PBSA/GBSA binding free energies are overestimated because of the application of MM calculations. Therefore, more precise QM calculations might be performed to improve these inaccuracies. This technique is called QM-PBSA/GBSA [523] and was used in CD (and their derivatives)-containing systems: two flavanones, hesperetin and naringenin, by complexation with β -cyclodextrin (β -CD) and its methylated derivatives (2,6-di-O-methyl- β -cyclodextrin, DM- β -CD and randomly methylated- β -CD [525-528]).

2.2.5. Challenges of molecular modelling of GAG and metal ions-containing systems

While computational approaches offer many advantages to scientists, including reduced costs of experiments, no need for sophisticated equipment, and the capability of inspecting the analysed systems at the atomic level unfeasible for the experiments, they have limitations that researchers shall not disregard. Since every investigated system has peculiarities, particular computational challenges must be considered.

Challenges of modelling GAG-containing systems

- GAGs possess great conformational freedom [530]. The longer the polysaccharide chain, the greater the number of degrees of freedom and the longer and more expensive the required calculations.
- Pyranose GAG rings (especially IdoA(2S)) undergo conformational changes. The chair and skewed conformations dominate [477], which might correspond to different binding free energies in complexes with interacting partners [514].
- GAGs interact with LYS and ARG's long flexible positively-charged side chains, contributing to the increase of the conformational space [531].
- Electrostatically-driven interactions are dominant between GAGs and proteins [532]. Therefore, the proper treatment of electrostatics is required.
- GAGs interact abundantly with solvent, essential for these polysaccharides' structure and biological function [533] and their interactions with proteins [498, 534].
- Sulfation code is still not understood [535]. The distribution of sulfate groups within a disaccharide unit and countless combinations of these units in a chain affect GAG's structural and dynamic properties, molecular recognition, and biological activity [536].
- Multipose binding is a feature of GAG-containing complexes. GAGs can bind to their protein target at the same binding site but in different poses with similar binding free energies [537-539].

Paiardi *et al.* summarised a distribution of computational studies among GAGs; almost half concerned HP [540], followed by CS > HS > HA > DS > KS. It can be explained, given the large array of biological functions played by HP and the significance of the design of HP-like drugs for the treatment of coagulation disorder, abnormal inflammatory or immune responses

and angiogenesis-dependent diseases. Moreover, the HP structure is more homogenous than any other GAGs and is more easily accessible. Therefore, in both computational and experimental studies, HP is often used as a structural analogue of HS/HSPGs. Surely, this has lowered the number of computational studies of HS, resulting in the number being significantly lower than that of studies of HP despite the former being more biologically relevant than the latter.

Challenges of modelling metal ion-containing species

- High angular momenta atomic orbitals. Due to the presence of d and/or f orbitals in transition metals (TM) ions, they possess more electrons and more complex shapes of the orbitals, which leads to more complex chemical bonding characteristics than the ones parameterized within molecular mechanics spherical formalism.
- Electronic state degeneracy. TM ions have complex electronic structures with various spin states possible with relatively close energies.
- Multiple oxidation states. A given TM ion can have multiple oxidation states that could be dynamically altered. For instance, Mn possesses oxidation states between +2 and +7, with +2, +4 and +7 being the most common. The higher oxidation states lead to a highly charged system with distinct long-range effects that require appropriate treatment.
- Complex chemical bonding and multiple coordination numbers (CNs). TM-containing species have more sophisticated chemical bonding patterns than their organic counterparts, resulting in their capability to possess a dynamic and flexible coordination environment. Whereas the CNs are pronounced for the leading group of metal elements, the more ionic chemical bonding character leads to higher and more diverse CNs compared to the main group of nonmetal elements. Ca^{2+} comprises CNs that span from 5 to 10 in an aqueous solution [541]. The higher value of CN corresponds to the higher number of compounds coordinating this metal centre. Due to flexible bonding, it is laborious to simulate various chemical bonds with just one modelling strategy/FF set of parameters.
- Experimental shortcomings. Limited experimental data about TM species slow the development of accurate theoretical methods needed for proper ion parameterisation.
- Polarisation. The charge on TM is not constant and is affected by factors like the environment, the nature of coordinated ligands, and the oxidation state of the TM. The

electronic cloud is usually distributed nonsymmetrically around the TM ion and can further change and redistribute in response to changes in the surrounding. There is no general solution to date on representing TM in molecular FF; therefore, it is still a matter of ongoing research and development [542].

In addition to the challenges mentioned earlier, another difficulty with modelling metal ion-containing systems remains the question of what ion parameters should be used. In the case of most ion models in molecular mechanics, they are simple spheres. There are only three basic parameters for a given ion in a spherical representation: the charge and the two parameters describing the Lennard-Jones potential. The charge is fixed to the actual value for the ion (i.e., +2 for calcium or magnesium), and the Lennard-Jones parameters are chosen to reproduce the experimental hydration free energy. The more advanced cationic dummy atom model (CDAM) was proposed, initially for the Mn^{2+} [543] and then extended to other metal ions. Nowadays, the parameters of this type are available for the Ca^{2+} and Mg^{2+} and some 3d TM ions, including Co^{2+} , Cu^{2+} , Fe^{2+} , Ni^{2+} , and Zn^{2+} . There are no CDAM parameters for monovalent ions because the nonbonded model is accurate enough for most scenarios. In the CDAM representation, several dummy sites are connected to the central metal ion. The number of dummy sites varies depending on the coordination number of the investigated metal ion. Duarte *et al.* indicated that CDAM simultaneously reproduces experimental hydration free energy (HFE) and ion-oxygen distance of the first solvation shell (IOD) and therefore provides an improvement over nonbonded models [544].

On the other hand, although CDAM possesses more centres, it does not automatically guarantee better results than standard (12-6 LJ) nonbonded metal ions [545]. The nonbonded model allocates an integral charge of the metal ion in a single centre. In contrast, the CDAM model approach distributes partial charges to the dummy sites, which may decrease the interaction strength between the ion and specific residues. An assumed CN is assigned to the ion; the CDAM may not be able to simulate processes involving CN changes.

CDAM parameters were developed for Ca^{2+} and Mg^{2+} by Saxena and Sept [546]. The pentagonal bipyramid is the most prevalent geometry for Ca^{2+} sites in biological systems. Therefore, seven dummy sites were placed around the central nuclei, whereas an octahedral geometry was used for the Mg^{2+} . Here, the dummy sites have uniform charges, while the central core is treated as neutral (Figure 5).

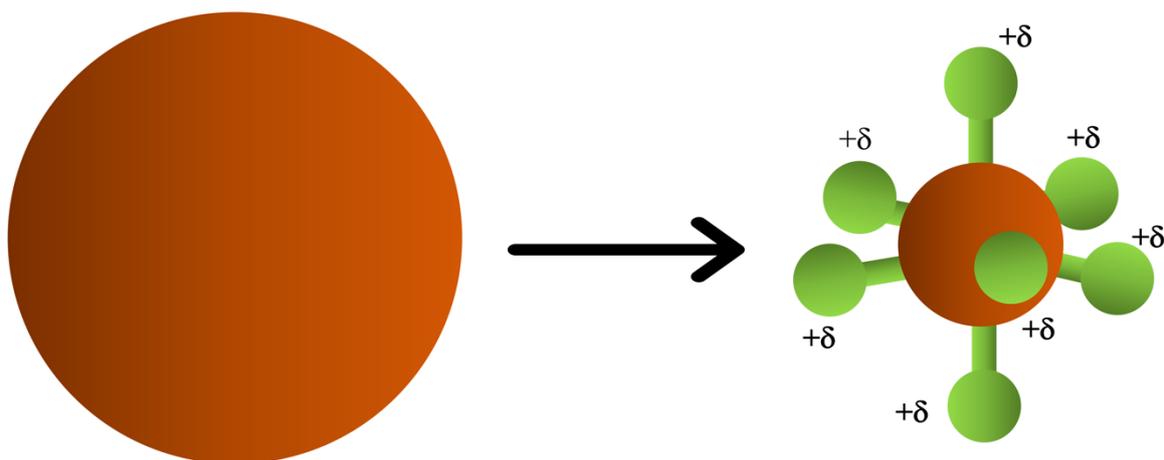


Figure 5. Calcium ion models. A simple sphere (left) is substituted with the distributed charge centres to multiple sites depending on the coordination geometry of the ion (right). The figure was adapted from Saxena and Sept's work [546].

The parametrisation attempts aimed to reproduce the experimental HFEs and also managed to reproduce experimental radial distribution functions (RDFs). Based on their calculations of relative binding energies of Ca^{2+} and Mg^{2+} to various systems, they demonstrated that the CDAM model outperformed all tested nonbonded models. In the following study, Saxena and Garcia refined the initial CDAM parameters for Ca^{2+} and Mg^{2+} since the original parameters could not simulate solutions with high salt concentrations. Duarte and others proposed CDAM parameters for Mg^{2+} , Fe^{2+} , Co^{2+} , Ni^{2+} and Zn^{2+} . They used an octahedral geometry with every dummy site assigned a $+0.5e$ charge and every nucleus with $-1.0 e$, respectively [544].

In the case of ions with strong ionic characteristics (e.g., Na^+ , K^+ , F^- , Cl^- , Ca^{2+}), the electrostatic term dominates its interactions with surrounding ligands. Li and Merz recommend using a nonbonded approach, including electrostatic and VDW interactions for these ions. In contrast, charge scaling might be required to reproduce better structural and dynamic properties under high salt concentrations [547]. There are a few classifications for nonbonded models: (a) The commonly used 12-6 nonbonded model usually provides excellent results for the alkali metal ions (except Li^+). This is because the partial charges on these ions are close to the formal charge ($+1e$), making a point charge a proper representation. (b) The CDAM, 12-6-4, Drude oscillator (DO), induce dipole moment (IDM), and ligand field molecular dynamics (LFMM) models are recommended to use for the remaining monovalent ions and divalent ions. This is because the 12-6 model fails to account for these ions' non-negligible polarisation and charge transfer (CT)

effects, making it challenging to reproduce multiple properties simultaneously. In the meantime, the models listed above incorporate the polarisation and CT interactions to some extent and offer significant improvement. (c) The 12-6-4 and induced dipole moment (IDM) are suggested for highly charged ions.

Accurate modelling of metal ion-containing systems, especially TM ones, is still challenging. Experimental studies, novel models and algorithms, advanced software development and rigorous application studies are crucial to further research in this area. The lack of experimental data holds back theoretical analysis of metal-containing systems. With more and better-quality experimental data, computational models (QM and MM) could be better parametrised. Additionally, with extensive experimental information available, benchmark studies to better understand the pros and cons of existing models can be performed to guide the development of the next-generation models.

Furthermore, it is essential to develop models that better encapsulate the physics of metal-containing systems at an affordable cost. For instance, models that accurately represent the structural dynamics of metal centres and bond-breaking/formation processes are needed to model TM-containing systems accurately. Advanced algorithms that bridge the gap between QM and MM models are required. Lastly, software development also plays a pivotal role in facilitating model development, enabling the rapid prototyping of various models to examine their applicability to a given problem.

3. Goals of the research

The main goal of my PhD research was to characterise the role of ions in carbohydrate- and protein-containing systems by computational approaches. To do this:

- the theoretical tools were applied to several chosen biologically relevant systems,
- novel protocols were developed and tested to model these systems more efficiently.

The Thesis is divided into several parts, each addressing particular questions:

1. Protein-ion-GAG and GAG-ion molecular systems

- *How are the most commonly deployed molecular modelling tools accurate and sensitive to investigate protein-ion-GAG systems?*
- *How do ions and their different parameters affect HP's structural and dynamic properties?*

2. The CD-ion molecular systems

- *How does the alkyl chain length affect the CD-SXS interactions (where $X=8,10$ and 12 denote the number of carbon atoms in the alkyl chain of the sulfate)?*
- *Can we predict/propose the mechanism of the entrance of the ion into the CD cavity using MD-based approaches?*
- *How does an initial SXS orientation affect the formed inclusion complex?*

3. The BSA-containing molecular systems

- *Can we observe the influence of various experimental conditions (pH and temperature) on the BSA-SDS interactions in the MD simulation?*
- *How powerful are the MD-based methods to localise binding sites for particular BSA-ligand systems?*
- *How can computational tools help account for the ligand displacement from BSA to HSA (human serum albumin)?*

4. Summary of the publications included in the PhD Thesis

The following section consists of short summaries of publications included in my PhD Thesis. Each unit is dedicated to a particular study, and detailed results are included at the end of this PhD Thesis as full-length publications.

4.1. Computational insights into the calcium ions role in protein-glycosaminoglycan systems [D1, from here on D with indices corresponding to the manuscript included in my PhD Thesis and listed after the Literature section].

In this work, a computational study was performed to inspect the impact of calcium ions (Ca^{2+}) in annexin- Ca^{2+} -HP systems. Annexins belong to the family of homologous proteins that are widely distributed and ubiquitous in eukaryotes; they are involved in many physiological processes, including apoptosis, cell signalling, ion transport, homeostasis, and membrane trafficking [548-552]. Most functional properties of annexins are attributed to their characteristic ability to bind to membrane phospholipids and other anionic polymers, including GAGs, in a Ca^{2+} -dependent manner [553]. The HP-binding properties of the annexin II tetramer have been investigated, and a potential HP-binding site was identified [223]. Several *in vitro* studies reported and described the binding of GAGs, including HP, to annexins IV, V, and VI [554-557] and confirmed that this family of proteins expresses different GAG preferences, and that GAG binding may be Ca^{2+} -dependent. So far, there are only three available crystal structures of both annexin V and II with HP, including Ca^{2+} and characterised binding pockets (PDB ID: 1G5N, 2HYU, 2HYV). This fact served as a motivation to perform this computational study. The goal was to get a deeper insight into protein-ion-GAG interactions using *in silico* approaches and to verify the accuracy and sensitivity of the most commonly deployed molecular modelling tools in the investigated annexin- Ca^{2+} -HP systems. As a first step, the electrostatic potential maps for annexin V and II were calculated in the presence and absence of Ca^{2+} . The obtained results revealed that in both cases, the lack of Ca^{2+} , as expected, notably affected the electrostatic potential in the proximity of the protein's surface, which would consequently reduce the strength of protein-GAG interactions (Figure D1). Molecular docking was performed to investigate whether AutoDock3 can recognise GAG binding sites (reported by the crystallographic studies) for annexins V and II. Additionally, the novelty of this work aimed to determine whether the removal of the chosen Ca^{2+} can impact the protein-

GAG interactions and whether AutoDock3 is sensitive enough to account for that. The most representative structures of annexin V/II-HP complexes have been calculated-obtained and used for further MD and free binding energy analyses. As an alternative to AutoDock3, I tested a new approach of replica-exchange MD with repulsive scaling. Initially, this method was developed for protein-protein complexes [351] and applied to various protein-GAG systems [352]. I also included an MM-GBSA vs LIE comparison study for various protein-GAG/alginate systems to determine whether a linear relationship between these two end-point methods could be established. Although the analysis confirmed a significant correlation between MM-GBSA and LIE, it strongly depends on the investigated system. Finally, the HP dissociation pathway was examined using the Umbrella Sampling approach in the presence and absence of Ca^{2+} .

To sum up, this study presented a systematic and detailed description of the annexins-HP interactions, which enabled a better understanding of the presence of Ca^{2+} in the investigated complexes. Reported findings indicate the limitations of the computational methodologies applicable to protein-GAG systems and contribute to the common knowledge of the physicochemical basis underlying the interactions between proteins, ions, and GAGs, as well as their potential specificity.

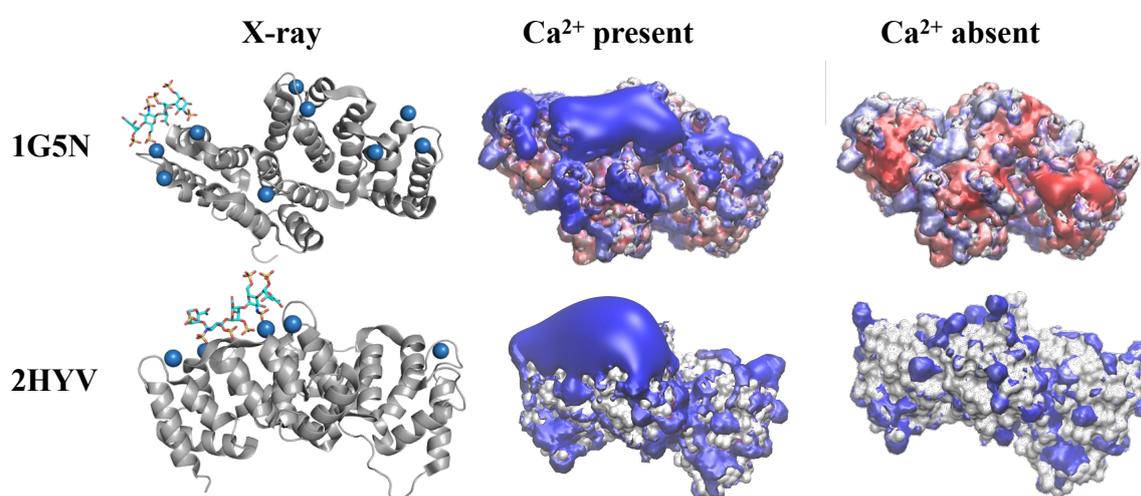


Figure D1. X-ray structures of annexins V and II (PDB ID: 1G5N and 2HYV, respectively), protein: grey cartoon, HP dp4: licorice, Ca^{2+} : blue VDW spheres. Electrostatic potential isosurfaces were calculated in the presence and absence of Ca^{2+} and are illustrated by surface representation (red, $-4 \text{ kcal mol}^{-1} e^{-1}$; blue, $+4 \text{ kcal mol}^{-1} e^{-1}$, respectively).

4.2. Impact of calcium ions on the structural and dynamic properties of heparin oligosaccharides by computational analysis [D2]

This work aimed to elaborate on the findings reported in D1. In this study, I characterised the structural and dynamic properties of HP oligosaccharides of various lengths (dp10 and dp18) by performing an extensive computational analysis of these HP molecules to investigate the impact of Ca^{2+} on the oligosaccharide conformation in a microsecond-timescale MD simulation. Since one of the key challenges in modelling protein-ion-GAG or GAG-ion interactions is the choice of the appropriate parameters for ions and the limited knowledge of their influence on specific ion effects, two models, namely spherical (referred to as type I) and cationic dummy atomic model (CDAM, referred to as type II) were used to account for their differences in HP- Ca^{2+} interactions [D2].

In this in-depth computational analysis, I concentrated on investigating the radius of gyration (R_g), sugar ring puckering, the distribution of dihedral angles, per-residue root mean square fluctuations, fractions and lifetimes of intermolecular hydrogen bond and thermodynamic parameters of HP molecules. Additionally, I monitored the coordination of Ca^{2+} to HP's coordination partners: 2 types of sulfate groups (NS and 6S) from GlcMS6S and one sulfate (S) and one carboxylate group (C6) from IdoA(2S). Based on the MD analysis, I concluded that the flexibility of the monosaccharides, the glycosidic linkages and ring puckering were not significantly affected by the presence of Ca^{2+} in contrast to H-bond propensities and the theoretically determined R_g for a fraction of the oligosaccharides' populations in both HP dp10 and HP dp18. Furthermore, the findings indicated the crucial importance of the Ca^{2+} parameters for modelling protein-HP interactions involving Ca^{2+} and highlighted the limitation of the currently available parameters to model binary (GAG-ion) and ternary (protein-ion-GAG) complexes. Nevertheless, increased accuracy and amount of structural data, as well as further computational and experimental analyses of other GAG-ion systems, are required to draw valid conclusions about the most appropriate divalent ion parameters for modelling ion-dependent protein-GAG interactions.

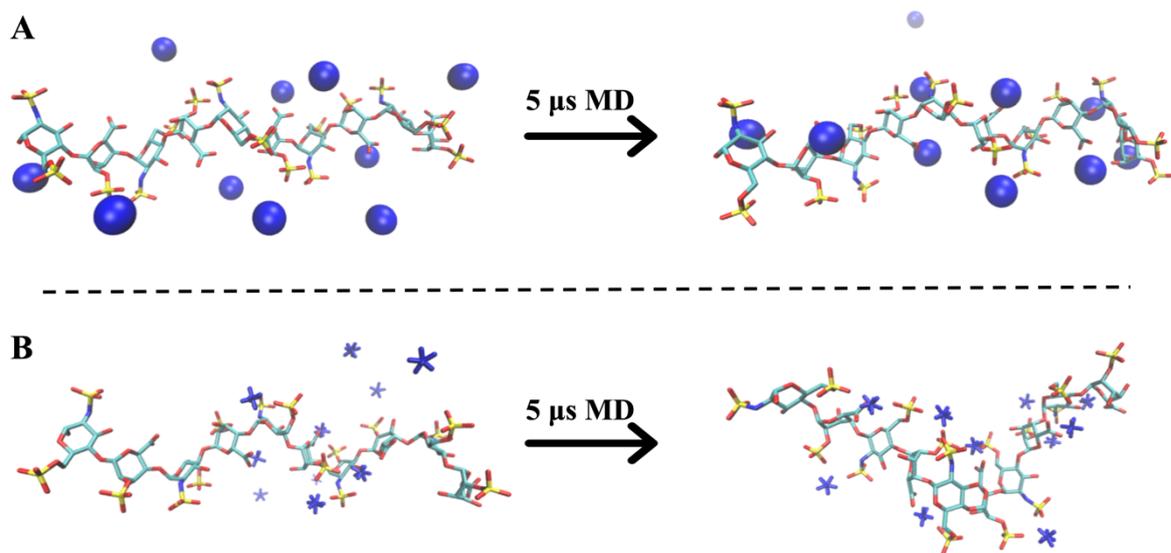


Figure D2. A: Starting structure for HP dp10 in the presence of Ca^{2+} type I (left) and final structure after 5 μ s MD simulation. B: Starting structure for HP dp10 in the presence of Ca^{2+} type I (left) and final structure after 5 μ s MD simulation.

4.3. Affinity and putative entrance mechanism of alkyl sulfates into the β -CD cavity [D3]

This computational study complemented the experimental analysis of interactions between β -cyclodextrin (β -CD) and three anionic surfactants of different alkyl chain lengths [D3].

Cyclodextrins (CDs) are novel pharmaceutical excipients and are extensively used in the pharmaceutical industry and drug delivery because of their ability to form host-guest complexes. Numerous structural features of guest molecules are responsible for the efficiency of guest-CD interactions. Among these factors, a guest's hydrophobic moieties and size are commonly considered to govern its binding affinity to CD molecules. Nevertheless, it is often challenging experimentally to extract structural and dynamic information at the atomic scale about CD inclusion complexes, and that is when MD simulations prove useful.

The surfactants chosen for this study were sodium dodecyl sulfate (S12S), sodium decyl sulfate (S10S) and sodium octyl sulfate (S8S). This set allowed us to assess the effect of the length of the hydrophobic carbon chain of the guest on the stability of inclusion complexes and the binding mode of the guest molecules, as well as on the detailed mechanism of the complex formation.

The study showed that the thermodynamic stability of the investigated complexes increased in proportion to the elongation of the guest alkyl chain. Both experimental and computational analysis revealed that apart from the hydrophobicity, electrostatic and van der Waals interactions should be considered when investigating a structure-based host-guest complexes design. The contribution of electrostatic interactions is more pronounced for the shorter alkyl chain sulfates. In contrast, van der Waals's interactions were essential for the complexation efficiency of the long-tail guests. Moreover, four putative mechanisms of alkyl sulfates' entrance into the β -CD cavity were proposed and inspected for the first time by MD-based computational approaches (Figure D3). Although a different entrance mode leads to different types of inclusion complexes, it can be presumed that the resulting complexes exist in a conformation equilibrium.

In conclusion, this work describes how the size of hydrophobic guest moiety affects the host-guest complex formation, proposes potential entrance mechanisms, and highlights the predictive power of computational approaches. All above could contribute to successful and faster drug design concerning β -CD.

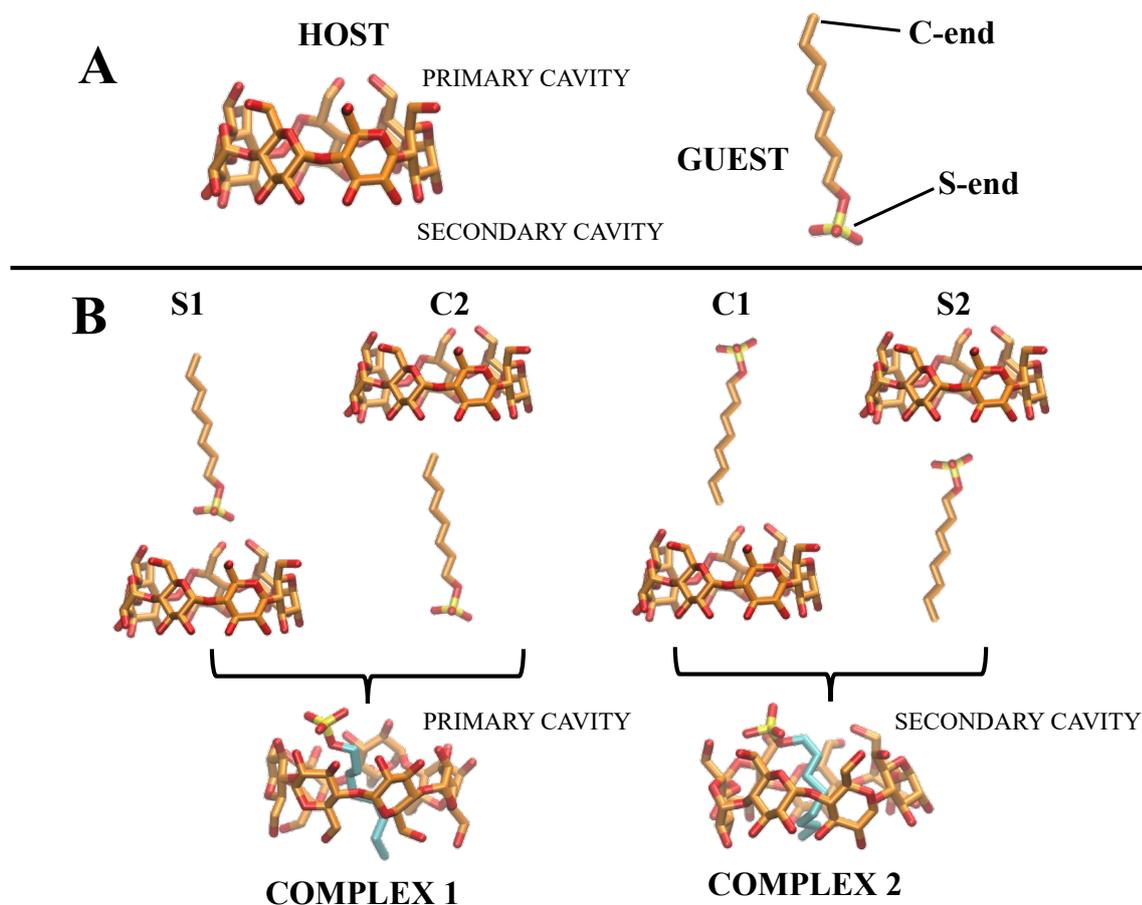


Figure D3. A: host (β -CD) and guest (S8S). B: Four potential mechanisms for a guest to enter β -CD: S1 = S-end first via primary cavity, C2 = first via secondary cavity resulting in Complex 1; C1 = C-end first via primary cavity, S2 = S-end first via secondary cavity resulting in Complex 2. For clarity, the colour of the guest molecule in Complex 1 and Complex 2 was changed compared to the unbound guest, and hydrogen atoms were omitted.

4.4. Physicochemical nature of sodium dodecyl sulfate interactions with bovine serum albumin revealed by interdisciplinary approaches [D4]

In this work, computational approaches were deployed to inspect the impact of pH and temperature on the interactions between sodium dodecyl sulfate (SDS), a common anionic surfactant, with bovine serum albumin (BSA) and to localise potential binding sites.

Surfactants are the subject of interest to many researchers due to their countless applications in industry, technological applications, and developments in life sciences. Furthermore, much attention has been paid to the studies of surfactant interactions with biologically relevant macromolecules such as proteins. Because surfactants tend to self-associate and form ordered structures (micelles), their concentration was kept below critical micelle concentration. Due to the presence of two tryptophan (TRP) residues in the BSA structure (TRP 134 and TRP 213), fluorescence spectroscopy was applied to localise the first binding site next to TRP 134 successfully. First, I performed molecular docking simulations to localise potential binding sites at two different pH values (5 and 7). However, this analysis did not show any conclusive results. Therefore, I conducted MD simulations for four systems (pH: 5 and 7, temperature: 290 and 300 K) where the surfactant molecules were initially placed randomly around the protein. This approach allowed me to identify the potential “second” binding site, which was more pronounced at lower pH and temperature. Still, the “first” binding site was not found in this MD simulation, and another strategy was proposed. Based on the steady-state fluorescence spectroscopy results and a known structure of HSA with myristic acid (PDB ID: 2XVW), the SDS ion was inserted in a similar binding mode close to TRP 134, followed by 100 ns MD simulation in the presence of 15 randomly distributed SDS ions. Finally, it was observed that during this MD simulation, another SDS ion approached the “inserted” one. Consequently, binding free energies were calculated with LIE and MM-GBSA. The results showed that the “inserted” and the “neighbouring” SDS ions had the most favourable energies. Hence, they were the most stable in binding to BSA compared to all obtained ΔG values for other SDS ions, including the ones corresponding to the putative “second” binding site.

The formation of the BSA-SDS complexes is an enthalpy-driven process in which the van der Waals interactions play a pivotal role. The first binding site, located close to the TRP 134 residue within the sub-domain IA, is pH-independent and binds two SDS molecules per BSA molecule. In contrast, the total number of SDS molecules bound to the second site of BSA is affected by both temperature and pH (Figure D4). To sum up, the MD-based computational

analysis corroborated the experimental findings. It provided atomistic details explaining the differences in the binding of SDS molecules in the high-affinity pH-independent and the low-affinity pH-dependent binding site.

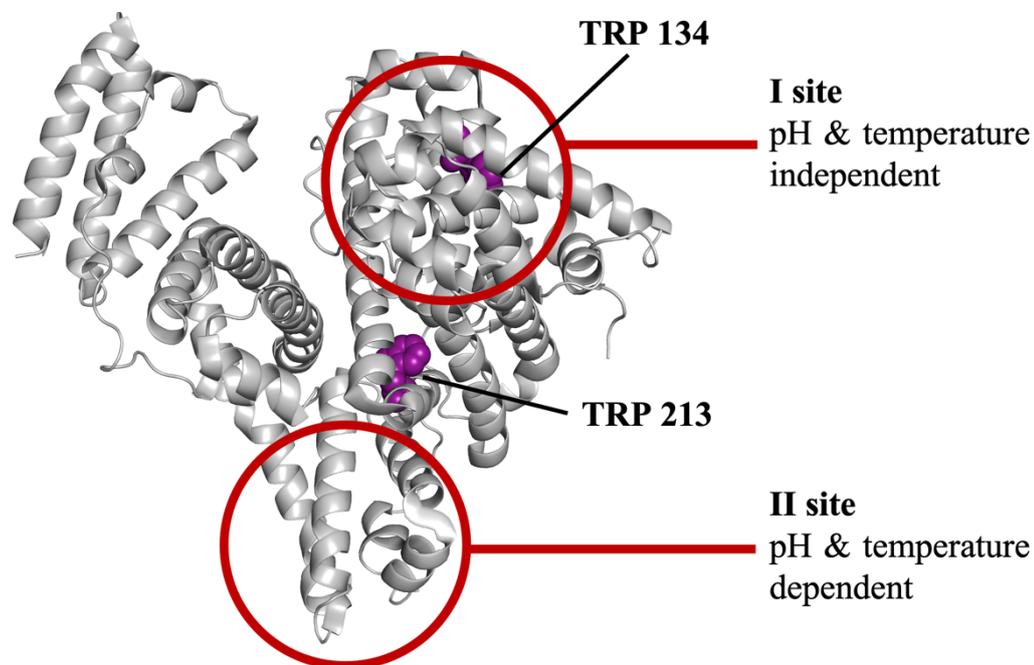


Figure D4. The secondary structure of BSA (grey cartoon) with marked TRP residues (magenta sphere representation). The first and second potential binding sites of SDS are marked with red circles and described accordingly.

4.5. Effect of Tetraphenylborate on Physicochemical Properties of Bovine Serum Albumin [D5]

In this work, the previously gathered information on BSA-SDS interactions [D4] were used to study the interactions of tetraphenylborate ions $[\text{B}(\text{Ph})_4]^-$ with BSA. Sodium tetraphenylborate is one of the numerous organo-metallic compounds for uses requiring non-aqueous solubilities, such as recent solar energy and water treatment applications. The choice of $[\text{B}(\text{Ph})_4]^-$ was motivated by the fact that these ions represent the low-molecular-weight compound with four bulky hydrophobic phenyl moieties and a negative charge capable of binding through hydrophobic interactions and a combination of hydrophobic and electrostatic forces.

This study aimed to use computational approaches to verify experimentally obtained results. Experimental methods, including ITC, DSC, and CD, were deployed to examine the physicochemical nature of these interactions and the effect of $[\text{B}(\text{Ph})_4]^-$ binding to the protein structure. Subsequently, I applied the MD approach to predict potential binding sites of $[\text{B}(\text{Ph})_4]^-$ on the BSA surface. Afterwards, I characterised and compared them regarding the binding affinities of these ions. Firstly, the protein was surrounded by 15 randomly distributed $[\text{B}(\text{Ph})_4]^-$ and the simulation was performed for 100 ns. Then, after careful examination of the MD trajectory frames for each $[\text{B}(\text{Ph})_4]^-$, the last 40% of the simulation was used for the following LIE analysis. Based on the above, two binding sites were localised for three $[\text{B}(\text{Ph})_4]^-$. One binding site was in proximity to TRP 134, in subdomain IA, as observed in the BSA-SDS system [D4]. Two ions bound to site I and one ion to site II (subdomain IIIA). According to LIE calculations, site I showed a slightly higher affinity towards $[\text{B}(\text{Ph})_4]^-$ than site II because of the more favourable van der Waals component of the free binding energy. Moreover, the theoretical calculations confirmed that hydrophobic rings of phenyl groups contributed to stabilising interactions in the BSA- $[\text{B}(\text{Ph})_4]^-$ system.

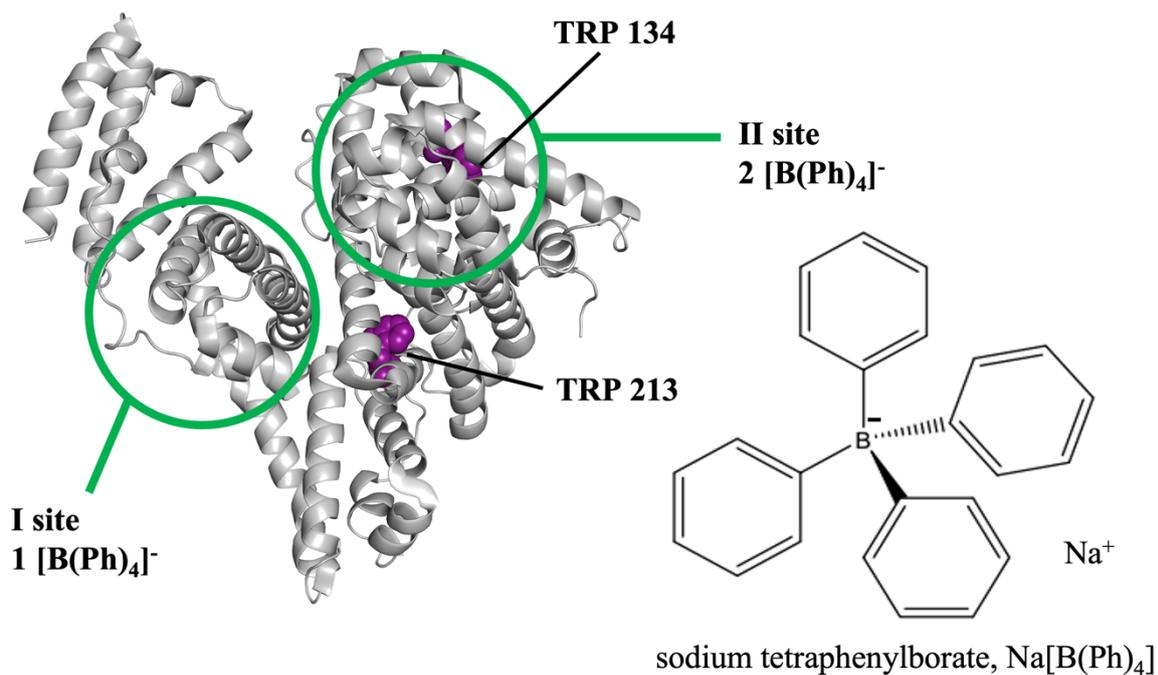


Figure D5. On the left: the secondary structure of BSA (grey cartoon) with marked TRP residues (magenta sphere representation). The first and second potential binding sites of $[\text{B}(\text{Ph})_4]^-$ are marked with green circles and described accordingly. On the right: the molecular structure of $[\text{B}(\text{Ph})_4]^-$.

4.6. Induced circular dichroism as a tool to monitor the displacement of ligands between albumins [D6]

The chirality induction in a ligand is a valuable analytical tool for studying protein-ligand interactions. This technique was applied to monitor the inversion of the induced circular dichroism (ICD) spectrum when ligands move between human and bovine serum albumin proteins (HSA and BSA, respectively). Subsequently, the experimental findings were verified with a set of computational tools.

The induction of optical activity in non-chiral compounds is intrinsically associated with intermolecular forces and the microenvironment. This photophysical phenomenon is broadly used to study host-guest complexation and supramolecular chemistry [558-560]. In protein-ligand systems, the chirality induction shows complexation [561-562]. ICD is the effect of the symmetry of non-chiral molecules perturbations resulting in the generation of non-zero rotational strengths, the theoretical counterpart of the circular dichroism phenomena [563]. The biaryl compounds are a family of molecules susceptible to ICD in host-guest complexes. This class of molecules, even without a chiral centre, can be optically active due to the energetic restraint for the free rotation of the single bond that connects the aryl rings to present axial chirality (Figure D6).

The experimental part studied the interactions of divanillate (DVT) and divanillin (DVN) with albumins and aimed at monitoring the movement of these ligands between the proteins. It was found that the ICD spectral shape of DVT and DVN depended on the albumin type, and HSA has a higher affinity for DVT and DVN than BSA.

As a first step in the computational part, molecular docking was performed for four systems, BSA-DVT, BSA-DVN, HSA-DVT and HSA-DVN, to verify binding sites. The analysis confirmed that the ligand binds to the priori established site I or site II in each case. The representative structures for each system were obtained and used for subsequent MD simulations and free energy analysis. MM-GBSA was used to calculate the binding free energy of the conformers, and with the US calculations, a dissociation pathway was proposed for each system. I applied the US protocol to investigate the energetic profile DVT/DVN orientation change in terms of the dihedral angle between the planes defined by the aromatic moieties of DVT/DVN on the BSA/HSA surfaces or in the absence of the protein. Finally, four energy minima dihedral angles conformers were identified, and the corresponding CD spectra were

calculated using the QM approach. The weighted spectra for the conformationally accessible conformers were obtained based on each conformer's Boltzmann probability distribution.

In conclusion, it was possible to monitor the movement of DVT and DVN from BSA to HSA using the inversion of their ICD signals. In other words, a competition of two proteins by the ligand was monitored by CD in real time. Additionally, deploying MD to estimate the contribution of the minima energy conformers and quantum mechanics to calculate the weighted CD spectra, the inversion of the sign was consistent with the experimental for DVT and qualitatively supported by my calculations.

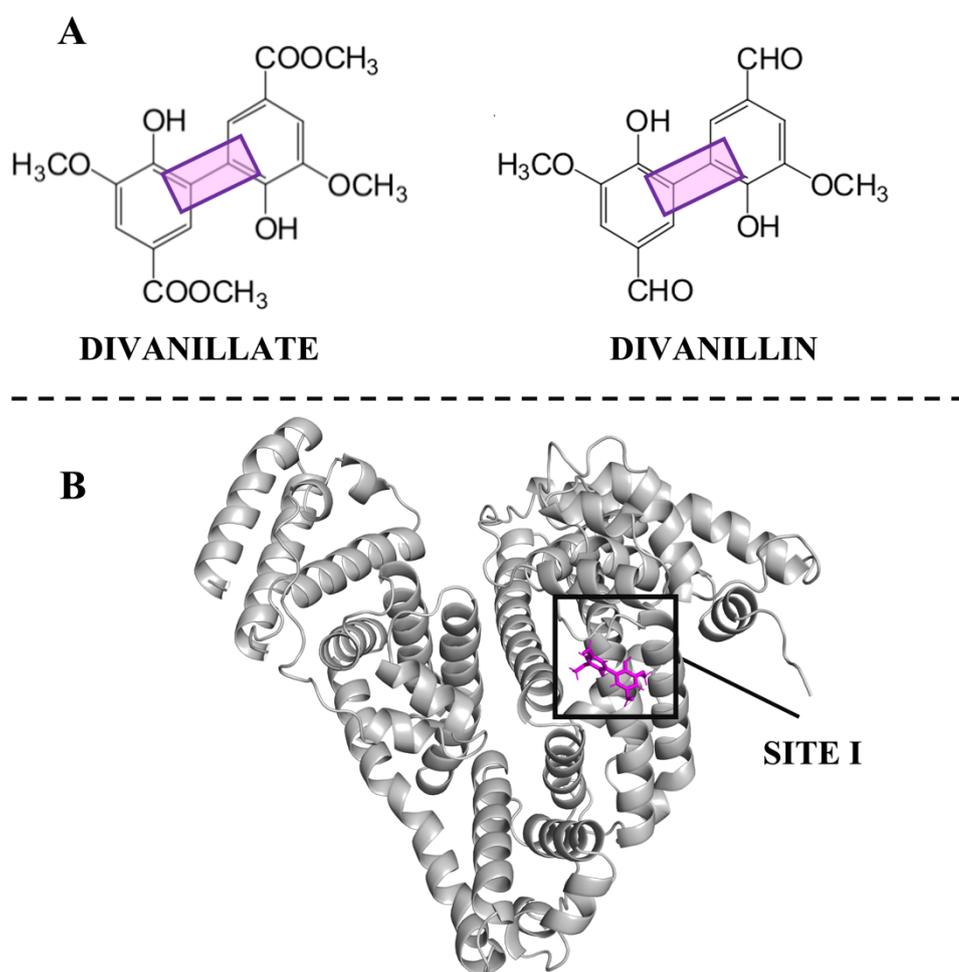


Figure D6. A: Molecular structures of divanillate (DVT) and divanillin (DVN). The purple box denotes restrained rotational chiral. B: The secondary structure of BSA (grey cartoon) with DVN (magenta licorice), located in the first binding site.

5. Conclusions and outlook

In my PhD Thesis, I aimed to analyse the role of ions in carbohydrate- and protein-containing systems using computational approaches. As stated in the “Goals of the research” section, the Thesis was divided into several parts, and the following conclusions were obtained in each of them, respectively.

- *Protein-ion-GAG and GAG-ion molecular systems [D1, D2]*

Firstly, I computationally characterised protein-ion-GAG complexes for the systems where ions are directly involved in the GAG binding [D1]. The results confirmed that Ca^{2+} is crucial in how the GAG molecules interact with the protein. In the second study, I investigated the differences between distinct sets of parameters for Ca^{2+} and their impact on two HP oligosaccharides of different lengths (dp10 and dp18) [D2]. This in-depth analysis showed that the theoretically determined Rg and H-bond propensities for a fraction of the oligosaccharides' populations in both HP dp10 and dp18 were affected by the presence of Ca^{2+} , as opposed to the flexibility of monosaccharides, the conformational space of glycosidic linkages and ring puckering.

- *The CD-ion molecular systems [D3]*

With MD-based approaches, I confirmed the initial experimentally derived hypothesis that the longer the alkyl chains bind stronger to CD [D3]. Furthermore, I proposed four potential mechanisms of SXS molecules' entrance into the CD cavity and confirmed that the initial SXS orientation affects the thermodynamic properties of the formed inclusion complex.

- *The BSA-containing molecular systems [D4-D6]*

The application of computational tools allowed for identifying the potential binding pockets in various BSA-ligand systems. Regarding the BSA-SDS system, two potential binding sites were localised: site I, pH and temperature-independent, and site II, both pH and temperature-dependent [D4]. Further, I used the MD approach to predict potential binding sites of $[\text{B}(\text{Ph})_4]^-$ on the BSA surface, and I subsequently characterised and compared them in terms of binding affinities for these ions [D5]. Two ions were predicted to bind to site I and one ion to site II. The calculations also confirmed that hydrophobic rings of phenyl groups contributed to stabilising the interactions in the BSA- $[\text{B}(\text{Ph})_4]^-$ complex. At last, computational tools were deployed to identify the binding sites in DVT/DVN-BSA/HSA systems. Two types of US calculations were

used: to calculate the dissociation pathway for each complex and to identify the conformations corresponding to the free energy minima in the absence and presence of the protein as a function of the intramolecular dihedral angles. Finally, the corresponding CD spectra were calculated using the quantum chemistry approach.

The data obtained during my PhD research show the robust nature of the computational approaches in analysing particular biologically active systems. Moreover, my research was not only complementary to the experimental findings but also provided experimentally inaccessible details, including the putative mechanism of entrance and formation of two inclusion complexes [D3], localisation of binding sites in the BSA-ligand systems [D4, D5] and the preferred ligand's conformations in the BSA/HSA pockets [D6]. Additionally, I highlighted and discussed the limitations of the applied protocols pointing at challenges persisting in the state-of-the-art *in silico* tools to study protein-ion-GAG systems [D7]. All in all, my findings confirm that computational approaches have great potential, from understanding molecular interactions at the atomic level to drug design applications.

6. Literature

- [1] Kraemer WJ, Ratamess NA, Hymer WC, Nindl BC, Fragala MS. *Frontiers in endocrinology*, **2020**:33.
- [2] Megha KB, Mohanan PV. *International Journal of Biological Macromolecules*, **2021**, 169:28-38.
- [3] Sankaran VG, Xu J, Orkin SH. *British journal of haematology*, **2010**, 149(2):181-194.
- [4] Morikawa K, Oseko F, Morikawa S. *Leukemia & lymphoma*, **1995**, 18(5-6):429-33.
- [5] Heissler SM, Sellers JR. *Journal of molecular biology*, **2016**, 428(9):1927-46.
- [6] Rose DR, Chaudet MM, Jones K. *Journal of pediatric gastroenterology and nutrition*, **2018**, 66:S11-3.
- [7] Dominguez R, Holmes KC. *Annual review of biophysics*, **2011**, 40:169.
- [8] Fischer E. *Berichte der deutschen chemischen Gesellschaft*, **1894**, 27(3):2984–2993.
- [9] Koshland Jr DE. *Proceedings of the National Academy of Sciences*, **1958**, 44(2): 98–104.
- [10] Tobi D, Bahar I. *Proceedings of the National Academy of Sciences*, **2005**, 102(52): 18908-18913.
- [11] Csermely P, Palotai R, Nussinov R. *Nature Precedings*, **2010**, 35, 539–546.
- [12] Janin J, Rodier F. *Proteins: Structure, Function, and Bioinformatics*, **1995**, 23(4): 580-587.
- [13] Otlewski J, Apostoluk W. *Acta Biochimica Polonica*, **1997**, 44(3):367-387.
- [14] Müller-Dethlefs K, Hobza P. *Chemical Reviews*, **2000**, 100(1):143-168.
- [15] McCammon JA. *Current opinion in structural biology*, **1998**, 8(2):245–249.
- [16] Chu WT, Yan Z, Chu X, Zheng X, Liu Z, Xu L, Zhang K, Wang J. *Reports on Progress in Physics*, **2021**, 9.
- [17] Cohen NC. *Gulf Professional Publishing*, **1996**, p.1–361.
- [18] Warren, G. L. et al. *Computational and Structural Approaches to Drug Discovery: Ligand-Protein Interactions*. RSC Publishing (**2008**).
- [19] Bolon DN, Grant RA, Baker TA, Sauer RT. *Proceedings of the National Academy of Sciences*, **2005**, 102(36):12724–12729.
- [20] Perozzo R, Folkers G, Scapozza L. *Journal of Receptors and Signal Transduction*, **2004**, 24(1-2):1–52.
- [21] Chaires JB. *Annual review of biophysics*, **2008**, 37(1):135–151.

- [22] Zhang Y, Sun T, Jiang C. *Acta Pharmaceutica Sinica B*, **2018**, 8(1):34-50.
- [23] Cao Y, Wang B. *International journal of molecular sciences*, **2009**, 10(4):1514-24.
- [24] Elzoghby AO, Samy WM, Elgindy NA. *Journal of controlled release*, **2012**, 161(1):38-49.
- [25] Mizrahy S, Peer D. *Chemical Society Reviews*, **2012**, 41(7):2623-2640.
- [26] Malafaya PB, Silva GA, Reis RL. *Advanced drug delivery reviews*, **2007**, 59(4-5):207-233.
- [27] Yoo JW, Irvine DJ, Discher DE, Mitragotri S. *Nature reviews Drug discovery*, **2011**, 10(7):521-35.
- [28] Nitta SK, Numata K. *International journal of molecular sciences*, **2013**, 14(1):1629-1654.
- [29] Sansonno DE, DeTomaso P, Papanice MA, Manghisi OG. *Journal of immunological methods*, **1986**, 90(1):131-136.
- [30] Chu AC, Tsang SY, Lo EH, Fung KP. *Life sciences*, **2001**, 70(5):591-601.
- [31] Andersen JT, Dalhus B, Viuff D, Ravn BT, Gunnarsen KS, Plumridge A, Bunting K, Antunes F, Williamson R, Athwal S, Allan E. *Journal of Biological Chemistry*, **2014**, 289(19):13492-13502.
- [32] Cummings RD, Pierce JM. *Handbook of glycomics*, Academic Press, **2009**.
- [33] Cummings RD, Pierce JM. *Chemistry & biology*, **2014**, 21(1):1-15.
- [34] Hudak JE, Bertozzi CR. *Chemistry & biology*, **2014**, 21(1):16-37.
- [35] Sharon N. *Pure and Applied Chemistry*, **1998**, 60(9):1389-1394.
- [36] Perrimon N, Bernfield M. *Seminars in cell & developmental biology*, **2001**, 12(2):65-67.
- [37] Ruoslahti E, Yamaguchi Y. *Cell*, **1991**, 64(5):867-869.
- [38] Pellegrini L, Burke DF, von Delft F, Mulloy B, Blundell TL. *Nature*, **2000**, 407(6807):1029-1034.
- [39] Proudfoot AE. *Biochemical Society Transactions*, **2006**, 34(3):422-426.
- [40] Misra S, Hascall VC, Atanelishvili I, Moreno Rodriguez R, Markwald RR, Ghatak S. *International journal of cell biology*, **2015**.
- [41] Maiti S. *Polysaccharide Carriers for Drug Delivery*, **2019**, 1:49-77.
- [42] Berdiaki A, Neagu M, Giatagana EM, Kuskov A, Tsatsakis AM, Tzanakakis GN, Nikitovic D. *Biomolecules*, **2021**, 11(3):395.

- [43] Wieboldt R, Läubli H. *American Journal of Physiology-Cell Physiology*, **2022**, 322(6):C1187-200.
- [44] Mao HQ, Roy K, Troung-Le VL, Janes KA, Lin KY, Wang Y, August JT, Leong KW. *Journal of controlled release*, **2001**, 70(3):399-421.
- [45] Belinskaia DA, Voronina PA, Shmurak VI, Jenkins RO, Goncharov NV. *International Journal of Molecular Sciences*, **2021**, 22(19):10318.
- [46] Fasano M, Curry S, Terreno E, Galliano M, Fanali G, Narciso P, Notari S, Ascenzi P. *IUBMB life*, **2005**, 57(12):787-796.
- [47] Bteich M. *Heliyon*, **2019**, 5(11): e02879.
- [48] Afratis N, Gialeli C, Nikitovic D, Tsegenidis T, Karousou E, Theocharis AD, Pavão MS, Tzanakakis GN, Karamanos NK. *The FEBS Journal*, **2012**, 279(7):1177-1197,
- [49] Price RD, Berry MG, Navsaria HA. *Journal of Plastic, Reconstructive & Aesthetic Surgery*, **2007**, 60(10):1110-1119.
- [50] Sugahara K, Mikami T, Uyama T, Mizuguchi S, Nomura K, Kitagawa H. *Current opinion in structural biology*, **2003**, 13(5):612-620.
- [51] Kreuger J, Spillmann D, Li JP, Lindahl U. *The Journal of cell biology*, **2006**, 174(3):323-327.
- [52] Saravanan R. *Advances in food and nutrition research*, **2014**, 72:45-60.
- [53] Funderburgh JL. *Glycobiology*, **2000**, 10(10):951-958.
- [54] Varki A, Cummings RD, Aebi M, Packer NH, Seeberger PH, Esko JD, Stanley P, Hart G, Darvill A, Kinoshita T, Prestegard JJ. *Glycobiology*, **2015**, 25(12):1323-1324.
- [55] Gandhi NS, Mancera RL. *Chemical biology & drug design*, **2008**, 72(6):455-482.
- [56] Sasarman F, Maftai C, Campeau PM, Brunel-Guitton C, Mitchell GA, Allard P. *Journal of inherited metabolic disease*, **2016**, 39(2):173-188.
- [57] Ghiselli G. *Medicinal research reviews*, **2017**, 37(5):1051-1094.
- [58] Necas JB, Bartosikova L, Brauner P, Kolar JJ. *Veterinarni medicina*, **2008**, 53(8), 397-411.
- [59] Mende M, Bednarek C, Wawryszyn M, Sauter P, Biskup MB, Schepers U, Bräse S. *Chemical Reviews*, **2016**, 116(14):8193-8255.
- [60] Neuman MG, Nanau RM, Oruña-Sanchez L, Coto G. *Journal of pharmacy & pharmaceutical sciences*, **2015**, 18(1):53-60.

- [61] Bukhari SNA, Roswandi NL, Waqas M, Habib H, Hussain F, Khan S, Sohail M, Ramli NA, Thu HE, Hussain Z. *International journal of biological macromolecules*, **2018**, 120:1682-1695.
- [62] Chen WYJ. *Wound Repair and Regeneration*, **1999**, 7(2):79-89.
- [63] Lamari FN, Theocharis AD, Asimakopoulou AP, Malavaki CJ, Karamanos NK. *Biomedical Chromatography*, **2006**, 20(6-7):539-550.
- [64] Martel-Pelletier J, Farran A, Montell E, Vergés J, Pelletier JP. *Molecules*, **2015**, 20(3):4277-4289.
- [65] Trowbridge JM, Gallo RL. *Glycobiology*, **2002**, 12(9):117R-125R.
- [66] Fransson LÅ. *Journal of Biological Chemistry*, **1968**, 243(7):1504-1510.
- [67] Casu B, Guerrini M, Torri G. *Current pharmaceutical design*, **2004**, 10(9):939-949.
- [68] Rabenstein DL. *Natural product reports*, **2002**, 19(3):312-331.
- [69] Vallet SD, Clerc O, Ricard-Blum S. *Journal of Histochemistry & Cytochemistry*, **2021**, 69(2):93-104.
- [70] Capila I, Linhardt RJ. *Angewandte Chemie International Edition*, **2002**, 41(3):390-412.
- [71] Häcker U, Nybakken K, Perrimon N. *Nature reviews Molecular cell biology*, **2005**, 6(7):530-541.
- [72] Shriver Z, Capila I, Venkataraman G, Sasisekharan R. *Heparin-A Century of Progress*, **2012**, 159-176.
- [73] Vlodaysky I, Abboud-Jarrous G, Elkin M, Naggi A, Casu B, Sasisekharan R, Ilan N. *Pathophysiology of haemostasis and thrombosis*, **2006**, 35(1-2):116-127.
- [74] Sasisekharan R, Raman R, Prabhakar V. *Annual Review of Biomedical Engineering*, **2006**, 8:181-231.
- [75] Pomin VH. *International journal of biological macromolecules*, **2015**, 72:282-289.
- [76] Caterson B, Melrose J. *Glycobiology*, **2018**, 28(4):182-206.
- [77] Belting M. *Thrombosis research*, **2014**, 133:S95-S101.
- [78] Hansen C, Otto E, Kuhlemann K, Förster G, Kahaly GJ. *Clinical and experimental rheumatology*, **1996**, 14:S59-S67.
- [79] Ariga T, Miyatake T, Yu RK. *Journal of neuroscience research*, **2010**, 88(11):2303-2315.
- [80] Lehri-Boufala S, Ouidja MO, Barbier-Chassefière V, Henault E, Raisman-Vozari R, Garrigue-Antar L, Papy-Garcia D, Morin C, *PloS one*, **2015**, 10(1):e0116641.

- [81] Wang JY, Roehrl MH. *Proceedings of the National Academy of Sciences*, **2002**, 99(22):14362-14367.
- [82] Kim SY, Jin W, Sood A, Montgomery DW, Grant OC, Fuster MM, Fu L, Dordick JS, Woods RJ, Zhang F, Linhardt RJ. *BioRxiv*, **2020**, 2020-04.
- [83] Scott RA, Panitch A. *Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology*, **2013**, 5(4):388-398.
- [84] Scharnweber D, Hübner L, Rother S, Hempel U, Anderegg U, Samsonov SA, Pisabarro MT, Hofbauer L, Schnabelrauch M, Franz S, Simon J. *Journal of Materials Science: Materials in Medicine*, **2015**, 26(9):1-10.
- [85] Köwitsch A, Zhou G, Groth T. *Journal of tissue engineering and regenerative medicine*, **2018**, 12(1):e23-e41.
- [86] Zhang J, Xu X, Rao NV, Argyle B, McCoard L, Rusho WJ, Kennedy TP, Prestwich GD, Krueger G. *PloS one*, **2011**, 6(2):e16658.
- [87] Lee WY, Savage JR, Zhang J, Jia W, Oottamasathien S, Prestwich GD. *PloS one*, **2013**, 8(10):e77854.
- [88] Savage JR, Pulsipher A, Rao NV, Kennedy TP, Prestwich GD, Ryan ME, Lee WY. *PloS one*, **2016**, 11(6):e0157310.
- [89] Bojarski KK, Becher J, Riemer T, Lemmnitzer K, Möller S, Schiller J, Schnabelrauch M, Samsonov SA. *Journal of Molecular Structure*, **2019**, 1197:401-416.
- [90] Putignano V, Rosato A, Banci L, Andreini C. **2018**, 46(D1):D459-64.
- [91] Chen AY, Adamek RN, Dick BL, Credille CV, Morrison CN, Cohen SM. *Chemical reviews*, **2018**, 119(2):1323-4155.
- [92] Liu J, Chakraborty S, Hosseinzadeh P, Yu Y, Tian S, Petrik I, Bhagi A, Lu Y. *Chemical reviews*, **2014**, 114(8):4366-4469.
- [93] Casalino L, Magistrato A. *Inorganica Chimica Acta*, **2016**, 452:73-81.
- [94] Hunsaker EW, Franz KJ. *Inorganic chemistry*, **2019**, 58(20):13528-13545.
- [95] Li P, Merz Jr KM. *Chemical reviews*, **2017**, 117(3):1564-1686.
- [96] Kuang Y, Wang Q. *Cancer letters*, **2019**, 464:56-61.
- [97] Da Silva JF, Williams RJ. Oxford University Press: **2001**.
- [98] Hambley TW. *Science*, **2007**, 318(5855):1392–1393.
- [99] Anastassopoulou J, Theophanides T. *Bioinorganic Chemistry*, **1995**, 459:209–218.
- [100] Lippard SJ. *Nature chemical biology*, **2006**, 2(10):504–507.
- [101] Jacoby, M. *Chemical & Engineering News*, **2015**, 93:8–11.

- [102] Zhang Y, Cremer PS. *Current opinion in chemical biology*, **2006**, 10(6):658-663.
- [103] Smith RA, Meade K, Pickford CE, Holley RJ, Merry CL. *Biochemical Society Transactions*, **2011**, 39:383-387.
- [104] Gasimli L, Hickey AM, Yang B, Li G, dela Rosa M, Nairn AV, Kulik MJ, Dordick JS, Moremen KW, Dalton S, Linhardt RJ. *Biochimica et Biophysica Acta (BBA)-General Subjects*, **2014**, 1840(6):1993-2003.
- [105] Napier MA, Hadler NM. *Proceedings of the National Academy of Sciences*, **1978**, 75(5):2261-2265.
- [106] Gabriel DA, Carr Jr ME. *The American journal of the medical sciences*, **1989**, 298(1):8-14.
- [107] Gribbon P, Heng BC, Hardingham TE. *Biochemical Journal*, **2000**, 350(1):329-335.
- [108] Chen X, Richter RP. *Journal of the Royal Society Interface Focus*, **2019**, 9(2):20180061.
- [109] Stevic I, Parmar N, Paredes N, Berry LR, Chan AK. *Cell biochemistry and biophysics*, **2011**, 59(3):171-178.
- [110] Bhavanandan VP, Davidson EA. *Proceedings of the National Academy of Sciences of the United States of America*, **1975**, 72(6):2032-2036.
- [111] Morris ER, Rees DA, Sanderson GR, Thom D. *Journal of the Chemical Society, Perkin Transactions 2*, **1975**, (13):1418-1425.
- [112] Casu B, Petitou M, Provasoli M, Sinay P. *Trends in biochemical sciences*, **1988**, 13(6):221-225.
- [113] Mulloy B, Forster MJ. *Glycobiology*, **2000**, 10(11): 1147-1156.
- [114] Rudd TR, Guimond SE, Skidmore MA, Duchesne L, Guerrini M, Torri G, Cosentino C, Brown A, Clarke DT, Turnbull JE, Fernig DG. *Glycobiology*, **2007**, 17(9):983-993.
- [115] Yates EA, Santini F, De Cristofano B, Payre N, Cosentino C, Guerrini M, Naggi A, Torri G, Hricovini M. *Carbohydrate Research*, **2000**, 329(1), 239-247.
- [116] Williams AJP. *Quarterly Review of the Chemical Society*, **1970**, 24:331-365.
- [117] Bush AI, Multhaup G, Moir RD, Williamson TG, Small DH, Rumble B, Pollwein P, Beyreuther K, Masters CL. *Journal of Biological Chemistry*, **1993**, 268(22):16109-16112.
- [118] Hesse L, Beher D, Masters CL, Multhaup G. *FEBS letters*, **1994**, 349(1):109-116.
- [119] Multhaup G, Bush AI, Pollwein P, Masters CL. *FEBS letters*, **1994**, 355(2):151-154.

- [120] Bush AI, Pettingell WH, Multhaup G, d.Paradis M, Vonsattel JP, Gusella F, Beyreuther K, Masters CI, Tanzi RE. *Science*, **1994**, 265:1464-1467.
- [121] Eckert R, Ragg H. *FEBS letters*, **2003**, 541(1-3):121-125.
- [122] Ricard-Blum S, Féraud O, Lortat-Jacob H, Rencurosi A, Fukai N, Dkhissi F, Vittet D, Imberty A, Olsen BR, Van Der Rest M. *Journal of Biological Chemistry*, **2004**, 279(4):2927-2936.
- [123] Zhang F, Liang X, Beaudet JM, Lee Y, Linhardt RJ. *Journal of biomedical technology and research*, **2014**, 1(1).
- [124] Srinivasan SR, Radhakrishnamurthy B, Berenson GS. *Archives of Biochemistry and Biophysics*, **1975**, 170:334-340.
- [125] Chevalier F, Lucas R, Angulo J, Martin-Lomas M, Nieto PM. *Carbohydrate research*, **2004**, 339(5):975-983.
- [126] Shao C, Zhang F, Kemp MM, Linhardt RJ, Waisman DM, Head JF, Seaton BA. *Journal of biological chemistry*, **2006**, 281(42):31689-31695.
- [127] Capila I, Hernáiz MJ, Mo YD, Mealy TR, Campos B, Dedman JR, Linhardt RJ, Seaton BA. *Structure*, **2001**, 9(1):57-64.
- [128] Weiss T, Ricard-Blum S, Moschovich L, Wineman E, Mesilaty S, Kessler E. *Journal of Biological Chemistry*, **2010**, 285(44):33867-33874.
- [129] Davis ME, Brewster ME. *Nature reviews Drug discovery*, **2004**, 3(12):1023-1035.
- [130] Carrier RL, Miller LA, Ahmed I. *Journal of Controlled Release*, **2007**, 123(2):78-99.
- [131] Cal K, Centkowska K. *European Journal of Pharmaceutics and Biopharmaceutics*, **2008**, 68(3):467-478.
- [132] Kurkov S, Loftsson T. *International journal of pharmaceutics*, **2013**, 453(1):167-180.
- [133] Del Valle EM. *Process biochemistry*, **2004**, 39(9):1033-1046.
- [134] Loftsson T, Jarho P, Másson M, Järvinen T. *Expert opinion on drug delivery*, **2005**, 2(2):335-351.
- [135] Loftsson T, Hreinsdóttir D, Másson M. *International journal of pharmaceutics*, **2005**, 302(1-2):18-28.
- [136] Iványi R, Jicsinszky L, Juvancz Z, Roos N, Otta K, Szejtli J. *Electrophoresis*, **2004**, 25(16):2675-2686.
- [137] Challa R, Ahuja A, Ali J, Khar RK. *Aaps Pharmscitech*, **2005**, 6(2):E329-E357.
- [138] Fifere A, Marangoci N, Maier S, Coroaba A, Maftei D, Pinteala M. *Beilstein journal of organic chemistry*, **2012**, 8(1):2191-201.

- [139] Whitesides GM, Grzybowski B. *Science*, **2002**, 295(5564): 2418–2421.
- [140] Seeman NC, Sleiman HF. *Nature Reviews Materials*, **2018**, 3(1):1-23.
- [141] Grzelczak M, Liz-Marzán LM, Klajn R. *Chemical Society Reviews*, **2019**, 48(5):1342–1361.
- [142] Sabadini E, Cosgrove T, Egidio FDC. *Carbohydrate Research*, **2006**, 341(2):270–274.
- [143] Raffaini G, Ganazzoli F. *Chemical physics*, **2007**, 333(2-3):128–134.
- [144] Saokham P, Muankaew C, Jansook P, Loftsson T. *Molecules*, **2018**, 23(5):1161.
- [145] Loftsson T, Vogensen SB, Brewster ME, Konráðsdóttir F. *Journal of pharmaceutical sciences*, **2007**, 96(10):2532-2546.
- [146] López-Nicolás JM, Rodríguez-Bonilla P, García-Carmona F. *Critical reviews in food science and nutrition*, **2014**, 54(2):251-276.
- [147] Zhong Y, Li W, Ran L, Hou R, Han P, Lu S, Wang Q, Zhao W, Zhu Y, Dong J. *Journal of food science*, **2020**, 85(4):1105-1113.
- [148] Tian B, Hua S, Liu J. *Carbohydrate polymers*, **2020**, 232:115805.
- [149] Szejtli J, Szenté L. *European journal of pharmaceuticals and biopharmaceutics*, **2005**, 61(3):115-125.
- [150] dos Santos C, Buera P, Mazzobre F. *Current Opinion in Food Science*, **2017**, 16:106-113.
- [151] Loftsson T, Brewster ME. *Journal of Pharmacy and Pharmacology*, **2010**, 62(11):1607–1621.
- [152] Prochowicz D, Kornowicz A, Lewiński J. *Chemical reviews*, **2017**, 117(22):13461–13501.
- [153] Liu Y, Chen Y, Gao X, Fu J, Hu L. *Critical Reviews in Food Science and Nutrition*, **2022**, 62(10):2627-2640.
- [154] Koizumi K, Sanbe H, Kubota Y, Terada Y, Takaha T. *Journal of Chromatography A*, **1999**, 852(2):407-16.
- [155] Larsen KL. *Journal of Inclusion Phenomena*, **2002**, 43:1–13.
- [156] Otzen DE. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, **2017**, 1859(4):639-649.
- [157] Otzen D. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, **2011**, 1814(5):562-591.
- [158] Khan TA, Mahler HC, Kishore RS. *European journal of pharmaceuticals and biopharmaceutics*, **2015**, 97:60-67.

- [159] Liu L, Guo QX. *Journal of inclusion phenomena and macrocyclic chemistry*, **2002**, 42(1):1–14.
- [160] Nilsson M, Valente AJ, Olofsson G, Söderman O, Bonini M. *The Journal of Physical Chemistry B*, **2008**, 112(36):11310–11316.
- [161] Benko M, Király Z. *The Journal of Chemical Thermodynamics*, **2012**, 54:211–216.
- [162] Valente AJ, Söderman O. *Advances in Colloid and Interface Science*, **2014**, 205:156–176.
- [163] Ondo D, Costas M. *Journal of colloid and interface science*, **2017**, 505:445–453.
- [164] Martín VI, Lopez-Cornejo P, Lopez-Lopez M, Blanco-Arevalo D, Moreno-Vargas AJ, Angulo M, Laschewsky A, Moyá ML. *Arabian Journal of Chemistry*, **2020**, 13(1):2318–2330.
- [165] Rekharsky MV, Inoue Y. *Chemical Reviews*, **1998**, 98(5):1875–1917.
- [166] Jiang L, Yan Y, Huang J. *Advances in colloid and interface science*, **2011**, 169(1):13–25.
- [167] Funasaki N, Ishikawa S, Neya S. *The Journal of Physical Chemistry B*, **2003**, 107(37):10094–10099.
- [168] Funasaki N, Ishikawa S, Neya S. *The Journal of Physical Chemistry B*, **2004**, 108(28):9593–9598.
- [169] Shen J, Song L, Xin X, Wu D, Wang S, Chen R, Xu G. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, **2016**, 509:512–520.
- [170] De Lisi R, Lazzara G, Milioto S, Muratore N. *The Journal of Physical Chemistry B*, **2003**, 107(47):13150–13157.
- [171] Okamoto H, Ino S, Nihei N, Ikuta N, Ueno C, Itoi A, Yoshikawa Y, Terao K, Sakamoto N. *Journal of Clinical Biochemistry and Nutrition*, **2019**, 65(2):99–108.
- [172] Gürten B, Yenigül E, Sezer AD, Malta S. *Brazilian Journal of Pharmaceutical Sciences*, **2018**, 54:1–11.
- [173] Vyas A, Saraf S, Saraf S. *Journal of inclusion phenomena and macrocyclic chemistry*, **2008**, 62(1):23–42.
- [174] Jansook P, Ogawa N, Loftsson T. *International journal of pharmaceutics*, **2018**, 535(1–2):272–284.
- [175] Brocos P, Díaz-Vergara N, Banquy X, Pérez-Casas S, Costas M, Piñeiro Á. *The Journal of Physical Chemistry B*, **2010**, 114(39):12455–12467.

- [176] Liveri VT, Cavallaro G, Giammona G, Pitarresi G, Puglisi G, Ventura C. *Thermochimica Acta*, **1992**, 199:125–132.
- [177] Saito Y, Ueda H, Abe M, Sato T, Christian SD. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, **1998**, 135(1-3):103–108.
- [178] Topchieva I, Karezin K. *Journal of colloid and interface science*, **1999**, 213(1):29–35.
- [179] Popova EI, Topchieva IN. *Russian Chemical Bulletin*, **2001**, 50(4):620–625.
- [180] Wilson LD, Verrall RE. *The Journal of Physical Chemistry B*, **1997**, 101(45):9270–9279.
- [181] Rafati AA, Bagheri A, Iloukhani H, Zarinehzad M. *Journal of molecular liquids*, **2005**, 116(1):37–41.
- [182] Benkő M, Tabajdi R, Király Z. *Journal of thermal analysis and calorimetry*, **2013**, 112(2):969–976.
- [183] Petek A, Krajnc M, Petek A. *Journal of Inclusion Phenomena and Macrocyclic Chemistry*, **2016**, 86(3):221–229.
- [184] Junquera E, Peña L, Aicart E. *Langmuir*, **1995**, 11(12):4685–4690.
- [185] Machperson Y, Palepu R, Reinsborough V. *Journal of inclusion phenomena and molecular recognition in chemistry*, **1990**, 9(2):137–143.
- [186] Sammalkorpi M, Karttunen M, Haataja M. *The Journal of Physical Chemistry B*, **2007**, 111(40):11722–11733.
- [187] Sun C, Yang J, Wu X, Huang X, Wang F, Liu S. *Biophysical journal*, **2005**, 88(5):3518–3524.
- [188] De S, Girigoswami A, Das S. *Journal of colloid and interface science*, **2005**, 285(2):562–573.
- [189] El Kadi N, Taulier N, Le Huérou JY, Gindre M, Urbach W, Nwigwe I, Kahn PC, Waks M. *Biophysical journal*, **2006**, 91(9):3397–3404.
- [190] Mackie A, Wilde P. *Advances in Colloid and Interface Science*, **2005**, 117(1-3):3–13.
- [191] Li Y, Lee JS. *Analytica Chimica Acta*, **2019**, 1063:18–39.
- [192] Bordbar AK, Sohrabi N, Gharibi H. *Bulletin of the Korean Chemical Society*, **2004**, 25(6):791–795.
- [193] Højgaard C, Sørensen HV, Pedersen JS, Winther JR, Otzen DE. *Biophysical journal*, **2018**, 115(11):2081–2086.
- [194] Shaw BF, Schneider GF, Whitesides GM. *Journal of the American Chemical Society*, **2012**, 134(45):18739–18745.

- [195] Otzen DE, Sehgal P, Westh P. *Journal of colloid and interface science*, **2009**, 329(2):273–283.
- [196] Winogradoff D, John S, Aksimentiev A. *Nanoscale*, **2020**, 12(9):5422-5434.
- [197] Lee HJ, McAuley A, Schilke KF, McGuire J. *Advanced drug delivery reviews*, **2011**, 63(13):1160-1171.
- [198] Tucker IM, Petkov JT, Penfold J, Thomas RK, Li P, Cox AR, Hedges N, Webster JR. *The Journal of Physical Chemistry B*, **2014**, 118(18):4867-4875.
- [199] Ali A, Malik NA, Uzair S, Ali M. *Molecular Physics*, **2014**, 112(20):2681-2693.
- [200] Mehan S, Aswal VK, Kohlbrecher J. *Physical Review E*, **2015**, 92(3):032713.
- [201] Maldonado F, Almela M, Otero A, Costa-López J. *Journal of protein chemistry*, **1991**, 10(2):189–192.
- [202] Mandal BB, Kund SC. *Biotechnology and bioengineering*, **2008**, 99(6):1482–1489.
- [203] Park JH, Kim MH, Jeong L, Cho D, Kwon OH, Park WH. *Journal of sol-gel science and technology*, **2014**, 71(2):364–371.
- [204] Kezwoń A, Góral I, Frączyk T, Wojciechowski K. *Colloids and Surfaces B: Biointerface*, **2016**, 148:238–248.
- [205] Kezwoń A, Wojciechowski K. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, **2016**, 509:390–400.
- [206] Pan F, Lu Z, Tucker I, Hosking S, Petkov J, Lu JR. *Journal of Colloid and Interface Science*, **2016**, 484:125–134.
- [207] Dubey P, Kumar S, Ravindranathan S, Vasudevan S, Aswal VK, Rajamohanan PR, Nisal A, Prabhune A. *Materials Chemistry and Physics*, **2018**, 203:9–16.
- [208] Diaz X, Abuin E, Lissi E. *Journal of Photochemistry and Photobiology A: Chemistry*, **2003**, 155(1-3):157–162.
- [209] Malik NA. *Applied biochemistry and biotechnology*, **2015**, 176(8):2077–2106.
- [210] Deep S, Ahluwalia JC. *Physical Chemistry Chemical Physics*, **2001**, 3(20):4583–4591.
- [211] Hansted JG, Wejse PL, Bertelsen H, Otzen DE. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, **2011**, 1814(5):713–723.
- [212] Mozzi A, Forni D, Cagliani R, Pozzoli U, Vertemara J, Bresolin N, Sironi M. *Genome biology and evolution*, **2014**, 6(11):2983–2997.
- [213] Li S, Cao Y, Geng F. *Evolutionary Bioinformatics*, **2017**, 13:1–6.
- [214] Jahanban-Esfahlan A, Panahi-Azar V, Sajedi S. *Biopolymers*, **2015**, 103(11):638-645.
- [215] Peters Jr T. *Academic press*, **1995**, Dec 21.

- [216] Ni Y, Su S, Kokot S. *Analytica chimica acta*, **2006**, 580(2):206-215.
- [217] Bujacz A. *Acta Crystallographica Section D: Biological Crystallography*, **2012**, 68(10):1278-1289.
- [218] Ameh ES. *The International Journal of Advanced Manufacturing Technology*, **2019**, 105(7):3289-3302.
- [219] Pye DA, Vivès RR, Hyde P, Gallagher JT. *Glycobiology*, **2000**, 10(11):1183-1192.
- [220] Ostrovsky O, Berman B, Gallagher J, Mulloy B, Fernig DG, Delehedde M, Ron D. *Journal of Biological Chemistry*, **2002**, 277(4):2444-2453.
- [221] Taylor KR, Rudisill JA, Gallo RL. *Journal of Biological Chemistry*, **2005**, 280(7):5300-5306.
- [222] Agostino M, Gandhi NS, Mancera RL. *Glycobiology*, **2014**, 24(9):840-851.
- [223] Shao C, Zhang F, Kemp MM, Linhardt RJ, Waisman DM, Head JF, Seaton BA. *Journal of Biological Chemistry*, **2006**, 281(42):31689-31695.
- [224] Baron MJ, Filman DJ, Prophete GA, Hogle JM, Madoff LC. *Journal of Biological Chemistry*, **2007**, 282(14):10526-10536.
- [225] Perez S, Bonnardel F, Lisacek F, Imberty A, Ricard-Blum S, Makshakova O. *Biomolecules*, **2020**, 10(12):1660.
- [226] Harata K, Endo T, Ueda H, Nagai T. *Supramolecular Chemistry*, **1998**, 9(2):143-150.
- [227] Christoforides E, Papaioannou A, Bethanis K. *Beilstein Journal of Organic Chemistry*, **2018**, 14:838-848.
- [228] Sala A, Hoossen Z, Bacchi A, Caira MR. *Molecules*, **2021**, 26(15):4427.
- [229] Takashima Y, Oizumi Y, Sakamoto K, Miyauchi M, Kamitori S, Harada A. *Macromolecules*, **2004**, 37(11):3962-3964.
- [230] Wang E, Chen G, Han C. *Chinese Journal of Chemistry*, **2011**, 29(4):617-622.
- [231] Braga SS, Gonçalves IS, Herdtweck E, Teixeira-Dias JJ. *New Journal of Chemistry*, **2003**, 27(3):597-601.
- [232] Mura P, Zerrouk N, Faucci MT, Maestrelli F, Chemtob C. *European journal of pharmaceuticals and biopharmaceutics*, **2002**, 54(2):181-191.
- [233] Li N, Zhang YH, Wu YN, Xiong XL, Zhang YH. *Journal of pharmaceutical and biomedical analysis*, **2005**, 39(3-4):824-829.
- [234] Figueiras A, Carvalho RA, Ribeiro L, Torres-Labandeira JJ, Veiga FJ. *European Journal of Pharmaceutics Biopharmaceutics*, **2007**, 67(2):531-539.
- [235] Guo P, Su Y, Cheng Q, Pan QQ, Li H. *Carbohydrate research*, **2011**, 346(7):986-990.

- [236] Speight JG, *Environmental Inorganic Chemistry for Engineers*, **2017**, 153-201.
- [237] Battistel MD, Azurmendi HF, Yu B, Freedberg DI. *Progress in nuclear magnetic resonance spectroscopy*, **2014**, 79:48-68.
- [238] Rabenstein DL, Millis KK. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology*, **1995**, 1249(1):29-36.
- [239] Lerner L, Torchia DA. *Journal of Biological Chemistry*, **1986**, 261(27):12706-11274.
- [240] Chevalier F, Angulo J, Lucas R, Nieto PM, Martín-Lomas M. *European Journal of Organic Chemistry*, **2002**, 14:2367-2376.
- [241] Guerrero-Martínez A, González-Gaitano G, Viñas MH, Tardajos G. *The Journal of Physical Chemistry B*, **2006**, 110(28):13819–13828.
- [242] Wimmer R, Aachmann FL, Larsen KL, Petersen SB. *Carbohydrate Research*, **2002**, 337(9):841–849.
- [243] Cabaleiro-Lago C, Nilsson M, Söderman O. *Langmuir*, **2005**, 21(25):11637–11644.
- [244] Krueger S. *Physica B: Condensed Matter*, **1998**, 241:1131–1137.
- [245] Zhou C, Cheng X, Zhao Q, Yan Y, Wang J, Huang J. *Langmuir*, **2013**, 29(43): 13175–81312.
- [246] Wang J, Yao M, Li Q, Yi S, Chen X. *Soft Matter*, **2016**, 12(48):9641–9648.
- [247] Naggar EF, Costello CE, Zaia J. *Journal of the American Society for Mass Spectrometry*, **2004**, 15(11):1534-1544.
- [248] Danikiewicz W. *Cyclodextrins and their complexes*, **2006**.
- [249] Brodbelt JS. *International Journal of Mass Spectrometry*, **2000**, 200(1– 3):57–69.
- [250] Jürg DM, Friess SD, Rajagopalan S, Wendt S, Zenobi R. *International Journal of Mass Spectrometry*, **2002**, 216:1–27.
- [251] Lebrilla CB. *Accounts of Chemical Research*, **2001**, 34(8):653–661.
- [252] Grigorean G, Cong X, Lebrilla CB. *International Journal of Mass Spectrometry*, **2004**, 234(1–3):71–77.
- [253] Miles AJ, Janes RW, Wallace BA. *Chemical Society Reviews*, **2021**.
- [254] Matsuo K, Namatame H, Taniguchi M, Gekko K. *Bioscience, biotechnology, and biochemistry*, **2009**, 73(3):557-561.
- [255] Zsila F. *Physical Chemistry Chemical Physics*, **2015**, 17(38):24560-24565.
- [256] Rudd TR, Skidmore MA, Guimond SE, Guerrini M, Cosentino C, Edge R, Brown A, Clarke DT, Torri G, Turnbull JE, Nichols RJ. *Carbohydrate research*, **2008**, 343(12):2184-2193.

- [257] Mura P. *Journal of pharmaceutical and biomedical analysis*, **2014**, 101:238-250.
- [258] Bertucci C, Pistolozzi M, De Simone A. *Analytical and bioanalytical chemistry*, **2010**, 398(1):155-166.
- [259] Jaffer FA, Kim DE, Quinti L, Tung CH, Aikawa E, Pande AN, Kohler RH, Shi GP, Libby P, Weissleder R. *Circulation*, **2007**, 115(17):2292-2298.
- [260] Wankar J, Salzano G, Pancani E, Benkovics G, Malanga M, Manoli F, Gref R, Fenyvesi E, Manet I. *International Journal of Pharmaceutics*, **2017**, 531(2):568-576.
- [261] Salzano G, Wankar J, Ottani S, Villemagne B, Baulard AR, Willand N, Brodin P, Manet I, Gref R. *International Journal of Pharmaceutics*, **2017**, 531(2):577-87.
- [262] Zhou X, Liang JF. *Journal of Photochemistry and Photobiology A: Chemistry*, **2017**, 349:124-128.
- [263] Ameen F, Siddiqui S, Jahan I, Nayeem SM, ur Rehman S, Tabish M. *Journal of Molecular Liquids*, **2020**, 303:112671.
- [264] Papadopoulou A, Green RJ, Frazier RA. *Journal of agricultural and food chemistry*. **2005**, 53(1):158-163.
- [265] Jahanban-Esfahlan A, Ostadrahimi A, Jahanban-Esfahlan R, Roufegarinejad L, Tabibiazar M, Amarowicz R. *International journal of biological macromolecules*, **2019**, 138:602-617.
- [266] Liedberg B, Nylander C, Lundström I. *Biosensors and Bioelectronics*, **1995**, 10(8):1-9.
- [267] Salamon Z, Brown MF, Tollin G. *Trends in biochemical sciences*, **1999**, 24(6):213-219.
- [268] Patching SG. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, **2014**, 1838(1):43-55.
- [269] Poltronieri P, Mezzolla V, Primiceri E, Maruccio G. *Foods*, **2014**, 3(3):511-526.
- [270] Yesudasu V, Pradhan HS, Pandya RJ. *Heliyon*, **2021**, 7(3):e06321.
- [271] Osmond RI, Kett WC, Skett SE, Coombe DR. *Analytical biochemistry*, **2002**, 310(2):199-207.
- [272] Cochran S, Li CP, Ferro V. *Glycoconjugate journal*, **2009**, 26(5):577-587.
- [273] Hintze V, Moeller S, Schnabelrauch M, Bierbaum S, Viola M, Worch H, Scharnweber D. *Biomacromolecules*, **2009**, 10(12):3290-3297.
- [274] Hintze V, Miron A, Moeller S, Schnabelrauch M, Wiesmann HP, Worch H, Scharnweber D. *Acta biomaterialia*, **2012**, 8(6):2144-2152.

- [275] Hintze V, Samsonov SA, Anselmi M, Moeller S, Becher J, Schnabelrauch M, Scharnweber D, Pisabarro MT. *Biomacromolecules*, **2014**, 15(8):3083-3092.
- [276] Pan T, Wong BS, Liu T, Li R, Petersen RB, Sy MS. *Biochemical Journal*, **2002**, 368(1):81-90.
- [277] Holdgate GA, Ward WH. *Drug discovery today*, **2005**, 10(22):1543-1550.
- [278] Saboury AA. *Journal of the Iranian Chemical Society*, **2006**, 3(1):1-21.
- [279] Bouchemal K. *Drug discovery today*, **2008**, 13(21-22):960-972.
- [280] Sepuru KM, Nagarajan B, Desai UR, Rajarathnam K. *Journal of Biological Chemistry*, **2018**, 293(46):17817-28.
- [281] Köhling S, Blaszkiewicz J, Ruiz-Gómez G, Fernández-Bachiller MI, Lemmnitzer K, Panitz N, Beck-Sickinger AG, Schiller J, Pisabarro MT, Rademann J. *Chemical science*, **2019**, 10(3):866-878.
- [282] Walkowiak JJ, Ballauff M, Zimmermann R, Freudenberg U, Werner C. *Biomacromolecules*, **2020**, 21(11):4615-4625.
- [283] Dutta AK, Sepuru KM, Rösgen J, Rajarathnam K. *In Glycosaminoglycans*, **2022** (pp. 307-317). Humana, New York, NY.
- [284] Bouchemal K, Agnely F, Koffi A, Ponchel G. *Journal of colloid and interface science*, **2009**, 338(1):169-76.
- [285] Daoud-Mahammed S, Couvreur P, Bouchemal K, Chéron M, Lebas G, Amiel C, Gref R. *Biomacromolecules*, **2009**, 10(3):547-554.
- [286] Othman M, Bouchemal K, Couvreur P, Gref R. *International journal of pharmaceuticals*, **2009**, 379(2):218-225.
- [287] Daoud-Mahammed S, Agnihotri S, Bouchemal K, Kloters S, Couvreur P, Gref R. *Current Nanoscience*, **2010**, 6(6):654-665.
- [288] Sajeesh S, Bouchemal K, Marsaud V, Vauthier C, Sharma CP. *Journal of Controlled Release*, **2010**, 147(3):377-384.
- [289] Mazzaferro S, Bouchemal K, Gallard JF, Iorga BI, Cheron M, Gueutin C, Steinmesse C, Ponchel G. *International journal of pharmaceuticals*, **2011**, 416(1):171-180.
- [290] Vial F, Cousin F, Bouteiller L, Tribet C. *Langmuir*, **2009**, 25(13):7506-7513.
- [291] Roques C, Bouchemal K, Ponchel G, Fromes Y, Fattal E. *Journal of controlled release*, **2009**, 138(1):71-77.
- [292] Bouchemal K, Agnely F, Koffi A, Djabourov M, Ponchel G. *Journal of molecular recognition*, **2010**, 23(4):335-342.

- [293] Bellot M, Bouteiller L. *Langmuir*, **2008**, 24(24):14176-14182.
- [294] Feng X, Leduc M, Pelton R. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, **2008**, 317(1-3):535-542.
- [295] Scrocco E, Tomasi J. In *New concepts II*, **1973**, (pp. 95-170). Springer, Berlin, Heidelberg.
- [296] Petrongolo C, Preston HJ, Kaufman JJ. *International Journal of Quantum Chemistry*, **1978**, 13(4):457-468.
- [297] Weiner PK, Langridge R, Blaney JM, Schaefer R, Kollman PA. *Proceedings of the National Academy of Sciences*, **1982**, 79(12):3754-3758.
- [298] Warwicker J, Watson HC. *Journal of molecular biology*, **1982**, 157(4):671-679.
- [299] Klapper I, Hagstrom R, Fine R, Sharp K, Honig B. *Proteins*, **1986**, 1:47-59.
- [300] Samsonov SA, Gehrecke JP, Pisabarro MT. *Journal of Chemical Information and Modeling*, **2014**, 54(2):582-592.
- [301] Rother S, Samsonov SA, Hofmann T, Blaszkiewicz J, Köhling S, Moeller S, Schnabelrauch M, Rademann J, Kalkhof S, von Bergen M, Pisabarro MT, Scharnweber D, Hintze V. *Acta biomaterialia*, **2016**, 45:143-154.
- [302] Lengauer T, Rarey M. *Current opinion in structural biology*, **1996**, 6(3):402-406.
- [303] Guedes IA, de Magalhães CS, Dardenne LE. *Biophysical reviews*, **2014**, 6(1):75-87.
- [304] Ferreira LG, Dos Santos RN, Oliva G, Andricopulo AD. *Molecules*, **2015**, 20(7):13384-13421.
- [305] Goodford PJ. *Journal of medicinal chemistry*, **1985**, 28(7):849-857.
- [306] Kastenholz MA, Pastor M, Cruciani G, Haaksma EE, Fox T. *Journal of medicinal chemistry*, **2000**, 43(16):3033-3044.
- [307] Levitt DG, Banaszak LJ. *Journal of molecular graphics*, **1992**, 10(4):229-234.
- [308] Laskowski RA. *Journal of molecular graphics*, **1995**, 13(5):323-330.
- [309] Glaser F, Morris RJ, Najmanovich RJ, Laskowski RA, Thornton JM. *PROTEINS: Structure, Function, and Bioinformatics*, **2006**, 62(2):479-488.
- [310] Brady GP, Stouten PF. *Journal of computer-aided molecular design*, **2000**, 14(4):383-401.
- [311] Mezei M. *Journal of Molecular Graphics and Modelling*, **2003**, 21(5):463-472.
- [312] Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE. *Journal of molecular biology*, **1982**, 161(2):269-288.

- [313] Rarey M, Kramer B, Lengauer T, Klebe G. *Journal of molecular biology*, **1996**, 261(3):470-489.
- [314] Kramer B, Rarey M, Lengauer T. *Proteins: Structure, Function, and Bioinformatics*, **1999**, 37(2):228-241.
- [315] Liu M, Wang S. *Journal of computer-aided molecular design*, **1999**, 13(5):435-451.
- [316] Kirkpatrick S, Gelatt Jr CD, Vecchi MP. *Science*, **1983**, 220(4598):671-680.
- [317] Goodsell DS, Olson AJ. *Proteins: Structure, Function, and Bioinformatics*, **1990**, 8(3):195-202.
- [318] Moré JJ, Wu Z. *Journal of Global Optimization*, **1999**, 15(3):219-234.
- [319] Fogel LJ, Owens AJ, Walsh MJ. *Behavioral science*, **1966**, 11(4):253-272.
- [320] Eiben AE, Smith JE. *Berlin: springer*, **2003**.
- [321] Kollman P. *Chemical reviews*, **1993**, 93(7):2395-2417.
- [322] Carlson HA, Jorgensen WL. *The Journal of Physical Chemistry*, **1995**, 99(26):10667-10673.
- [323] Åqvist J, Luzhkov VB, Brandsdal BO. *Accounts of chemical research*, **2002**, 35(6):358-365.
- [324] Verkhivker G, Appelt K, Freer ST, Villafranca JE. *Protein Engineering, Design and Selection*, **1995**, 8(7):677-691.
- [325] Wallqvist A, Jernigan RL, Covell DG. *Protein Science*, **1995**, 4(9):1881-1903.
- [326] Muegge I, Martin YC. *Journal of medicinal chemistry*, **1999**, 42(5):791-804.
- [327] Mitchell JB, Laskowski RA, Alex A, Thornton JM. *Journal of Computational Chemistry*, **1999**, 20(11):1165-1176.
- [328] Ishchenko AV, Shakhnovich EI. *Journal of medicinal chemistry*, **2002**, 45(13):2770-2780.
- [329] Feher M, Deretey E, Roy S. *Journal of chemical information and computer sciences*, **2003**, 43(4):1316-1327.
- [330] Gehlhaar DK, Verkhivker GM, Rejto PA, Sherman CJ, Fogel DR, Fogel LJ, Freer ST. *Chemistry & biology*, **1995**, 2(5):317-24.
- [331] Head RD, Smythe ML, Oprea TI, Waller CL, Green SM, Marshall GR. *Journal of the American Chemical society*, **1996**, 118(16):3959-3969.
- [332] Jain AN. *Journal of computer-aided molecular design*, **1996**, 10(5):427-440.
- [333] Böhm HJ. *Journal of computer-aided molecular design*, **1998**, 12(4):309.

- [334] Verkhivker GM, Bouzida D, Gehlhaar DK, Rejto PA, Arthurs S, Colson AB, Freer ST, Larson V, Luty BA, Marrone T, Rose PW. *Journal of computer-aided molecular design*, **2000**, 14(8):731-751.
- [335] Griffith AR, Rogers CJ, Miller GM, Abrol R, Hsieh-Wilson LC, Goddard WA. *Proceedings of the National Academy of Sciences*, **2017**, 114(52):13697–13702.
- [336] Uciechowska-Kaczmarzyk U, de Beauchene IC, Samsonov SA. *Journal of Molecular Graphics and Modelling*, **2019**, 90:42-50.
- [337] Frank M. *Carbohydrates as Drugs*, **2014**, 12:53-72.
- [338] Nivedha AK, Thieker DF, Makeneni S, Hu H, Woods RJ. *Journal of chemical theory and computation*, **2016**, 12(2):892–901.
- [339] Boittier ED, Burns JM, Gandhi NS, Ferro V. *Journal of Chemical Information and Modeling*, **2020**, 60(12):6328–6343.
- [340] Eberhardt J, Santos-Martins D, Tillack AF, Forli S. *Journal of Chemical Information and Modeling*, **2021**, 61(8):3891–3898.
- [341] Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelly M, Perry JK, Shaw DE, Francis P, Shenkin PS. *Journal of medicinal chemistry*, **2004**, 47(7):1739–1749.
- [342] Desta IT, Porter KA, Xia B, Kozakov D, Vajda S. *Structure*, **2020**, 28(9):1071–1081.
- [343] Mottarella SE, Beglov D, Beglova N, Nugent MA, Kozakov D, Vajda S. *Journal of chemical information and modelling*, **2014**, 54(7):2068-2078.
- [344] Van Zundert GCP, Rodrigues JPGLM, Trellet M, Schmitz C, Kastiris PL, Karaca E, Melquionhorougvan Dijk M, deVries SJ, Bonvin AMJJ. *Journal of molecular biology*, **2016**, 428(4):720–725.
- [345] Grosdidier A, Zoete V, Michielin O. *Nucleic acids research*, **2011**, 39(suppl_2):W270-W277.
- [346] Clausen TM, Sandoval DR, Spliid CB, Pihl J, Perrett HR, Painter CD, Narayanan A, Majowicz SA, Kwong EM, McVicar RN, Thacker BE, Glass CA, Yang Z, Torres JL, Golden GJ, Bartles PL, Porell RN, garretson AF, Laubach L, Feldman J, Yin X, Pu Y, Hauser BM, Cardonna TM, Kellman BP, Martino C, Gordts PLSM, Chanda SK, Schmidt AG, Godula K, Leibel SL, Jose J, Corbett KD, Ward AB, Carlin AF, Esko JD. *Cell*, **2020**, 183(4):1043-1057.
- [347] Joseph PRB, Mosier PD, Desai UR, Rajarathnam K. *Biochemical Journal*, **2015**, 472(1):121–133.

- [348] Vuorio J, Vattulainen I, Martinez-Seara H. *PLoS computational biology*, **2017**, 13(7):e1005663.
- [349] Sandoval Daniel R, Toledo Alejandro Gomez, Painter Chelsea D, Tota Ember M, Osman Sheikh XM, West Alan MV, Frank Martin M, Wells Lance, Xu Ding, Bicknell Roy, Corbett Xde Kevin D. *Journal of Biological Chemistry*, **2020**, 295(9):2804–2821.
- [350] Samsonov SA, Zacharias M, Beauchene IC. *Journal of Computational Chemistry*, **2019**, 40(14):1429–1439.
- [351] Siebenmorgen T, Engelhard M, Zacharias M. *Journal of Computational Chemistry*, **2020**, 41(15):1436–1447.
- [352] Maszota-Zieleniak M, Marcisz M, Kogut MM, Siebenmorgen T, Zacharias M, Samsonov SA. *Journal of Computational Chemistry*, **2021**, 42(15):1040–1053.
- [353] Samsonov SA, Teyra J, Pisabarro MT. *Journal of computer-aided molecular design*, **2011**, 25(5):477-489.
- [354] Marcisz M, Gaardløs M, Bojarski KK, Siebenmorgen T, Zacharias M, Samsonov SA. *Journal of Computational Chemistry*, **2022**, 43(24):1633-1640
- [355] Marcisz M, Zacharias M, Samsonov SA. *Journal of chemical information and modelling*, **2021**, 61(9):4475-4485.
- [356] Tao X, Huang Y, Wang C, Chen F, Yang L, Ling L, Che Z, Chen X. *International Journal of Food Science and Technology*, **2020**, 55(1):33–45.
- [357] Illapakurthy AC, Sabnis YA, Avery BA, Avery MA, Wyandt CM. *Journal of pharmaceutical sciences*, **2003**, 92(3):649–655.
- [358] Wang Z, Sun H, Yao X, Li D, Xu L, Li Y, Tian S, Hou T. *Physical Chemistry Chemical Physics*, **2016**, 18(18):12964-12975.
- [359] Alder BJ, Wainwright TE. *Transport processes in statistical mechanics*, **1958**, 27:97-131.
- [360] Van Gunsteren WF. *Computer simulation of biomolecular system: theoretical and experimental applications*, **1993**, 2:3-6.
- [361] Frenkel D, Smit B. *Elsevier*, **2001**.
- [362] Scheraga HA, Khalili M, Liwo A. *Annual review of physical chemistry*, **2007**, 58(1):57-83.
- [363] Durrant JD, McCammon JA. *BMC biology*, **2011**, 9(1):1-9.
- [364] Paquet E, Viktor HL. *BioMed research international*, **2015**.
- [365] Karplus M, Petsko GA. *Nature*, **1990**, 347(6294):631-639.

- [366] Wang J, Cieplak P, Kollman PA. *Journal of Computational Chemistry*, **2000**, 21(12):1049-1074.
- [367] Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. *Journal of chemical theory and computation*, **2015**, 11(8):3696-3713.
- [368] Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I, Mackerell Jr AD. *Journal of Computational Chemistry*, **2010**, 31(4):671-690.
- [369] Schmid N, Eichenberger AP, Choutko A, Riniker S, Winger M, Mark AE, van Gunsteren WF. *European Biophysics Journal*, **2011**, 40(7):843-856.
- [370] Lindorff-Larsen K, Piana S, Dror RO, Shaw DE. *Science*, **2011**, 334(6055):517-520.
- [371] Feig M, Yu I, Wang PH, Nawrocki G, Sugita Y. *The Journal of Physical Chemistry B*, **2017**, 121(34):8009-8025.
- [372] Takahashi KZ, Nishimura R, Yamato N, Yasuoka K, Masubuchi Y. *Scientific reports*, **2017**, 7(1):1-7.
- [373] Marrink SJ, Tieleman DP. *Chemical Society Reviews*, **2013**, 42(16):6801-6822.
- [374] Alessandri R, Souza PC, Thallmair S, Melo MN, De Vries AH, Marrink SJ. *Journal of chemical theory and computation*, **2019**, 15(10):5448-5460.
- [375] Souza PC, Alessandri R, Barnoud J, Thallmair S, Faustino I, Grünewald F, Patmanidis I, Abdizadeh H, Bruininks BM, Wassenaar TA, Kroon PC. *Nature methods*, **2021**, 18(4):382-388.
- [376] Maisuradze GG, Senet P, Czaplewski C, Liwo A, Scheraga HA. *The Journal of Physical Chemistry A*, **2010**, 114(13):4471-4485.
- [377] Kar P, Gopal SM, Cheng YM, Predeus A, Feig M. *Journal of chemical theory and computation*, **2013**, 9(8):3769-88.
- [378] Davtyan A, Schafer NP, Zheng W, Clementi C, Wolynes PG, Papoian GA. *The Journal of Physical Chemistry B*, **2012**, 116(29):8494-8503.
- [379] Basdevant N, Borgis D, Ha-Duong T. *Journal of Chemical Theory and Computation*, **2013**, 9(1):803-813.
- [380] Darré L, Machado MR, Brandner AF, González HC, Ferreira S, Pantano S. *Journal of chemical theory and computation*, **2015**, 11(2):723-739.
- [381] Machado MR, Barrera EE, Klein F, Sónora M, Silva S, Pantano S. *Journal of Chemical Theory and Computation*, **2019**, 15(4):2719-2733.

- [382] Marrink SJ, Corradi V, Souza PC, Ingólfsson HI, Tieleman DP, Sansom MS. *Chemical reviews*, **2019**, 119(9):6184–6226.
- [383] López CA, Rzepiela AJ, de Vries AH, Dijkhuizen L, Hünenberger PH, Marrink SJ. *Journal of Chemical Theory and Computation*, **2009**, 5(12):3195-3210.
- [384] Qi Y, Ingólfsson HI, Cheng X, Lee J, Marrink SJ, Im W. *Journal of chemical theory and computation*, **2015**, 11(9):4486-4494.
- [385] Liwo A, Baranowski M, Czaplewski C, Gołaś E, He Y, Jagieła D, Krupa P, Maciejczyk M, Makowski M, Mozolewska MA, Niadzvedtski A. *Journal of molecular modelling*, **2014**, 20(8):1-5.
- [386] Liwo A, Czaplewski C, Sieradzan AK, Lubecka EA, Lipska AG, Golon Ł, Karczyńska A, Krupa P, Mozolewska MA, Makowski M, Ganzynkiewicz R. *Progress in Molecular Biology and Translational Science*, **2020**, 170:73-122.
- [387] Liwo A, Czaplewski C, Pillardy J, Scheraga HA. *The Journal of Chemical Physics*, **2001**, 115(5):2323-47.
- [388] Liwo A, Sieradzan AK, Lipska AG, Czaplewski C, Joung I, Żmudzińska W, Hałabis A, Ołdziej S. *The Journal of Chemical Physics*, **2019**, 150(15):155104.
- [389] He Y, Maciejczyk M, Ołdziej S, Scheraga HA, Liwo A. *Physical review letters*, **2013**, 110(9):098101.
- [390] Lubecka EA, Liwo A. *The Journal of Chemical Physics*, **2017**, 147(11):115101.
- [391] Samsonov SA, Lubecka EA, Bojarski KK, Ganzynkiewicz R, Liwo A. *Biopolymers*, **2019**, 110(8):e23269.
- [392] Germer A, Peter MG, Kleinpeter E. *The Journal of Organic Chemistry*, **2002**, 67(18):6328-6338.
- [393] Wen X, Yuan Y, Kuntz DA, Rose DR, Pinto BM. *Biochemistry*, **2005**, 44(18):6729-6737.
- [394] Kawatkar SP, Kuntz DA, Woods RJ, Rose DR, Boons GJ. *Journal of the American Chemical Society*, **2006**, 128(25):8310-8319.
- [395] Pedatella S, De Nisco M, Ernst B, Guaragna A, Wagner B, Woods RJ, Palumbo G. *Carbohydrate research*, **2008**, 343(1):31-38.
- [396] Stortz CA. *Journal of computational chemistry*, **2005**, 26(5):471-483.
- [397] Stortz CA, Johnson GP, French AD, Csonka GI. *Carbohydrate research*, **2009**, 344(16):2217-2228.
- [398] Fadda E, Woods RJ. *Drug discovery today*, **2010**, 15(15-16):596-609.

- [399] Foley BL, Tessier MB, Woods RJ. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, **2012**, 2(4):652-697.
- [400] Taha HA, Richards MR, Lowary TL. *Chemical Reviews*, **2013**, 113(3):1851-1876.
- [401] Xiong X, Chen Z, Cossins BP, Xu Z, Shao Q, Ding K, Zhu W, Shi J. *Carbohydrate research*, **2015**, 401:73-81.
- [402] Guvench O, Greene SN, Kamath G, Brady JW, Venable RM, Pastor RW, Mackerell Jr AD. *Journal of computational chemistry*, **2008**, 29(15):2543-2564.
- [403] Guvench O, Hatcher E, Venable RM, Pastor RW, MacKerell Jr AD. *Journal of chemical theory and computation*, **2009**, 5(9):2353-2370.
- [404] Mallajosyula SS, Guvench O, Hatcher E, MacKerell Jr AD. *Journal of chemical theory and computation*, **2012**, 8(2):759-776.
- [405] Cloutier T, Sudrik C, Sathish HA, Trout BL. *The Journal of Physical Chemistry B*, **2018**, 122(40):9350-9360.
- [406] He X, Lopes PE, MacKerell Jr AD. *Biopolymers*, **2013**, 99(10):724-738.
- [407] Jana M, MacKerell Jr AD. *The Journal of Physical Chemistry B*, **2015**, 119(25):7846-7859.
- [408] Patel DS, He X, MacKerell Jr AD. *The Journal of Physical Chemistry B*, **2015**, 119(3):637-652.
- [409] Yang M, Aytenfisu AH, MacKerell Jr AD. *Carbohydrate research*, **2018**, 457:41-50.
- [410] Aytenfisu AH, Yang M, MacKerell Jr AD. *Journal of chemical theory and computation*, **2018**, 14(6):3132-3143.
- [411] Kognole AA, Lee J, Park SJ, Jo S, Chatterjee P, Lemkul JA, Huang J, MacKerell Jr AD. *Journal of computational chemistry*, **2022**, 43(5):359-375.
- [412] Singh A, Tessier MB, Pederson K, Wang X, Venot AP, Boons GJ, Prestegard JH, Woods RJ. *Canadian journal of chemistry*, **2016**, 94(11):927-935.
- [413] Kirschner KN, Yongye AB, Tschampel SM, González-Outeiriño J, Daniels CR, Foley BL, Woods RJ. *Journal of computational chemistry*, **2008**, 29(4):622-655.
- [414] DeMarco ML, Woods RJ. *Glycobiology*, **2009**, 19(4):344-355.
- [415] DeMarco ML. *In Glycoinformatics*, **2015** (pp. 379-390). Humana Press, New York, NY.
- [416] Kirschner KN, Lins RD, Maass A, Soares TA. *Journal of chemical theory and computation*, **2012**, 8(11):4719-4731.

- [417] Tessier MB, DeMarco ML, Yongye AB, Woods RJ. *Molecular Simulation*, **2008**, 34(4):349-364.
- [418] Pol-Fachin L, Rusu VH, Verli H, Lins RD. *Journal of chemical theory and computation*, **2012**, 8(11):4681-4690.
- [419] Lins RD, Hünenberger PH. *Journal of computational chemistry*, **2005**, 26(13):1400-1412.
- [420] Hansen HS, Hünenberger PH. *Journal of computational chemistry*, **2011**, 32(6):998-1032.
- [421] Plazinski W, Lonardi A, Hünenberger PH. *Journal of computational chemistry*, **2016**, 37(3):354-365.
- [422] Naumov VS, Ignatov SK. *Journal of Molecular Modeling*, **2017**, 23(8):1-5.
- [423] Nester K, Gaweda K, Plazinski W. *Journal of Chemical Theory and Computation*, **2019**, 15(2):1168-1186.
- [424] Pol-Fachin L, Fernandes CL, Verli H. *Carbohydrate research*, **2009**, 344(4):491-500.
- [425] Fernandes CL, Sachett LG, Pol-Fachin L, Verli H. *Carbohydrate research*, **2010**, 345(5):663-671.
- [426] Jorgensen WL, Maxwell DS, Tirado-Rives J. *Journal of the American Chemical Society*, **1996**, 118(45):11225-11236.
- [427] Damm W, Frontera A, Tirado-Rives J, Jorgensen WL. *Journal of computational chemistry*, **1997**, 18(16):1955-1970.
- [428] Jamali SH, Westen TV, Moulton OA, Vlugt TJ. *Journal of chemical theory and computation*, **2018**, 14(12):6690-6700.
- [429] Plazinska A, Plazinski W. *Journal of Chemical Theory and Computation*, **2021**, 17(4):2575-2585.
- [430] Lazar RD, Akher FB, Ravenscroft N, Kuttel MM. *Journal of Chemical Theory and Computation*, **2022**, 18(2):1156-1172.
- [431] Molinero V, Goddard WA. M3B: *The Journal of Physical Chemistry B*, **2004**, 108(4):1414-1427.
- [432] Liu P, Izvekov S, Voth GA. *The Journal of Physical Chemistry B*, **2007**, 111(39):11566-11575.
- [433] Yu Z, Lau D. *Journal of Molecular Modeling*, **2015**, 21(5):1-9.
- [434] Benner SW, Hall CK. *The Journal of Physical Chemistry B*, **2016**, 120(29):7253-7264.
- [435] Tsereteli L, Grafmüller A. *PLoS One*, **2017**, 12(7):e0180938.

- [436] Schmalhorst PS, Deluweit F, Scherrers R, Heisenberg CP, Sikora M. *Journal of Chemical Theory and Computation*, **2017**, 13(10):5039-5053.
- [437] Srinivas G, Cheng X, Smith JC. *Journal of chemical theory and computation*, **2011**, 7(8):2539-2548.
- [438] Srinivas G, Cheng X, Smith JC. *The Journal of Physical Chemistry B*, **2014**, 118(11):3026-3034.
- [439] Glass DC, Moritsugu K, Cheng X, Smith JC. *Biomacromolecules*, **2012**, 13(9):2634-2644.
- [440] Poma AB, Chwastyk M, Cieplak M. *The Journal of Physical Chemistry B*, **2015**, 119(36):12028-12041.
- [441] Nagarajan B, Holmes SG, Sankaranarayanan NV, Desai UR. *Current Opinion in Structural Biology*, **2022**, 74:102356.
- [442] Verli H, Guimarães JA. *Carbohydrate research*, **2004**, 339(2):281-290.
- [443] Uciechowska-Kaczmarzyk U, Babik S, Zsila F, Bojarski KK, Beke-Somfai T, Samsonov SA. *Journal of Molecular Graphics and Modelling*, **2018**, 82:157-166.
- [444] Potthoff J, Bojarski KK, Kohut G, Lipska AG, Liwo A, Kessler E, Ricard-Blum S, Samsonov SA. *International Journal of Molecular Sciences*, **2019**, 20(20):5021.
- [445] Sepuru KM, Nagarajan B, Desai UR, Rajarathnam K. *Journal of Biological Chemistry*, **2016**, 291(39):20539-20550.
- [446] Krieger E, Geretti E, Brandner B, Goger B, Wells TN, Kungl AJ. *Bioinformatics*, **2004**, 54(4):768-775.
- [447] Salbach-Hirsch J, Samsonov SA, Hintze V, Hofbauer C, Picke AK, Rauner M, Gehrcke JP, Moeller S, Schnabelrauch M, Scharnweber D, Pisabarro MT. *Biomaterials*, **2015**, 67:335-345.
- [448] Ruiz-Gómez G, Vogel S, Möller S, Pisabarro MT, Hempel U. *Scientific reports*, **2019**, 9(1):1-15.
- [449] Marcisz M, Huard B, Lipska AG, Samsonov SA. *Glycobiology*, **2021**, 31(7):772-86.
- [450] Marcisz M, Maszota-Zieleniak M, Huard B, Samsonov SA. *Biomolecules*, **2021**, 11(9):1349.
- [451] Nieto PM, Canales A, Fernández IS, Santillana E, González-Corrochano R, Redondo-Horcajo M, Cañada FJ, Nieto P, Martín-Lomas M, Gimenez-Gallego G, Jiménez-Barbero, J. *ChemBioChem*, **2013**, 14:1732–1744.

- [452] Muñoz-García JC, Chabrol E, Vivès RR, Thomas A, de Paz JL, Rojo J, Imberty A, Fieschi F, Nieto PM, Angulo J. *Journal of the American Chemical Society*, **2015**, 137(12):4100-4110.
- [453] Bojarski KK, Samsonov SA. *Journal of Chemical Information and Modeling*, **2021**, 61(1):455-466.
- [454] Gandhi NS, Mancera RL. *Journal of chemical information and modelling*, **2011**, 51(2):335-358.
- [455] Nordsieck K, Pichert A, Samsonov SA, Thomas L, Berger C, Pisabarro MT, Huster D, Beck-Sickinger AG. *ChemBioChem*, **2012**, 13(17):2558-2566.
- [456] Pichert A, Samsonov SA, Theisgen S, Thomas L, Baumann L, Schiller J, Beck-Sickinger AG, Huster D, Pisabarro MT. *Glycobiology*, **2012**, 22(1):134-145.
- [457] Möbius K, Nordsieck K, Pichert A, Samsonov SA, Thomas L, Schiller J, Kalkhof S, Pisabarro MT, Beck-Sickinger AG, Huster D. *Glycobiology*, **2013**, 23(11):1260-1269.
- [458] Singh A, Kett WC, Severin IC, Agyekum I, Duan J, Amster IJ, Proudfoot AE, Coombe DR, Woods RJ. *Journal of Biological Chemistry*, **2015** 290(25):15421-15436.
- [459] Künze G, Köhling S, Vogel A, Rademann J, Huster D. *Journal of Biological Chemistry*, **2016**, 291(6):3100-3113.
- [460] Penk A, Baumann L, Huster D, Samsonov SA. *Glycobiology*, **2019**, 29(10):715-725.
- [461] Sapay N, Cabannes E, Petitou M, Imberty A. *Glycobiology*, **2011**, 21(9):1181-1193.
- [462] Panitz N, Theisgen S, Samsonov SA, Gehrcke JP, Baumann L, Bellmann-Sickert K, Köhling S, Pisabarro MT, Rademann J, Huster D, Beck-Sickinger AG. *Glycobiology*, **2016**, 26(11):1209-1221.
- [463] Vallet SD, Miele AE, Uciechowska-Kaczmarzyk U, Liwo A, Duclos B, Samsonov SA, Ricard-Blum S. *Scientific reports*, **2018**, 8(1):1-6.
- [464] Lecaille F, Chazeirat T, Bojarski KK, Renault J, Saidi A, Prasad VG, Samsonov S, Lalmanach G. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, **2020**, 1868(2):140318.
- [465] Faller CE, Guvench O. *The Journal of Physical Chemistry B*, **2015**, 119(20):6063-6073.
- [466] Whitmore EK, Vesenka G, Sihler H, Guvench O. *Biomolecules*, **2020**, 10(4):537.
- [467] Whitmore EK, Martin D, Guvench O. *International journal of molecular sciences*, **2020**, 21(20):7699.
- [468] Guvench O, Whitmore EK. *ACS omega*, **2021**, 6(20):13204-13217.

- [469] Giubertoni G, Pérez de Alba Ortíz A, Bano F, Zhang X, Linhardt RJ, Green DE, DeAngelis PL, Koenderink GH, Richter RP, Ensing B, Bakker HJ. *Macromolecules*, **2021**, 54(3):1137-1146.
- [470] Samantray S, Olubiyi OO, Strodel B. *International journal of molecular sciences*, **2021**, 22(21):11529.
- [471] Joshi HV, Jedrzejak MJ, de Groot BL. *Proteins: Structure, Function, and Bioinformatics*, **2009**, 76(1):30-46.
- [472] Torrent M, Nogues MV, Andreu D, Boix E. *PLoS One*, **2012**; 7:e42692.
- [473] Sankaranarayanan, N.V.; Desai, U.R. *Glycobiology*, **2014**, 24:1323–1333.
- [474] Samsonov SA, Pisabarro MT. *Glycobiology*, **2016**, 26(8):850-861.
- [475] Li Z, Yasuda Y, Li W, Bogyo M, Katz N, Gordon RE, Fields GB, Brömme D. *Journal of Biological Chemistry*, **2004**, 279(7):5470-5479.
- [476] Bathe M, Rutledge GC, Grodzinsky AJ, Tidor B. *Biophysical journal*, **2005**, 88(6):3870-3887.
- [477] Sattelle BM, Shakeri J, Almond A. *Biomacromolecules*, **2013**, 14(4):1149-1159.
- [478] Sattelle BM, Shakeri J, Cliff MJ, Almond A. *Biomacromolecules*, **2015**, 16(3):951-961.
- [479] Samsonov SA, Bichmann L, Pisabarro MT. *Journal of Chemical Information and Modeling*, **2015**, 55(1):114-124.
- [480] Mazurek AH, Szeleszczuk Ł, Gubica T. *International journal of molecular sciences*, **2021**, 22(17):9422.
- [481] Cézard C, Trivelli X, Aubry F, Djedaïni-Pilard F, Dupradeau FY. *Physical Chemistry Chemical Physics*, **2011**, 13(33):15103-15121.
- [482] He J, Chipot C, Shao X, Cai W. *The Journal of Physical Chemistry C*, **2013**, 117(22):11750-11756.
- [483] Da Silva AJ, Dos Santos ES. *European Biophysics Journal*, **2020**, 49(7):571-589.
- [484] Hotarat W, Nutho B, Wolschann P, Rungrotmongkol T, Hannongbua S. *Molecules*. **2020**, 25(11):2532.
- [485] Zhao Q, Miriyala N, Su Y, Chen W, Gao X, Shao L, Yan R, Li H, Yao X, Cao D, Wang Y. *Molecular Pharmaceutics*, **2018**, 15(4):1664-1673.
- [486] Wadhwa R, Nigam N, Bhargava P, Dhanjal JK, Goyal S, Grover A, Sundar D, Ishida Y, Terao K, Kaul SC. *Journal of Cancer*, **2016**, 7(13):1755.
- [487] Mahalapbutr P, Wonganan P, Charoenwongpaiboon T, Prousoontorn M, Chavasiri W, Rungrotmongkol T. *Biomolecules*, **2019**, 9(10):545.

- [488] Sun Q, Tang P, Zhao L, Pu H, Zhai Y, Li H. *Carbohydrate polymers*, **2018**, 194:294-302.
- [489] Rescifina A, Chiacchio U, Iannazzo D, Piperno A, Romeo G. *Journal of agricultural and food chemistry*, **2010**, 58(22):11876-11882.
- [490] Kamigauchi M, Kawanishi K, Ohishi H, Ishida T. *Chemical and pharmaceutical bulletin*, **2007**, 55(5):729-733.
- [491] López CA, de Vries AH, Marrink SJ. *Scientific reports*, **2013**, 3(1):1-6.
- [492] Cieplak M, Thompson D. *The Journal of chemical physics*, **2008**, 128(23):234906.
- [493] Škvára J, Nezbeda I. *Journal of Molecular Liquids*, **2018**, 265:791-6.
- [494] Zwanzig RW. *The Journal of Chemical Physics*, **1954**, 22(8):1420-1426.
- [495] Yoo J, Cui Q. *Biophysical journal*, **2008**, 94(8):L61-L63.
- [496] Chandrasekhar J, Shariffskul S, Jorgensen WL. *The Journal of Physical Chemistry B*, **2002**, 106(33):8078-8085.
- [497] Li Z, Li X, Huang YY, Wu Y, Zhou L, Liu R, Wu D, Zhang L, Liu H, Xu X, Zhang Y, Cui J, Wang X, Luo HB. *BioRxiv*, **2020**, epub.
- [498] Sarkar A, Yu W, Desai UR, MacKerell AD, Mosier PD. *Glycobiology*, **2016**, 26(10):1041-1047.
- [499] Babik S, Samsonov SA, Pisabarro MT. *Glycoconjugate journal*, **2017**, 34(3):427-440.
- [500] Aqvist J, Marelius J. *Combinatorial chemistry & high throughput screening*, **2001**, 4(8):613-626.
- [501] Jones-Hertzog DK, Jorgensen WL. *Journal of medicinal chemistry*, **1997**, 40(10):1539-1549.
- [502] Lamb ML, Tirado-Rives J, Jorgensen WL. *Bioorganic & medicinal chemistry*, **1999**, 7(5):851-860.
- [503] Park H, Merz KM Jr. *Journal of the American Chemical Society*, **2003**, 125(4):901-911.
- [504] You W, Tang Z, Chang CA. *Journal of chemical theory and computation*, **2019**, 15(4):2433-2443.
- [505] Torrie GM; Valleau JP. *Journal of Computational Physics*, **1977**, 23(2):187-199.
- [506] Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA. *Journal of computational chemistry*, **1992**, 13(8):1011-1021.
- [507] Souaille M, Roux B. *Computer physics communications*, **2001**, 135(1):40-57.
- [508] Homeyer N, Gohlke H. *Molecular Informatics*, **2012**, 31(2):114-122.

- [509] Gandhi NS, Mancera RL. *Glycobiology*, **2009**, 19(10):1103-1115.
- [510] Miller III BR, McGee Jr TD, Swails JM, Homeyer N, Gohlke H, Roitberg AE. *Journal of chemical theory and computation*, **2012**, 8(9):3314-3321.
- [511] Srinivasan J, Cheatham TE, Cieplak P, Kollman PA, Case DA. *Journal of the American Chemical Society*, **1998**, 120(37):9401-9409.
- [512] Weis A, Katebzadeh K, Söderhjelm P, Nilsson I, Ryde U. *Journal of medicinal chemistry*, **2006**, 49(22):6596-6606.
- [513] Karplus M, Kushick JN. *Macromolecules*, **1981**, 14(2):325-332.
- [514] Samsonov SA, Pisabarro MT. *Carbohydrate Research*, **2013**, 381:133–137.
- [515] Zsila F, Samsonov SA. *Carbohydrate research*, **2018**, 462:28-33.
- [516] Zsila F, Samsonov SA, Maszota-Zieleniak, M. *The Journal of Physical Chemistry B*, **2020**, 124(51):11625-11633.
- [517] Núñez-Agüero CJ, Escobar-Llanos CM, Díaz D, Jaime C, Garduño-Juárez R. *Tetrahedron*, **2006**, 62(17), 4162-4172.
- [518] Beà I, Gotsev MG, Ivanov PM, Jaime C, Kollman PA. *The Journal of Organic Chemistry*, **2006**, 71(5):2056-2063.
- [519] Triwahyuningtyas D, Megantara S, Hong TT, Yusuf M, Muchtaridi M. *Journal of Advanced Pharmaceutical Technology & Research*, **2021**, 12(3):250.
- [520] Sankareswari VG, Vinod D, Mahalakshmi A, Alamelu M, Kumaresan G, Ramaraj R, Rajagopal S. *Dalton Transactions*, **2014**, 43(8):3260-3272.
- [521] Fan F, Zhao Y, Cao Z. *Physical Chemistry Chemical Physics*, **2019**, 21(14):7429-7439.
- [522] Niu X, Gao X, Wang H, Wang X, Wang S. *Journal of molecular modeling*, **2013**, 19(3):1039-1047.
- [523] Genheden S, Ryde U. *Expert opinion on drug discovery*, **2015**, 10(5):449-461.
- [524] Mishra SK, Koča J. *The Journal of Physical Chemistry B*, **2018**, 122(34):8113-8121.
- [525] Sangpheak W, Khuntawee W, Wolschann P, Pongsawasdi P, Rungrotmongkol T. *Journal of Molecular Graphics and Modelling*, **2014**, 50:10–15.
- [526] Sangpheak W, Kicuntod J, Schuster R, Rungrotmongkol T, Wolschann P, Kungwan N, Viernstein H, Mueller M, Pongsawasdi P. *Beilstein journal of organic chemistry*, **2015**, 11(1):2763–2773.
- [527] Hanpaibool C, Chakcharoensap T, Hijikata Y, Irle S, Wolschann P, Kungwan N, Pongsawasdi P, Ounjai P, Rungrotmongkol T. *Journal of Molecular Liquids*, **2018**, 265:16–23.

- [528] Norrby PO, Liljefors T. *Journal of computational chemistry*, **1998**, 19(10):1146-66.
- [529] Allinger NL, Geise HJ, Pyckhout W, Paquette LA, Gallucci JC. *Journal of the American Chemical Society*, **1989**, 111(3):1106-1114.
- [530] Casu B, Petitou M, Provasoli M, Sinaÿ P. *Trends in Biochemical Sciences*, **1988**, 13:221–225.
- [531] Imberty A, Lortat-Jacob H, Perez S. *Carbohydrate Research*, **2007**, 342:430–439.
- [532] Mulloy, B. *Anais da Academia Brasileira de Ciencias*, **2005**, 77(4):651-664.
- [533] Kaliannan P, Gromiha MM, Ramamurthi K, Elanthiraiyan M. *Biophysical chemistry*, **1998**, 74(1):11-22.
- [534] Mosier PD, Krishnasamy C, Kellogg GE, Desai UR. *PLoS One*, **2012**, 7(11):e48632.
- [535] Habuchi H, Habuchi O, Kimata K. *Glycoconjugate Journal*, **2004**, 21:47–52.
- [536] Gama CI, Tully SE, Sotogaku N, Clark PM, Rawat M, Vaidehi N, Goddard III WA, Nishi A, Hsieh-Wilson LC. *Nature chemical biology*, **2006**, 2(9):467-473.
- [537] Atkovska K, Samsonov SA, Paszkowski-Rogacz M, Pisabarro MT. *International journal of molecular sciences*, **2014**, 15:2622–2645.
- [538] Joseph PRB, Mosier PD, Desai UR, Rajarathnam K. *Biochemical Journal*, **2015**, 472:121–133.
- [539] Bojarski KK, Samsonov SA. *Journal of Chemical Information and Modeling*, **2020**, 61(1):455-466.
- [540] Paiardi G, Milanese M, Wade RC, D’Ursi P, Rusnati M. *Biomolecules*, **2021**, 11(5):739.
- [541] Megyes T, Grósz T, Radnai T, Bakó I, Pálinkás G. *The Journal of Physical Chemistry A*, **2004**, 108(35):7261–7271.
- [542] Rahnamoun A, O’Hearn KA, Kaymak MC, Li Z, Merz Jr KM, Aktulga HM. *The Journal of Physical Chemistry Letters*, **2022**, 13:5334-5340
- [543] Åqvist, J. *The Journal of Physical Chemistry*, **1990**, 94(21):8021–8024.
- [544] Duarte F, Bauer P, Barrozo A, Amrein BA, Purg M, Åqvist J, Kamerlin SC. *The Journal of Physical Chemistry B*, **2014**, 118(16):4351–4362.
- [545] Lu SY, Huang ZM, Huang WK, Liu XY, Chen YY, Shi T, Zhang J. *Proteins: Structure, Function, and Bioinformatics*, **2013**, 81(5):740–753.
- [546] Saxena A, Sept D. *Journal of chemical theory and computation*, **2013**, 9(8):3538–3542.
- [547] Li P, Song LF, Merz Jr KM. *Journal of chemical theory and computation*, **2015**, 11(4):1645-1657.

- [548] Raynal P, Pollard HB. *Biochimica et Biophysica Acta (BBA)-Reviews on Biomembranes*, **1994**, 1197(1):63-93.
- [549] Swairjo MA, Seaton BA. *Annual review of biophysics and biomolecular structure*, **1994**, 23(1):193-213.
- [550] Gerke V, Moss SE. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, **1997**, 1357(2):129-154.
- [551] Seaton BA, Dedman JR. *Biometals*, **1998**, 11(4):399-404.
- [552] Gerke V, Creutz CE, Moss SE. *Nature reviews Molecular cell biology*, **2005**, 6(6):449-461.
- [553] Waisman DM. *Signal Transduction Mechanisms*, **1995**:301-322.
- [554] Kojima K, Yamamoto K, Irimura T, Osawa T, Ogawa H, Matsumoto I. *Journal of Biological Chemistry*, **1996**, 271(13):7679-7685.
- [555] Kassam G, Manro A, Braat CE, Louie P, Fitzpatrick SL, Waisman DM. *Journal of Biological Chemistry*, **1997**, 272(24):15093-15100.
- [556] Ishitsuka R, Kojima K, Utsumi H, Ogawa H, Matsumoto I. *Journal of Biological Chemistry*, **1998**, 273(16):9935-9941.
- [557] Fitzpatrick SL, Kassam G, Manro A, Braat CE, Louie P, Waisman DM. *Biochemistry*, **2000**, 39(9):2140-2148.
- [558] Tedesco D, Bertucci C. *Journal of pharmaceutical and biomedical analysis*, **2015**, 113:34-42.
- [559] Holmgaard List N, Knoops J, Rubio-Magnieto J, Ide J, Beljonne D, Norman P, Surin M, Linares M. *Journal of the American Chemical Society*, **2017**, 139(42):14947-14953.
- [560] Tramarin A, Tedesco D, Naldi M, Baldassarre M, Bertucci C, Bartolini M. *Journal of Pharmaceutical and Biomedical Analysis*, **2019**, 162:171-178.
- [561] de Carvalho Bertozo L, Philot EA, Lima AN, de Resende Lara PT, Scott AL, Ximenes VF. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, **2019**, 208: 243-254.
- [562] Losytskyy M, Chornenka N, Vakarov S, Meier-Menches SM, Gerner C, Potocki S, Arion VB, Gumienna-Kontecka E, Voloshin Y, Kovalska V. *Biomolecules*, **2020**, 10(12):1602.
- [563] Allenmark S. *Chirality: The Pharmacological, Biological, and Chemical Consequences of Molecular Asymmetry*, **2003**, 15(5):409-422.

7. Publications included in the PhD Thesis

- [D1] **Kogut MM**, Maszota-Zieleniak M, Marcisz M, Samsonov S. *Physical Chemistry Chemical Physics*. **2021**, 23(5).
- [D2] **Kogut MM**, Danielsson A, Ricard-Blum S, Samsonov SA. *Computational Biology and Chemistry*. **2022**, 99:107727.
- [D3] **Kogut MM**, Grabowska O, Wyrzykowski D, Samsonov SA. *Journal of Molecular Liquids*. **2022**, 364:119978.
- [D4] Tesmar A, **Kogut MM**, Żamojć K, Grabowska O, Chmur K, Samsonov SA, Makowska J, Wyrzykowski D, Chmurzyński L. *Journal of Molecular Liquids*. **2021**, 340:117185.
- [D5] Grabowska O, **Kogut MM**, Żamojć K, Samsonov SA, Makowska J, Tesmar A, Chmur K, Wyrzykowski D, Chmurzyński L. *Molecules*. **2021**, 26(21):6565.
- [D6] de Carvalho Bertozo L, **Kogut MM**, Maszota-Zieleniak M, Samsonov SA, Ximenes VF. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*. **2022**, 278:121374.
- [D7] **Kogut MM**, Marcisz M, Samsonov SA. *Current Opinion in Structural Biology*. **2022**, 73:102.

Other publications prepared during the PhD course

- Bojko M, Węgrzyn K, Sikorska E, Kocikowski M, Parys M, Battin C, Steinberger P, **Kogut MM**, Winnicki M, Sieradzan AK, Spodzieja M, Rodziewicz-Motowidło S. *Bioorganic Chemistry*. **2022**, 128:106047.
- Danielsson A, **Kogut MM**, Maszota-Zieleniak M, Chopra P, Boons GJ, Samsonov SA. *Computational Biology and Chemistry*. **2022**, 99:107716.
- Antoniak A, Biskupek I, Bojarski KK, Czaplewski C, Giełdoń A, Kogut M, **Kogut MM**, Krupa P, Lipska AG, Liwo A, Lubecka EA. *Journal of Molecular Graphics and Modelling*. **2021**, 108:108008.
- Maszota-Zieleniak M, Marcisz M, **Kogut MM**, Siebenmorgen T, Zacharias M, Samsonov SA. *Journal of Computational Chemistry*. **2021**, 42(15):1040-1053.

APPENDIX

PUBLICATION D1

**COMPUTATIONAL INSIGHTS INTO THE CALCIUM IONS
ROLE IN PROTEIN-GLYCOSAMINOGLYCAN SYSTEMS**



Cite this: *Phys. Chem. Chem. Phys.*,
2021, **23**, 3519

Computational insights into the role of calcium ions in protein–glycosaminoglycan systems†

Małgorzata M. Kogut, Martyna Maszota-Zieleniak, Mateusz Marcisz and
Sergey A. Samsonov *

Glycosaminoglycans (GAGs) are anionic, periodic, linear polysaccharides which are composed of periodic disaccharide units. They play a vital role in many biological processes ongoing in the extracellular matrix. In terms of computational approaches, GAGs are very challenging molecules due to their high flexibility, periodicity, predominantly electrostatic-driven nature of interactions with their protein counterparts and potential multipose binding. Furthermore, the molecular mechanisms underlying GAG-mediated interactions are not fully known yet, and experimental techniques alone are not always sufficient to gain insights into them. The aim of this study was to characterize protein–ion–GAG complexes for the systems where ions are directly involved in GAG binding. Molecular docking, molecular dynamics and free energy calculation approaches were applied to model and rigorously analyse the interactions between annexins (II and V), calcium ions (Ca^{2+}) and heparin (HP). The computational data were examined and discussed in the context of the structural data previously reported by the crystallographic studies. The computational results confirm that the presence of Ca^{2+} has a tremendous impact on the annexin–HP binding site. This study provides a general computational pipeline to discover the complexity of protein–GAG interactions and helps to understand the role of ions involved at the atomic level. The limitations of the applied protocols are described and discussed pointing at the challenges persisting in the state-of-the-art *in silico* tools to study protein–ion–GAG systems.

Received 16th October 2020,
Accepted 2nd January 2021

DOI: 10.1039/d0cp05438k

rsc.li/pccp

Introduction

Glycosaminoglycans (GAGs) represent a class of negatively charged linear polysaccharides composed of disaccharide units containing an amino sugar and uronic acid.¹ Depending on their glycosidic linkages and composition, GAGs are categorized into a few types such as heparin, heparan sulfate, dermatan sulfate, chondroitin sulfate, and keratan sulfate. Additionally, GAGs have different sulfation patterns, called the sulfation code,² which contribute to their structural variety, activity and recognition.³ Heparin (HP) consists of 2-*O*-sulfated iduronic acid and 6-*O*-sulfated, N-sulfated glucosamine with a -4 net charge per disaccharide unit. Due to its negative charge, most of the interactions are electrostatically-driven.⁴ Annexins are a family of homologous proteins that are widely distributed and ubiquitous in eukaryotes; they are associated with a wide range of physiological processes, including cell signalling, inflammation, apoptosis, homeostasis, membrane trafficking and ion transport.^{5–10} Most functional properties of annexins are attributed to their distinctive ability to bind to membrane phospholipids and

other anionic polymers, including GAGs in a Ca^{2+} -dependent manner.¹¹ The HP-binding properties of the annexin II tetramer have been characterized and a potential HP-binding site has been identified.¹² The binding of GAGs, including HP, to annexins IV, V and VI has been reported and described in several *in vitro* studies.^{13–17} It was confirmed that annexins express different GAG preferences and that GAG binding may be Ca^{2+} -dependent. At present, there are three available crystal structures of both annexins V and II with HP including Ca^{2+} , with characterized HP-binding pockets (PDB ID: 1G5N, 2HYU and 2HYV) that were chosen for this study.^{12,18} Current experimental techniques are insufficient to gain full insights into protein–GAG interactions that take place at single-molecule levels.¹⁹ Therefore, the theoretical approaches are not only complementary but can lead to a better understanding of the role of individual interaction partners (including GAGs and potentially ions) by bringing often new and experimentally inaccessible details. Formerly, methodologies like molecular docking and molecular dynamics proved to be successful in modelling protein–GAG interactions. In particular, the effects of GAG binding on chemokines,^{20,21} growth factors^{22,23} and other proteins^{24,25} were investigated. These studies examined the fundamental questions related to these interactions such as their specificity, the role of the multipose character of GAG binding and the polarity of the binding poses of these periodic molecules.^{24,26} Nevertheless, modelling interactions

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

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d0cp05438k

in the corresponding complexes still represent substantial challenges. One of the reasons is the highly charged nature of GAGs, which leads to the importance of electrostatic-driven interactions including abundant solvent-mediated ones²⁵ and ion-dependent both requiring appropriate and careful computational treatment. Binding ions to GAGs have been experimentally characterized for different types of ions such as Zn²⁺, Mn²⁺, Cu²⁺, Ca²⁺, Co²⁺, Na⁺, Mg²⁺, Fe³⁺, Ni²⁺, Al³⁺, and Sr²⁺.^{27,28} The crucial role of ions for protein–GAG interactions was shown experimentally for APP,²⁹ HP cofactor II,³⁰ endostatin,³¹ FGF1 and IL-7.³² However, neither a conformational basis of GAG-ion interactions nor the molecular mechanisms for protein–ion–GAG systems were proposed to explain the experimental findings at the atomic level. In a recent study, it was demonstrated that the Zn²⁺ attenuation of the binding affinity in a particular system containing endostatin and HP is electrostatic-driven.³³ The goal of this study was to get a deeper insight into protein–ion–GAG interactions using *in silico* techniques and to verify the accuracy and sensitivity of most commonly deployed molecular modelling tools in the investigated annexin–Ca²⁺–HP complexes. The calculation of the electrostatic potential map for annexins II and V provided information on the GAG-binding site in the presence of Ca²⁺ and how it can be changed upon Ca²⁺ removal. Using the molecular docking approach, various structures were calculated of annexin–Ca²⁺–GAG complexes (Table 1) addressing the aspect of the putative role of Ca²⁺ in these interactions and in the establishment of the HP-binding site. All-atom molecular dynamics simulations were conducted to study the dynamics of the above structures and complemented by free energy analysis including linear interaction energy (LIE) and molecular mechanics generalized Born surface area (MM-GBSA) approaches to characterize the stability of these complexes in time. Additionally, it was investigated how the potential mean force (PMF) is affected by the presence and absence of Ca²⁺ using the umbrella sampling (US) method. The results reported here contribute to a better understanding of how computational tactics might help to inspect the biologically relevant annexin–GAG interactions in the presence and absence of Ca²⁺ at the atomic level.

Methods

Structures

Protein structures. The structures of annexin II and annexin V were obtained from the Protein Data Bank (PDB ID: 2HYV,

1.42 Å; 2HYU, 1.86 Å; 1G5N, 1.90 Å). Based on these structures, the following systems with various numbers of Ca²⁺ were prepared as presented in Table 1.

GAG structures. Heparin (HP): dp4, dp5 and dp24 (dp stands for the degree of polymerization) was built from the building blocks of the sulfated GAG monomeric unit libraries.²⁰ Their charges were obtained from the GLYCAM06 force field³⁴ and from the literature for sulfate groups.³⁵

Electrostatic potential calculations

In order to calculate the electrostatic potential isosurfaces for monomers of annexins II and V (in the presence and absence of Ca²⁺), a Poisson–Boltzmann surface area (PBSA) program from AmberTools³⁶ was used with a 1 Å grid spacing step. The results of PBSA analysis have shown how the presence of Ca²⁺ may influence the GAG binding region. Additionally, the electrostatic potential was calculated for annexin V dimer to investigate how the neighbouring proteins available in the crystal may dictate the preferential GAG-binding site. The obtained electrostatic potential maps were visualized with VMD software.³⁷

Molecular docking

Autodock3 software³⁸ was used to perform molecular docking. Among other electrostatic potential-based calculation protocols,^{39,40} this approach was reported to be the most successful in predicting the GAG-binding regions on the protein surface.^{23,41}

Maximum gridbox size was used (126 Å × 126 Å × 126 Å) that covered half of the protein, including the known GAG-binding site with the default grid step of 0.375 Å. The Lamarckian genetic algorithm was deployed for 1000 independent runs. The size of 300 for the initial population and 10⁵ generations for termination conditions were chosen and 9995 × 10⁵ energy evaluations were performed. Clustering was performed using the DBSCAN algorithm⁴² on the top 50 docking poses. Two to three clusters of each GAG's docking solutions were chosen for further analysis. Each time clustering parameters were chosen individually (epsilon in the range of 1–4, minpoints ranging between 2 and 5). In total, 16 different systems have been analysed: 8 for 1G5N, 4 for 2HYU and 4 for 2HYV (Table 1).

RS-REMD

Alternatively to Autodock3, a new approach of replica-exchange MD with repulsive scaling was implemented in this study. The protocol used in the original work of Siebenmorgen *et al.*⁴³ was applied to the protein–GAG complexes. The ff14SBonlysc force field parameters for protein⁴⁴ and the GLYCAM06³⁴ for GAGs were used. The oligosaccharide ligand was placed at the opposite side of the protein with respect to the experimentally known binding site. MD simulations were performed in an implicit solvent with the model igb = 8⁴⁵ with an infinite cut-off for non-bonded interactions. The minimization was performed with 3000 steps of steepest descent and 3000 steps of conjugate gradient. This was followed by heating to 300 K for 10 ns with a Langevin thermostat ($\gamma = 5 \text{ ps}^{-1}$). The harmonic restraints of 0.05 kcal mol⁻¹ Å⁻² were applied on all the heavy atoms of the

Table 1 In each complex, various numbers of Ca²⁺ were removed, depending on how many Ca²⁺ ions were in the binding site

Name	Number of Ca ²⁺ ions removed (numbering as in PDB)		
	1G5N	2HYU	2HYV
All ions	None	None	None
Comb1	401	507	607
Comb2	402	508	608
Comb3	403	N/A	N/A
Comb4	401 & 402	N/A	N/A
Comb5	401 & 403	N/A	N/A
Comb6	402 & 403	N/A	N/A
No ions	All	All	All

protein and Ca^{2+} in the production run. In addition, in order to avoid ligand dissociation too far away from the receptor, a positional restraint of $1.0 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ between the centres of masses (COMs) of the receptor and the ligand was applied. The distance between COMs was calculated as a sum of the maximum distances between the COMs and the atoms of the surface of the receptor and ligand, respectively, increased by 10 \AA , to allow the free movement of the ligand on the complete surface of the protein. The ligand internal degrees of freedom were completely unrestrained during the production run. For each system, 16 replicas were used with different Lennard-Jones (LJ) parameters for atomic pairs from both the receptor and the ligand molecules. As in the original work of Siebenmorgen *et al.*,⁴³ the parameter d adjusting the effective van der Waals radius and the factor e changing the LJ potential well depths were assigned to 0.00 \AA , 0.01 \AA , 0.02 \AA , 0.04 \AA , 0.08 \AA , 0.12 \AA , 0.16 \AA , 0.20 \AA , 0.24 \AA , 0.28 \AA , 0.32 \AA , 0.38 \AA , 0.44 \AA , 0.50 \AA , 0.58 \AA and 0.68 \AA , and 0.000 , 0.015 , 0.030 , 0.0045 , 0.060 , 0.075 , 0.090 and 0.120 , respectively. In the production run, 50 000 MD exchange steps between the neighbouring replicas were attempted yielding 500 ns per replica in total.

Molecular dynamics simulations

Every all-atom molecular dynamics (MD) simulation of different annexin II and V combination complexes obtained from molecular docking was performed using the AMBER16 software package.³⁶ A truncated octahedron TIP3P periodic box of 8 \AA water layer from the box's border to the solute was used to solvate the complexes. The charge was neutralized with either Na^+ or Cl^- counterions. Energy minimization was carried out in two steps: beginning with 500 steepest descent cycles and 10^3 conjugate gradient cycles with $100 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ harmonic force restraints, continued with 3×10^3 steepest descent cycles and 3×10^3 conjugate gradient cycles without any restraints. Following the minimization steps, the system was heated up from 0 to 300 K for 10 ps with harmonic force restraints of $100 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$. Then, the system was equilibrated at 300 K and 10^5 Pa in an isothermal isobaric ensemble for 500 ps. Afterwards, the prediction MD run was carried out in the same isothermal isobaric ensemble, using the ff14SBonlysc force fields⁴⁶ for 20 ns. The particle mesh Ewald method for treating electrostatics and the SHAKE algorithm for all the covalent bonds containing hydrogen atoms were implemented in the MD simulations.

Binding free energy calculations

Energetic postprocessing of the trajectories and per-residue energy decomposition were performed for all the systems using linear interaction energy (LIE) with a dielectric constant of 80 and molecular mechanics generalized Born surface area (MM-GBSA)⁴⁷ using a model with surface area and Born radii as default parameters as implemented in the $\text{igb} = 2$ model in AMBER16.⁴⁶ All the frames from MD simulations were analysed.

The selectivity of the binding poses was estimated using the approach proposed by Siebenmorgen *et al.*⁴⁹ First, all the scores for the same complex are normalized as $S'_i = (S_i - \langle S \rangle) / S_{\min}$, where S_{\min} and $\langle S \rangle$ are the minimum and mean energy values

from all these scores, respectively. Then, a decoy of “correct binding poses” is defined if the RMSatd of a pose is within a cut-off of $\text{RMSatd}_{\min} + 1.5 \text{ \AA}$, where RMSatd_{\min} is the minimum RMSatd value. The rest of the poses are described as “incorrect binding poses”. The poses with the highest scores among “correct binding poses” and “incorrect binding poses” are denoted as true (T) and false (F) poses, respectively. Then, the selectivity is defined as $S'_T - S'_F$ and ranges from -1 to 1 . Values of 1 and -1 represent the maximum selectivity in the binding and outside the binding site, respectively, while a value close to 0 means the absence of the selectivity.

Free energy perturbation

Free energy perturbation (FEP) calculations were performed by the double decoupling method as described by Hamelberg *et al.*⁵⁰ Briefly, the perturbation was carried out in two steps: (1) the electrostatic interactions are gradually turned off by changing the atomic charges to zero; (2) the van der Waals radii of the atoms are decreased to zero. The thermodynamic integration approach is applied to calculate the free energy difference between two states. The coupling parameter is changed from 0 to 1 and backward with the step of 0.01 . For each value of the coupling parameter, 10 ps of equilibration and 10 ps of prediction MD run were performed. Another component of the free energy, ΔG_{constr} , corresponds to the entropic penalty due to the ion positional constraints in the site calculated based on the fluctuations of the ion in a corresponding unrestrained simulation. The free energy of the transfer from bulk solvent to vacuo ($\Delta G_{\text{bulk-vacuo}}$) was calculated to complete the thermodynamic cycle.

Umbrella sampling and the potential mean force

Umbrella sampling (US) was performed to compute the free energy along the dissociation pathway in AMBER.⁴⁸ In US, a series of windows were evenly located along the reaction coordinates, and conformational sampling in these windows was achieved by enforcing an external biasing potential. The samplings in each window overlapped with those in the adjacent windows so that the unbiased potential mean force (PMF) was reproduced by removing the biasing potential. Once all the US MD simulations were finished, the data collected from the separate simulation windows were combined along the reaction coordinates. Subsequently, these data were used to calculate the PMF profile for the whole structural transition process by the weighted histogram analysis method (WHAM)⁵¹ using the code developed by Alan Grossfield.⁵² In total, 8 independent US simulations were performed for two complexes of annexins II and V (PDB ID: 1G5N & 2HYV) in the presence and absence of Ca^{2+} and using two different force constants (Table S1, ESI†). In each case, the initial distance between the centres of masses of the protein and the GAG was 6.5 \AA , and the final value was 60.5 \AA with the step of 1 \AA , producing 55 windows in total (in each window, the MD simulation was carried out for 10 ns). In the WHAM, the tolerance of iteration was set to 0.001 and the temperature to 300 K .

Data analysis and its graphical representation were done with the R-package,⁵³ PyMOL⁵⁴ and VMD.³⁷

Results and discussion

HP-binding regions

The binding regions for the HP ligand on the surfaces of annexins II and V are known and well-characterized by multiple *in vitro* experiments and their crystallographic structures are available.^{12,18} Nevertheless, both the experimental studies only consider the scenarios where Ca^{2+} ions are present. Therefore, the purpose of these calculations was not only to verify those regions using a computational approach but also to examine how the electrostatic potential of both proteins will change upon Ca^{2+} removal and, in consequence, to what extent it might influence the annexin-HP interactions. The PBSA program from the AmberTools package was used to calculate the electrostatic potential isosurfaces corresponding to the protein (Fig. 1). This approach was reported to be successful in predicting the GAG-binding regions on the protein surface.⁴¹ The obtained results revealed that in both cases, the absence of Ca^{2+} notably diminishes the positive regions on the protein's surface and, in consequence, could lead to the reduced strength of protein-GAG interactions. It should be noted that the crystallographic study on annexin V was conducted in the presence of other protein monomers within the complex crystal, whereas in this study, the single monomer was taken to examine the interactions with HP in a 1:1 relation. Therefore, when looking at 1G5N in Fig. 1B, it is clearly visible that the potential HP-binding region is extended over a much greater area. In contrast, in the case of annexin II (2HYV in Fig. 1B), the binding region reported in the crystallographic studies appears to be the most favourable one, due to the most positive potential conserved in that area. Additionally, a closer look was taken at

the role of particular Ca^{2+} in cases where the ligand is present and absent. For annexin V, all three ions participating in the HP binding interface were also present in the protein unbound structure (PDB ID: 1A8B, 1.9 Å). In contrast, for annexin II, the Ca^{2+} observed close to the HP molecule in the protein-HP complexes (number 508 in 2HYU and number 608 in 2HYV) were not present in the protein unbound structure (PDB ID: 2HYW, 2.1 Å), suggesting that the binding of this Ca^{2+} is favourable only in the presence of an HP molecule. Once the Ca^{2+} was removed from the structures of both the annexin complexes, the ligands were significantly more mobile than in the presence of Ca^{2+} (Table S2, ESI[†]). Similarly, the Ca^{2+} fluctuations, which participate in the HP binding, are always lower when the ligand is bound (Table S3, ESI[†]). The Ca-508 in the annexin II system unbinds when HP is not present, which agrees with the X-ray data.

Predicting annexin V/II-GAG complex structures

Molecular docking was performed to investigate whether AutoDock3 is able to recognize the binding sites for both the proteins that were predicted by the crystallographic studies. Moreover, the novelty of this study was to determine whether the removal of the chosen Ca^{2+} can influence the protein-GAG interactions and whether Autodock3 is a sensitive-enough technique to probe for that. Previously in our studies, this software proved to be the most effective docking tool for GAGs.^{23,41} The representative structures of annexin II/V-HP complexes have been produced and they were used for further molecular dynamics (MD) and free energy analysis. In the case of annexin V with all Ca^{2+} present, a greater distribution of clusters along the potential binding site could be observed (Fig. 2 upper left), whereas for annexin II, both the clusters were in close proximity to the binding site. These results are in agreement with the findings from section "HP-binding regions," suggesting increased mobility and flexibility of HP in the interactions with annexin V due to the lack of steric hindrance of the neighbouring protein's monomers in the corresponding crystal. Furthermore, the data provided by

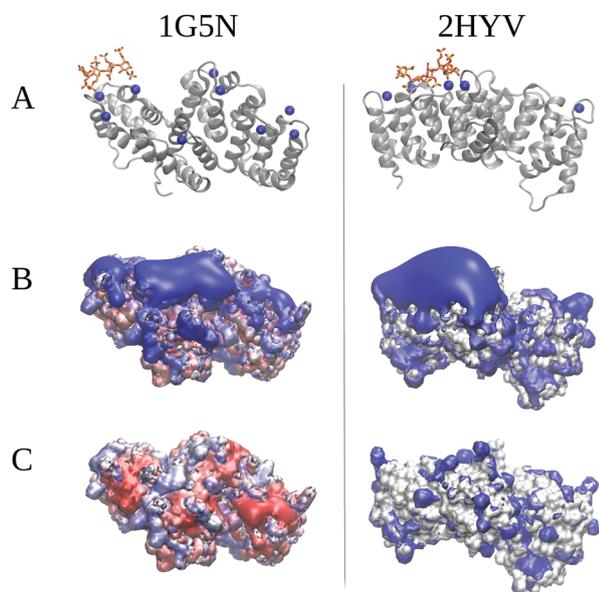


Fig. 1 (A) X-ray structures of annexins V and II (PDB ID: 1G5N and 2HYV, respectively), grey cartoon: protein, licorice: HP dp4, VDW spheres: calcium ions. (B) Electrostatic potential isosurfaces for annexins in the presence of calcium ions in surface representation (red, $-4 \text{ kcal mol}^{-1} \text{ e}^{-1}$; blue, $+4 \text{ kcal mol}^{-1} \text{ e}^{-1}$, respectively) and (C) in the absence of Ca^{2+} .

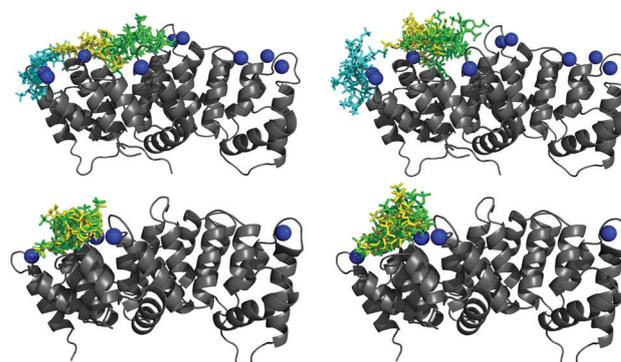


Fig. 2 "All ions": (1G5N) annexin V, 3 clusters (cyan, yellow and green), 4 structures in each cluster, after docking (top left) and after 20 ns MD (top right); (2HYU) annexin II, 2 clusters (yellow and green), 5 structures in each cluster, after AD3 and after MD. Clusters are in close proximity to the BS, and the results are in agreement with the PBSA predictions.

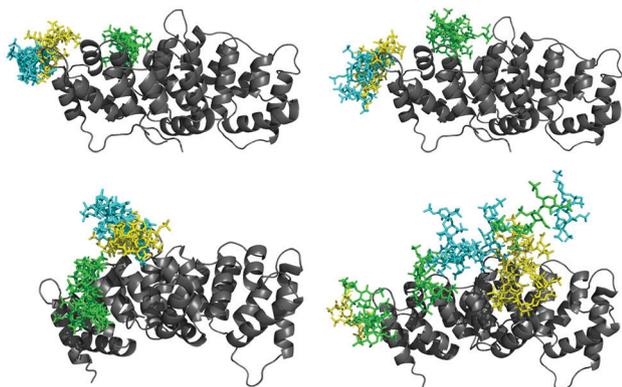


Fig. 3 “No ions”: 1G5N annexin V, three clusters, 4 structures in each cluster, after docking (top left) and after 10 ns MD (top right); 2HYU annexin II, three clusters (cyan, yellow and green), 5 structures in each cluster.

Autodock3 suggest that Autodock3 might not be the best tool for answering the question of how single Ca^{2+} influences the binding. Accordingly, the 1G5N complex has already been recognized to be computationally challenging in the previous studies.²³ Nevertheless, the differences between the “All ions” and the “No ions” scenarios appear obvious (Fig. 2 and 3). Autodock3 has found the clusters that are semi-close to the binding site; however, after 20 ns of MD simulations, several structures move further away from their initial location and spread over the protein’s surface, particularly in the case of annexin II. This indicates that Ca^{2+} ions do influence GAG’s behaviour and they are crucial to maintaining the HP molecule closely to the binding site.

Furthermore, similar to the study of Lie *et al.*,⁵⁵ in which molecular docking was performed with water molecules covalently attached to a ligand, here it was attempted to dock HP with covalently attached Ca^{2+} to annexin V. The Ca^{2+} were covalently attached to the HP’s oxygen atoms and the distance between them was chosen based on the MD simulation when the complex and these particular atoms interacting non-covalently with each other were stable. The goal was to investigate whether Ca^{2+} ions enhance the performance of Autodock3 in terms of predicting the annexin V-HP dp4 binding site. Unfortunately, this approach did not improve the performance of Autodock3. For six complexes of annexin II/V-HP, the RS-REMD (repulsive scaling replica-exchange molecular dynamics) method was used as an alternative to Autodock3 to identify the native complex geometry after starting from the distant initial location of HP. In the starting arrangement, HP was located on the opposite side of the receptor (Fig. 4 and Fig. S1, ESI[†]). In each case, 16 RS-REMD replicas with a different set of van der Waals parameters were performed using an implicit generalized Born (GB) solvent model. The results have shown that in the case of annexins V and II in the presence of Ca^{2+} , HP was able to find the binding site (Fig. 4 and Fig. S1, ESI[†]) unlike in the scenario where all the Ca^{2+} have been removed, HP was found to be too far away from the binding site. Moreover, annexin V was simulated with HP dp24 in the presence and

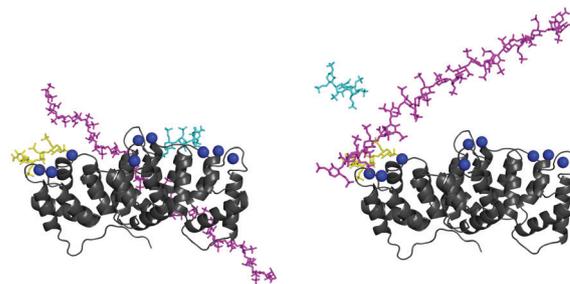


Fig. 4 1G5N: initial (on the left) and final (on the right) positions of GAGs. In yellow: X-ray structure of HP dp4, cyan: HP dp4 and magenta: HP dp24.

absence of Ca^{2+} . The aim of this experiment was to examine whether the longer GAG would approach more favourably the binding site or cover the entire upper surface of the protein. The latter turned out to be true, confirming the hypothesis that the binding site found in the crystallographic site could have been enforced by the neighbouring monomers, restricting the HP molecule from binding to other sites. When applied to annexin-HP systems, RS-REMD proved to be an efficient tool to complement other molecular docking schemes, also in terms of its sensitivity towards the presence of divalent cations in the system. Its potential advantages over conventional docking approaches such as the AD3 for protein–GAG complexes, in general, include allowing for full flexibility of the receptor side chains and ligand; direct access to force field-based scoring schemes; and independence of the quality of the results of the GAG’s length that could assist both in the search of GAG minimal binding sequences and in the characterization of particular GAG binding subsites on the protein surface. In terms of computational expenses, this new approach requires the use of several GPUs of the same type: nevertheless, the total wall-time of the simulation (typically several hours) is essentially less than the time demanded by AD3 to complete 1000 runs with the tuned parameters as it is used in our protocols (up to 12–16 hours for GAG dp6, for example, in the case of 100 parallel processes).

MD-based free energy analysis of annexin–GAG binding in the presence and absence of Ca^{2+}

From the docking solutions described in section “Predicting annexin V/II-GAG complex structures,” four to five structures from each cluster were picked for MD simulations and for subsequent free energy analysis. Binding free energy (ΔG) computation can play an important role in prioritizing compounds to be evaluated experimentally on their affinity for target proteins. In this study, two popular end-point methods were used: linear interaction energy (LIE) and molecular mechanics generalized Born surface area (MM-GBSA). The more negative the ΔG value, the more stable the complex. Based on the crystallographic studies, the description of single ions and their importance in annexin–GAG binding, it was hypothesised that the most stable complex was expected when all ions were present.

Annexin V-HP dp4. The data indicate that LIE provides inconclusive results on whether the presence of ions contributes

to the stability of annexin V-HP dp4 complexes, since there are no significant differences in the obtained energies between combinations (Fig. 5: top). On the contrary, it could be argued that MM-GBSA displays a significant difference between the “All ions” and the “No ions” scenarios with $\Delta G = -18.2 \pm 5.9 \text{ kcal mol}^{-1}$ and $\Delta G = -7.8 \pm 12.9 \text{ kcal mol}^{-1}$, respectively, and points towards the importance of Ca^{2+} in annexin-HP binding. Furthermore, the results show that in combinations 2 to 6, the removal of ions contributed to the formation of more stable annexin V-HP complexes. Nevertheless, this information should be carefully

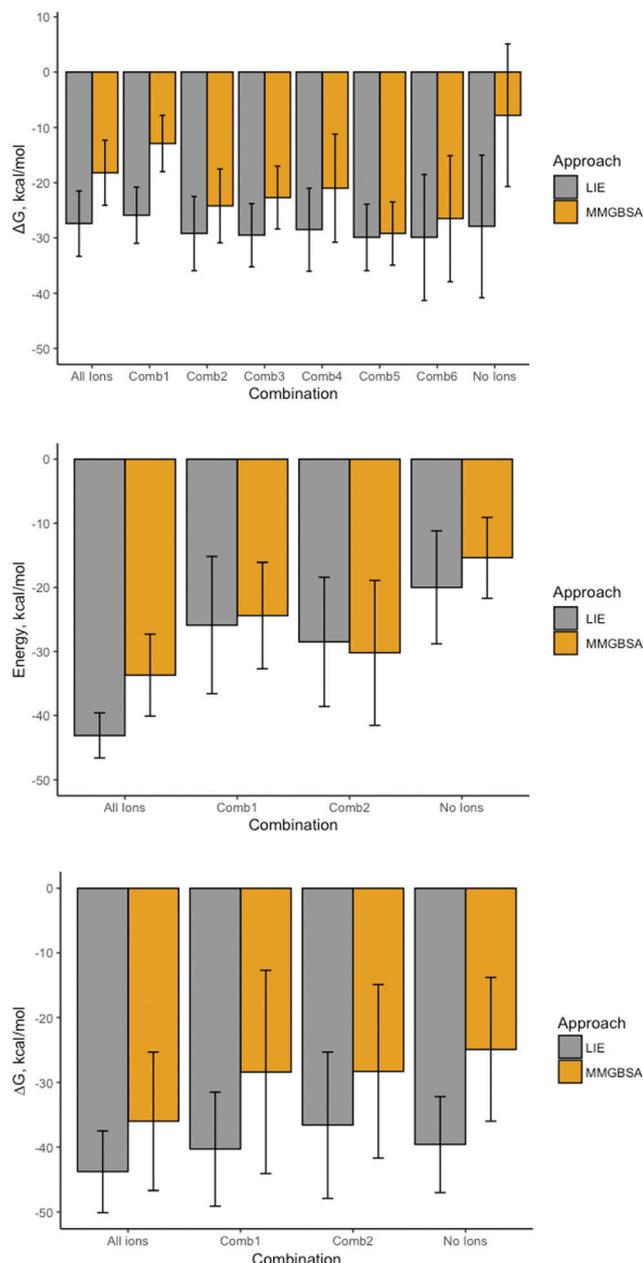


Fig. 5 Binding free energy dependence on the number of Ca^{2+} present in the investigated annexin-HP system predicted by Autodock3 and further simulated with the MD approach: annexin V-HPdp4 (top), annexin II-HP dp4 (middle), and annexin II-HPdp5 (bottom).

considered and suggests that it is rather the method being not sensitive enough to the probe for the effects of single ions. However, the data obtained in general are statistically insignificant due to the large margin of variance (Table S4, ESI†).

Annexin II-HP dp4/dp5. For the binding of both HP dp4 and HP dp5, the presence of Ca^{2+} seems to have a more noteworthy meaning compared to other complexes (Fig. 5: middle and bottom, respectively). In both cases, “All ions” have the lowest energetic values: for annexin II-HPdp4, $\Delta G = -43.1 \pm 3.5 \text{ kcal mol}^{-1}$, $\Delta G = -33.7 \pm 6.4 \text{ kcal mol}^{-1}$ and for annexin II-HP dp5, $\Delta G = -43.8 \pm 6.3 \text{ kcal mol}^{-1}$, $\Delta G = -36.0 \pm 10.7 \text{ kcal mol}^{-1}$ (LIE and MM-GBSA values, respectively; for details, see Tables S5 and S6, ESI†).

To complement these data, a 20 ns MD was run for the X-ray structures of each complex in the presence and absence of Ca^{2+} (Fig. 6). Consequently, LIE and MM-GBSA free energies were calculated (Tables S7–S9, ESI†). Both the end-point methods have shown that with the decreasing number of Ca^{2+} , the strength of interactions decreases. Among all the calculated annexin-HP complexes, the most stable ones were formed in the presence of Ca^{2+} with MM-GBSA being a more sensitive method. At the same time, despite this higher sensitivity of MM-GBSA for the MD simulations using experimental structures as initial points for calculations, we observed that the docked complexes of annexin II scored higher by the MM-GBSA method in comparison with the corresponding experimental

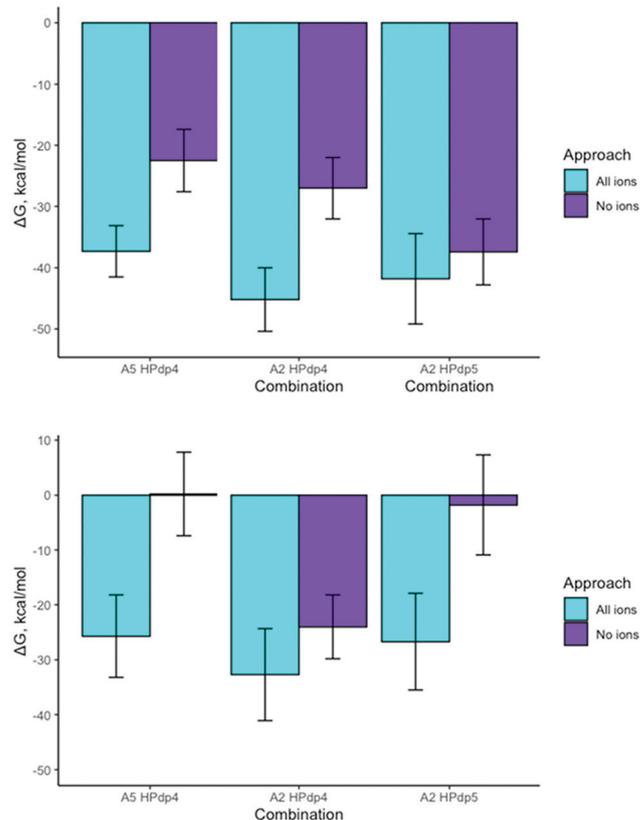


Fig. 6 Comparison of binding free energy dependence on the number of Ca^{2+} present in the investigated X-ray annexin-HP structures by two methods: LIE (top) and MM-GBSA (bottom).

structures, whereas LIE yielded more favourable binding free energies for the experimental complexes. We suggest that the potential reason for these findings could be the general inaccuracy of the implicit solvent model implemented in the MM-GBSA scheme for such highly charged intermolecular interfaces. The statistical significance of this difference is not obvious: there is a clear overlap between the values for docked and experimental solutions. More rigorous calibration analysis for a broader data set of protein–GAG complexes with many MD simulation repetitions is needed to conclude which putative reasons for such an observation could be decisive.

The low selectivity in these molecular complexes could also lead to challenges in reaching high performance by the application of AD3 and MD-based free energy calculations as described above. Therefore, we examined how selective the binding is in annexin-HP systems. Our data obtained from all the MD simulations performed for both the annexin complexes with all ions suggested very low to no selectivity (Tables S10–S15, ESI†).

Ca²⁺ binding analysis by free energy perturbation

FEP calculations were performed in order to estimate how strongly Ca²⁺ ions are bound in the interfaces of annexin-HP complexes and in the absence of HP ligand (Tables S16 and S17, ESI†). In the absence of HP, the following ranking of ion binding sites was obtained (from the strongest to the weakest): Ca-403, Ca-402, Ca-401 and Ca-507, Ca-508 for annexins V and II, respectively. The ranking changed upon HP binding: Ca-401, Ca-402, Ca-403 and Ca-508, Ca-507 for annexins V and II, respectively. The binding site affinity to Ca²⁺ also changed significantly in both systems. Furthermore, MM-GBSA calculations were performed to estimate whether the strength of these Ca²⁺ agree with the aforementioned findings, suggesting that MM-GBSA is able to rank the ion affinity properly. Data have shown that the *in vacuo* electrostatic energy (ΔG_{ele}) was highly negative and the following ranking of ions was recorded: Ca-401, Ca-402, Ca-403 and Ca-508, Ca-507 for annexins V and II, respectively (Tables S18–S27, ESI†). On the contrary, the per-ion decomposed energies were positive although all the Ca²⁺ were stable during the entire MD simulation performed in an explicit solvent. Since the structures were obtained from the X-ray experiments, these positive values are not due to any disparity in the initial conformation used for the MD. Therefore, using the implicit continuous solvent model as implemented in the MM-GBSA approach fails to properly account for the physically sound binding strength of these divalent ions in terms of the decomposed total binding free energy. Previously, we obtained similar results for other protein–GAG systems reported by Potthoff.⁵⁶ In addition, the transition between the explicit solvent model in the MD simulation and the implicit solvent model in the free energy calculations requires a reweighting procedure which is normally not carried out.⁵⁷ This also contributes to the positive values for MM-GBSA decomposed per-ion free energy, which would not be expected for stable molecular interactions. One of the reasons for observing the positive MM-GBSA per-ion decomposed energy could be the fact that there are still many challenges in the parameterization of

Ca²⁺ that are widely known.⁵⁸ For the analysis like the one in this work, it is, however, an important conclusion that the ranking of ions using MM-GBSA *in vacuo* electrostatic component is proper as it is shown when compared with the FEP results.

MM-GBSA vs. LIE comparison study for various protein–alginate/GAG systems and the potential impact of ions

Free-energy calculations have drawn more attention in the last few decades for their ability to provide an accurate estimate of the ligand-protein binding affinity and description of the ligand binding mechanism.^{59–67} Many studies compared MM-GBSA and LIE showing that for both, the quality of the results depends on the system investigated.^{68–71} Generally, LIE is faster and more precise than MM-GBSA, particularly when computationally intensive calculations of the entropic term are included and LIE parameters are calibrated for a particular class of ligands.⁷² Therefore, a comparative study between these two end-point methods was performed for GAG-containing systems with and without Ca²⁺ ions in GAG binding interfaces. The data were collected from various protein–GAG/alginate systems and grouped in terms of the protein's and sugar's charges (q) and sugar's length (Table S28, ESI†). Alginate is a periodic, linear, negatively charged polysaccharide made up of the same type of acidic monosaccharide units in different ring puckering conformations. The goal was to probe whether a linear relationship between LIE and MM-GBSA can be found as a function of $y = mx + c$. In other words, can an MM-GBSA value (y) be predicted based on the LIE outcome (x) and *vice versa*. If there were at least 10 MD repeats for a given protein–alginate/GAG system, the Pearson and Spearman correlation coefficients (r_p and r_s , respectively) between MM-GBSA and LIE were calculated, followed by the intercept (c) and slope (m) for each MM-GBSA vs. LIE function. Subsequently, the relationships between the protein's and sugar's charges vs. r_p , r_s , m and c were investigated. Interestingly, this study shows that the free energies calculated for protein–alginate complexes appear to correlate better than those for protein–GAG ones. Based on the analysis (Fig. S2, ESI†), it can be deduced that the longer (more negatively charged) the alginate is, the more negative the value of c and the smaller the value of m are observed. Additionally, the longer the alginate, the weaker the MM-GBSA vs. LIE correlation becomes (manifested by decreasing r_p and r_s coefficients). In the case of protein–GAG complexes (Fig. S3, ESI†), it seems that the correlation between MM-GBSA and LIE is not influenced by GAG's charge which could be attributed to the heterogeneity of the GAG-containing dataset. However, like in the situation of the protein–alginate complex, a more negative value of c and a smaller value of m are observed for more negatively charged GAGs. The final factor that was considered, protein's charge, has shown similar results (Fig. S4, ESI†). Therefore, this analysis confirms that to some degree, the correlation between MM-GBSA and LIE exists, although it substantially depends on the investigated system, and it could be possible to create a simple linear function that could be deployed to the free energy estimation. However, since the high variance was observed even for GAG-containing systems with

the same charge, such conclusions should be taken with caution, and a proper calibration should be performed for each particular system when possible. Besides, more systems should be carefully tested to make a general statement on the correlation between the two free energy calculation approaches. Additionally, the systems with various numbers of Ca^{2+} ions (for 1G5N, 2HYU and 2HYV) were inspected (Tables S4–S6, ESI†) to answer the question of whether the correlation between the MM-GBSA and LIE values is affected by the number of Ca^{2+} ions present in the complex. The results did not show any significant difference in the described relationships for the systems containing Ca^{2+} ions and other systems.

Hydrogen bonding in annexin-HP complexes

The crystallographic studies carried out by Capila *et al.*¹⁸ describe in detail the nature of annexin V-HP dp4 binding. In spite of the strict calcium-dependence of HP to annexin V and the importance of sulfate groups to the association, the sulfate groups do not interact directly with the annexin-bound Ca^{2+} . Instead, the oxygen atoms form hydrogen bonds to the backbone nitrogen atoms in the calcium binding loop. Additional hydrogen bonds are formed with water molecules that coordinate the bound Ca^{2+} in both the loops, and with a serine side chain in the IDE loop. Calcium coordination is independent of the HP presence. Another important determinant of the HP binding site is a basic cluster that includes Arg-23, Lys-27, Arg-61, and Arg-149 from the adjacent monomer. Hydrogen bonding contacts were calculated using X-ray structures with and without Ca^{2+} after 10 ns MD simulations (Tables S29–S31, ESI†). The results indicated that in the presence of Ca^{2+} , the contacts are in agreement with those that were reported by Capila *et al.*¹⁸ Additionally, there are substantial interactions of sulfate groups with Arg-23 and Arg-61. Upon Ca^{2+} removal, those contacts no longer exist although different hydrogen bonds are formed between annexin V and the sulfate groups of HP. That would support the hypothesis that indeed Ca^{2+} contributes to the establishment of loops in this area and those loops contribute to annexin V-HP binding and potentially to its specificity. Although hydrogen bonds are present between annexin II and HP, they are not established due to the presence of Ca^{2+} ions and therefore they were not analysed in this study.

The effect of tuning simulation parameters on the performance of the applied protocols in the annexin II-HP system

Taking into account difficulties experienced by the applied computational protocols, the simulation parameters were tuned for the annexin II-HP system, for which no crystal contact-related artifacts represented an obstacle. First, the impact of receptor conformation on molecular docking results was examined. To do this, 200 ns MD simulations were performed for the unbound annexin II with all, none, and the chosen Ca^{2+} . Five structures corresponding to the final frames for every 40 ns were chosen as receptors for docking for each system. This was done to evaluate the impact of receptor flexibility and to avoid any potential structural bias originating from the presence of Ca^{2+} on its surface. The comparison of the

docking results for the top 50 solutions, however, did not reveal any substantial impact of receptor conformational differences for either of the analysed complexes (Table S5, ESI†).

Furthermore, the potential effect of the MD length was investigated in the presence and absence of Ca^{2+} . The corresponding 20 ns long MD trajectories for X-ray complexes used as starting conformations were prolonged to 200 ns. All the complexes were stable already after 20 ns based on the RMSD values for the protein, ions, and ligand (Tables S6 and S7, ESI†) except for the ligand in the “No ions” scenario, for which the dissociation was observed after 20 ns and rebinding in a different binding pose (Table S7, ESI†). The LIE free energy values did slightly increase for both the systems when the results were compared between 20 ns and 200 ns MD simulations: $-41.8 \pm 7.4 \text{ kcal mol}^{-1}$ and $-37.4 \pm 5.4 \text{ kcal mol}^{-1}$ vs. $-36.2 \pm 8.0 \text{ kcal mol}^{-1}$ and $-35.2 \pm 9.1 \text{ kcal mol}^{-1}$ for “All ions” and “No ions”, respectively. For “All ions”, the total energy, electrostatic energy, and van der Waals components (Table S8, ESI†) revealed converged behaviour already after 20 ns, whereas for “No ions”, the total energy and the dominating van der Waals component dropped between 20 ns and 50 ns, then returned to a value similar to the first 20 ns and increased again after 150 ns. In the case of MM-GBSA, the energy values were also slightly lower for the longer simulation for “All ions,” $-26.7 \pm 8.8 \text{ kcal mol}^{-1}$ vs. $-3.5 \pm 8.4 \text{ kcal mol}^{-1}$, and decreased dramatically from -1.8 ± 9.1 to $-22.1 \pm 12.5 \text{ kcal mol}^{-1}$ for “No ions.” Such a drastic change of the energy corresponded to the dissociation in the first 20 ns followed by binding in a different binding pose in the long simulation (Fig. S9, ESI†). This, however, was not observed in the LIE results supporting the previous findings considering the differences in MM-GBSA and LIE sensitivities. Since protein–GAG interactions are usually weak and transient, it remains an open question whether such a change of a pose should be accounted for in the analysis of the initially evaluated binding pose. Therefore, it is up to the researcher to determine whether a longer simulation would rather provide more (and which particular) details to the understanding of a well-defined binding pose or whether the objective is to characterize the whole ensemble of possible conformations of the ligand in a binding site. In the latter, longer simulations are required while in the former, they do not provide much valuable information. Finally, the per-residue decomposition and visual structural inspection of a GAG allowed us to understand which specific changes in the energies and related conformations occurred in the longer MD simulation (Fig. S9 and Tables S32, S33, ESI†). For “All ions,” the decrease of binding strength corresponded to the weakening of the individual contribution of the oligosaccharide residue at the reducing end. In contrast, the dramatic change of conformation in the case of “No ions” led to the more favourable per-residue interactions in the new pose adapted by the ligand after 200 ns. To conclude, the MD simulation length is not a trivial parameter for the analysis of a protein–GAG system, and it should be set up depending on the scientific question to be answered in each case.

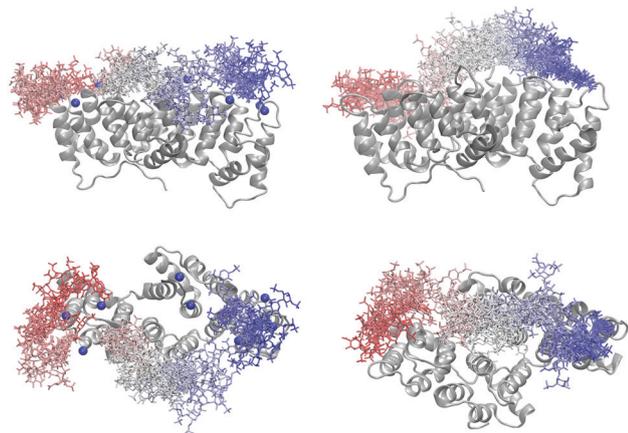


Fig. 7 HP dissociation pathway for the 1G5N complex at the force constant of $4 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$. In dark red the initial HP's position and the final one in dark blue. In the presence of Ca^{2+} (left) and without Ca^{2+} (right), side view (upper) and top view (lower).

HP dissociation pathway analysis by umbrella sampling

In US, 4 different systems have been investigated, each with two different force constants (Fig. S10–S17, ESI[†]), summing up to 8 independent US simulations (Table S34, ESI[†]). US was used to get an insight into a dissociation pathway of HP from annexin. Moreover, how the force constant and the presence/absence of Ca^{2+} influence the GAG's behaviour once pulled away from the protein was inspected (Fig. S14 and S15, ESI[†]). The initial distance between the centres of masses of the annexin and the HP was 5.5 \AA and the final value was 60.5 \AA .

Annexin V-HP dp4. In the case of annexin V, with the force constant of $4 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$, there is a substantial difference in energetic barrier for the complexes with and without Ca^{2+} with the corresponding PMF values of $64.2 \text{ kcal mol}^{-1}$ and $30.8 \text{ kcal mol}^{-1}$. This suggests that the energy required to pull away HP from the protein is twice as much for the scenario where Ca^{2+} ions are present. Fig. 7 represents the snapshots from each US window (55 in total) and combined together, they represent the HP dissociation pathway. Interestingly, in both cases, the HP molecule is not simply pulled away; in fact, it appears to travel along the upper surface of the protein. It can be observed that the HP chooses a different route when Ca^{2+} ions are present and probably it is their presence that dictates HP to take a more energetically demanding pathway (Fig. 7). On the other hand, when a force constant of $2 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ is applied, in the presence of Ca^{2+} , the HP is successfully pulled away from the protein (Fig. 8), whereas the pathway for the complex without Ca^{2+} remains the same. This observation that the protein's surface charge alone is enough to keep the GAG in close proximity to the receptor is in agreement with the previous findings from PBSA calculations and docking: HP's binding site might be dictated not only by the presence of Ca^{2+} but also by the lack of space due to other monomers being present in the crystal structure.

Annexin II-HP dp5. When a force constant of $2 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ is applied, there is a significant difference in the energetic barrier values: $103.7 \text{ kcal mol}^{-1}$ (Ca^{2+} present) and $8.3 \text{ kcal mol}^{-1}$ (Ca^{2+}

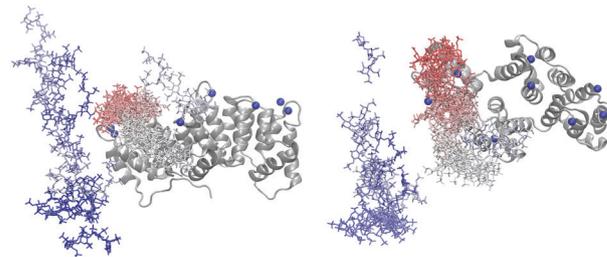


Fig. 8 HP dissociation pathway for the 1G5N complex in the presence of Ca^{2+} at the force constant of $2 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$. In dark red the initial HP's position and the final one in dark blue. On the left side view and on the right top view.

absent). From Fig. S7 (ESI[†]), it can be seen that the HP molecule is being pulled away and then, suddenly, at a distance of approximately 35.5 \AA , it moves in a discontinuous manner to the other patch on the protein surface, towards Ca^{2+} (number 605 in 2HYV), hence an increase in the PMF value. Without Ca^{2+} , the HP molecule appears to be almost effortlessly pulled away and this finding suggests that there are no other major annexin II-HP interactions that would try to keep the GAG close to the protein, unlike in the case of annexin V. Intriguingly, for a force constant of $4 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$, the energetic barrier has the same value of $23.5 \text{ kcal mol}^{-1}$ and takes place at the same distances of 35.5 \AA in the presence and absence of Ca^{2+} . Nevertheless, while looking at Fig. S8 (ESI[†]), HP's behaviour is entirely different in each case: Ca^{2+} forces the HP molecule to travel on a protein's surface, again towards Ca^{2+} (number 605 in 2HYV), whereas when Ca^{2+} ions were removed, the GAG dissociates straightaway. It is worth mentioning that although in both cases where Ca^{2+} ions are absent, an HP molecule is successfully pulled away with a lower force constant value, it is still interacting to some extent with the protein, unlike when $4 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ was used, and the GAG appeared to dissociate right away, causing an increase in the PMF value for the energetic barrier.

The reason why different energy barriers were calculated with different force constants used in the umbrella sampling MD simulations could be explained by the distinct dissociation pathways observed and, therefore, by distinct protein–ligand interactions established through these pathways.

Conclusion

In this study, for the first time, the role of Ca^{2+} on the annexin-HP binding was analysed computationally. The PBSA calculations have clearly shown that the presence/absence of Ca^{2+} has a substantial impact on the electrostatic potential of both the proteins' surfaces and, in consequence, could influence the HP-binding site. These findings were particularly important in the case of annexin V, where the location of the HP binding site, reported in the crystallographic studies,¹⁸ might have been partly dictated by the presence of neighbouring monomers within the crystal structure, suggesting that for some complexes, X-ray structures should be taken with a pinch of salt. Although Autodock3 has shown the difference between the “All ions” and the

“No ions” scenarios, this technique was not sensitive enough to probe for how the presence/removal of a single ion adds to the protein–GAG structure, and the binding and stability of the complex. At the same time, the RS-REMD approach to dock GAGs could be more powerful for the investigation of the ions’ impact on the molecular systems containing this class of oligosaccharides. Furthermore, the ranking of particular ions was performed using the FEP and MM-GBSA methods, confirming that the Ca^{2+} ions reported experimentally to be involved in the binding site contribute at most to the annexin-HP binding and these interactions are electrostatically-driven. An attempt was made to investigate whether there is a correlation between the two end-point free energy methods and how different factors (including sugar’s and protein’s charge or the number of Ca^{2+} ions present) might influence it. Umbrella sampling proved to be a valuable tool for examining the dissociation pathway of HP from annexin II/V in the presence/absence of Ca^{2+} . Based on our results, it is clear that Ca^{2+} ions play a role in how the GAG molecule interacts with the protein. Additionally, it is inconclusive which force constant should be chosen for the future, and it appears that each complex needs an individual treatment in terms of force constant choice. The obtained data provided a detailed and systemic description of the interactions between annexins and HP, which in turn allowed us to better understand the effect of Ca^{2+} ions present in the investigated complexes. To sum up, as described by Shao *et al.* and Capila *et al.*, the annexin-HP binding is determined by Ca^{2+} .^{12,18} Findings presented in this study might have a significant impact on the understanding of the limitations of the computational methodologies applicable to protein–GAG systems and they contribute to the general knowledge of the physicochemical basis underlying the interactions between proteins, ions and GAGs as well as their potential specificity.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This research was funded by the National Science Centre of Poland (Grant Number: UMO-2018/30/E/ST4/00037). Computational resources were provided by the Polish Grid Infrastructure (PL-GRID, Grants: plggagstr2 and plgca2gagphd), ZIH at TU Dresden (grant p_gag) and our cluster at the Faculty of Chemistry, University of Gdansk. The authors thank Krzysztof Bojarski for sharing his data on FGF-1-HP that contributed to MM-GBSA/LIE correlation analysis.

References

- 1 J. D. Esko, K. Kimata and U. Lindhal, Proteoglycans and Sulfated Glycosaminoglycans, 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

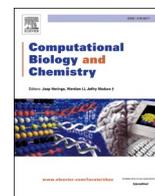
- SpringHarbor, NY, Cold Spring Harbor Laboratory Press, 2009.
- 2 H. Habuchi, O. Habuchi and K. Kimata, Sulfation pattern in glycosaminoglycan: Does it have a code?, *Glycoconjugate J.*, 2004, **21**, 47–52.
- 3 C. Gama, S. Tully, N. Sotogaku, P. Clark, M. Rawat, N. Vaidehi, W. Goddard, A. Nishi and L. Hsieh-Wilson, Sulfation patterns of glycosaminoglycans encode molecular recognition and activity, *Nat. Chem. Biol.*, 2006, **2**, 467–473.
- 4 A. Imberty, H. Lortat-Jacob and S. Pérez, Structural view of glycosaminoglycan-protein interactions, *Carbohydr. Res.*, 2007, **342**, 430–439.
- 5 P. Raynal and H. B. Pollard, Annexins: the problem of assessing the biological role for a gene family of multi-functional calcium- and phospholipid-binding proteins, *Biochim. Biophys. Acta*, 1994, **1197**, 63–93.
- 6 M. A. Swairjo and B. A. Seaton, Annexin structure and membrane interactions: a molecular perspective, *Annu. Rev. Biophys. Biomol. Struct.*, 1994, **23**, 193–213.
- 7 B. A. Seaton ed., *Annexins: Molecular Structure to Cellular Function*, R.G. Landes, Austin, TX, 1996.
- 8 V. Gerke and S. E. Moss, Annexins and membrane dynamics, *Biochim. Biophys. Acta*, 1997, **1357**, 129–154.
- 9 B. A. Seaton and J. R. Dedman, Annexins, *BioMetals*, 1998, **11**, 399–404.
- 10 V. Gerke, C. E. Creutz and S. E. Moss, Annexins: linking Ca^{2+} signalling to membrane dynamics, *Nat. Rev. Mol. Cell Biol.*, 2005, **6**, 449–461.
- 11 D. M. Waisman, Annexin II tetramer: structure and function, *Mol. Cell. Biochem.*, 1995, **149**, 301–322.
- 12 C. Shao, F. Zhang, M. M. Kemp, R. J. Linhardt, D. M. Waisman, J. F. Head and B. A. Seaton, Crystallographic analysis of calcium-dependent heparin binding to annexin A2, *J. Biol. Chem.*, 2006, **281**, 31689–31695.
- 13 K. Kojima, K. Yamamoto, T. Irimura, T. Osawa, H. Ogawa and I. Matsumoto, Characterization of carbohydrate binding protein p33/41: relation with annexin IV, molecular basis of the doublet forms (p33/41), and modulation of the carbohydrate binding activity by phospholipids, *J. Biol. Chem.*, 1996, **271**, 7679–7685.
- 14 G. Kassam, A. Manro, C. E. Braat, P. Louie, S. L. Fitzpatrick and D. M. Waisman, Characterization of the heparin binding properties of annexin II tetramer., *J. Biol. Chem.*, 1997, **272**, 15093–15100.
- 15 R. Ishitsuka, K. Kojima, H. Utsumi, H. Ogawa and I. Matsumoto, Glycosaminoglycan binding properties of annexin IV, V and VI., *J. Biol. Chem.*, 1998, **273**, 9935–9941.
- 16 I. Capila, V. A. VanderNoot, T. R. Mealy, B. A. Seaton and R. J. Linhardt, Interaction of heparin with annexin V., *FEBS Lett.*, 1999, **446**, 327–330.
- 17 S. L. Fitzpatrick, G. Kassam, A. Maro, C. E. Braat, P. Louie and D. M. Waisman, Fucoidan-dependent conformational changes in annexin II tetramer., *Biochemistry*, 2000, **39**, 2140–2148.
- 18 I. Capila, M. J. Hernaiz, Y. D. Mo, T. R. Mealy, B. Campos, J. R. Deadman, R. J. Linhardt and B. A. Seaton, Annexin V-heparin

- oligosaccharide complex suggests heparan sulfate-mediated assembly on cell surfaces, *Structure*, 2001, **9**, 57–64.
- 19 A. Almond, Multiscale modeling of glycosaminoglycan structure and dynamics: current methods and challenges, *Curr. Opin. Struct. Biol.*, 2018, **50**, 58–64.
 - 20 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 modelling, *Glycobiology*, 2012, **22**, 134–145.
 - 21 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.
 - 22 K. K. Bojarski, A. K. Sieradzan and S. A. Samsonov, Molecular Dynamics Insights into Protein-Glycosaminoglycan Systems from Microsecond-Scale Simulations, *Biopolymers*, 2019, **110**, e23252.
 - 23 U. Uciechowska-Kaczmarzyk, I. Chauvot de Beauchene and S. A. Samsonov, Docking software performance in protein-glycosaminoglycan system, *J. Mol. Graphics Modell.*, 2019, **90**, 42–50.
 - 24 S. Rother, S. A. Samsonov, T. Hofmann, J. Blaszkiewicz, S. Köhling, M. Schnabelrauch, S. Möller, J. Rademann, S. Kalkhof and M. von Bergen, 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.
 - 25 N. V. Sankaranarayanan, B. Nagarajan and U. R. Desai, U.R So you think computational approaches to understanding glycosaminoglycan-protein interactions are too dry and too rigid? Think again!, *Curr. Opin. Struct. Biol.*, 2018, **50**, 91–100.
 - 26 S. D. Vallet, A. E. Miele, U. Uciechowska-Kaczmarzyk, A. Liwo, D. Duclos, S. A. Samsonov and S. Ricard-Blum, Insights into the structure and dynamics of lysyl oxidase propeptide, a flexible protein with numerous partners, *Sci. Rep.*, 2018, **8**, 11768.
 - 27 N. E. Woodhead, W. F. Long and F. B. Williamson, Binding of zinc ions to heparin. Analysis by equilibrium dialysis suggests the occurrence of two, entropy-driven, processes, *Biochem. J.*, 1986, **237**, 281–284.
 - 28 I. Stevic, N. Parmar, N. Paredes, L. R. Berry and A. K. C. Chan, Binding of heparin to metals, *Cell Biochem. Biophys.*, 2011, **59**, 171–178.
 - 29 G. Multhaup, A. I. Bush, P. Pollwein and C. L. Masters, Interaction between the zinc (II) and the heparin binding site of the Alzheimer's disease beta A4 amyloid precursor protein (APP), *FEBS Lett.*, 1994, **335**, 151–154.
 - 30 R. Eckert and H. Raag, Zinc ions promote the interaction between heparin and heparin cofactor II, *FEBS Lett.*, 2003, **541**, 121–125.
 - 31 Q. Han, Y. Fu, H. Zhou, Y. He and Y. Lou, Contributions of Zn (II)-binding to the structural stability of endostatin, *FEBS Lett.*, 2007, **581**, 3027–3032.
 - 32 F. Zhang, X. Liang, J. M. Beaudet, Y. Lee and R. J. Linhardt, The effects of metal ions on heparin/heparin sulfate-protein interactions, *J. Biomed. Tech. Res.*, 2014, **1**, 1–7.
 - 33 U. Uciechowska-Kaczmarzyk, S. Babik, F. Zsila, K. K. Bojarski, T. Beke-Somfai and S. A. Samsonov, Molecular Dynamics-Based Model of VEGF-A and Its Heparin Interactions., *J. Mol. Graphics Modell.*, 2018, **82**, 157–166.
 - 34 K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. González-Outeiriño, C. R. Daniels, B. L. Foley and R. J. Woods, GLYCAM06: A Generalizable Biomolecular Force Field, Carbohydrates, *J. Comput. Chem.*, 2008, **29**, 622–655.
 - 35 C. J. M. Huige and C. Altona, Force Field Parameters for Sulfates and Sulfamates Based on Ab Initio Calculations: Extensions of AMBER and CHARMM Fields, *J. Comput. Chem.*, 1995, **16**, 56–79.
 - 36 D. A. Case, D. S. Cerutti, T. E. III Cheatham, T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Goetz, D. Greene, N. Homeyer, S. Izadi, A. Kovalenko, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. Mermelstein, K. M. Merz, G. Monard, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D. R. Roe, A. Roitberg, C. Sagui, C. L. Simmerling, W. M. Botello-Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, R. X. Wu, L. Xiao, D. M. York and P. A. Kollman, *AMBER 2017*, University of California, San Francisco, 2017.
 - 37 W. Humphrey, A. Dalke and K. Schulten, VMD - Visual Molecular Dynamics, *J. Mol. Graphics*, 1996, **14**, 33–38.
 - 38 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.
 - 39 A. Sarkar and U. R. Desai, A simple method for discovering druggable, specific glycosaminoglycan-protein systems. Elucidation of key principles from heparin/heparan sulfate-binding proteins, *PLoS One*, 2015, **10**(10), 1–18.
 - 40 V. Veverka, A. J. Henry, P. M. Slocombe, A. Ventom, B. Mulloy, F. W. Muskett, M. Muzylak, K. Greenslade, A. Moore, L. Zhang, J. Gong, X. Qian, C. Paszty, R. J. Taylor, M. K. Robinson and M. D. Carr, Characterization of the structural features and interactions of sclerostin: molecular insight into a key regulator of Wnt-mediated bone formation, *J. Biol. Chem.*, 2009, **284**(16), 10890–10900.
 - 41 S. A. Samsonov and T. M. Pisabarro, Computational Analysis of Interactions in Structurally Available Protein-Glycosaminoglycan Complexes, *Glycobiology*, 2016, **8**, 850–861.
 - 42 M. Ester, H. P. Kriegel, J. Sander and X. Xu, A Density-based algorithm for discovering clusters in large spatial databases with noise, KDD-96, 1996.
 - 43 T. Siebenmorgen, M. Engelhard and M. Zacharias, Prediction of protein-protein complexes using replica exchange with repulsive scaling, *J. Comput. Chem.*, 2020, **41**, 1436–1447.
 - 44 V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, Comparison of multiple Amber force fields and development of improved protein backbone parameters, *Proteins*, 2006, **65**(3), 712–725.

- 45 H. Nguyen, D. R. Roe and C. Simmerling, Improved Generalized Born Solvent Model Parameters for Protein Simulations, *J. Chem. Theory Comput.*, 2013, **9**(4), 2020–2034.
- 46 K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror and D. E. Shaw., Improved side-chain torsion potentials for the Amber ff99SB protein force field, *Proteins*, 2010, **78**, 1950–1958.
- 47 A. Tsui and D. A. Case, Theory and applications of the generalized Born solvation model in macromolecular simulations, *Biopolymers*, 2000–2001, **56**, 275–291.
- 48 A. Onufriev, D. Bashford and D. Case, Modification of the generalized Born model suitable for macromolecules, *J. Phys. Chem. B*, 2000, **104**, 3712–3720.
- 49 T. Siebenmorgen and M. Zacharias, Evaluation of Predicted Protein-Protein Complexes by Binding Free Energy Simulations, *J. Chem. Theory Comput.*, 2019, **15**(3), 2071–2086.
- 50 D. Hamelberg and J. A. McCammon, Standard free energy of releasing a localized water molecule from the binding pockets of proteins: double-decoupling method, *J. Am. Chem. Soc.*, 2004, **126**, 7683–7689.
- 51 S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman and J. M. Rosenberg, The weighted histogram analysis method for free-energy calculations on biomolecules. I: The method, *J. Comput. Chem.*, 1992, **13**, 1011–1021.
- 52 A. Grossfield, “WHAM: the weighted histogram analysis method”, version 2.0.10.2, http://membrane.urmc.rochester.edu/wordpress/?page_id=126.
- 53 R. Core Team (2020), R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, URL <https://www.R-project.org/>.
- 54 PyMOL: The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.
- 55 M. A. Lie, R. Thomsen, C. N. S. Pedersen, B. Schiøtt and M. H. Christensen, Molecular docking with ligand attached water molecules, *J. Chem. Inf. Model.*, 2011, **51**(4), 909–917.
- 56 J. Potthoff, K. K. Bojarski, G. Kohut, A. G. Lipska, A. Liwo, E. Kessler, S. Ricard-Blum and S. A. Samsonov, Analysis of Procollagen C-Proteinase Enhancer-1/Glycosaminoglycan Binding Sites and of the Potential Role of Calcium Ions in the Interaction, *Int. J. Mol. Sci.*, 2019, **20**(20), 5021.
- 57 F. Godschalk, S. Genheden, P. Soderhjelm and U. Ryde, Comparison of MM/GBSA calculations based on explicit and implicit solvent simulations, *Phys. Chem. Chem. Phys.*, 2013, **15**, 7731–7739.
- 58 A. Saxena and D. Sept, Multisite ion models that improve coordination and free energy calculations in molecular dynamics simulations, *J. Chem. Theory Comput.*, 2013, **9**(8), 3538–3542.
- 59 V. Limongelli, L. Marinelli and S. Cosconati, Sampling protein motion and solvent effect during ligand binding, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**(5), 1467–1472.
- 60 K. M. Merz and P. A. Kollman, Free-energy perturbation simulations of the inhibition of thermolysin—Prediction of the free-energy of binding of a new inhibitor, *J. Am. Chem. Soc.*, 1989, **111**(15), 5649–5658.
- 61 D. L. Mobley and M. K. Gilson, Predicting binding free energies: Frontiers and benchmarks, *Annu. Rev. Biophys.*, 2017, **46**, 531–558.
- 62 V. Limongelli, M. Bonomi and L. Marinelli, *et al.*, Molecular basis of cyclooxygenase enzymes (COXs) selective inhibition, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**(12), 5411–5416.
- 63 G. Grazionso, V. Limongelli and D. Branduardi, *et al.*, Investigating the mechanism of substrate uptake and release in the glutamate transporter homologue Glt(Ph) through metadynamics simulations, *J. Am. Chem. Soc.*, 2012, **134**(1), 453–463.
- 64 V. Limongelli, M. Bonomi and M. Parrinello, Funnel metadynamics as accurate binding free-energy method, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **111**(16), 6358–6363.
- 65 L. Wang, Y. J. Wu and Y. Deng, *et al.*, Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field, *J. Am. Chem. Soc.*, 2015, **137**(7), 2695–2703.
- 66 M. De Vivo and A. Cavalli, Recent advances in dynamic docking for drug discovery., *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2017, **7**(6), e1320.
- 67 M. De Vivo, M. Masetti, G. Bottegoni and A. Cavalli, Role of molecular dynamics and related methods in drug discovery, *J. Med. Chem.*, 2016, **59**(9), 4035–4061.
- 68 F. L. Gervasio, A. Laio and M. Parrinello, Flexible docking in solution using metadynamics, *J. Am. Chem. Soc.*, 2005, **127**(8), 2600–2607.
- 69 S. Genheden, MM/GBSA and LIE estimates of host-guest affinities: Dependence on charges and solvation model, *J. Comput.-Aided Mol. Des.*, 2011, **25**(11), 1085–1093.
- 70 S. Wong, R. E. Amaro and J. A. McCammon, MM-PBSA captures key role of intercalating water molecules at a protein-protein Interface, *J. Chem. Theory Comput.*, 2009, **5**(2), 422–429.
- 71 P. Mikulskis, S. Genheden, P. Rydberg, L. Sandberg, L. Olsen and U. Ryde, Binding affinities in the SAMPL3 trypsin and host-guest blind tests estimated with the MM/PBSA and LIE methods, *J. Comput.-Aided Mol. Des.*, 2012, **26**(5), 527–541.
- 72 R. R. 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.

PUBLICATION D2

**IMPACT OF CALCIUM IONS ON THE STRUCTURAL AND
DYNAMIC PROPERTIES OF HEPARIN
OLIGOSACCHARIDES BY COMPUTATIONAL ANALYSIS**



Impact of calcium ions on the structural and dynamic properties of heparin oligosaccharides by computational analysis

Małgorzata M. Kogut^a, Annemarie Danielsson^a, Sylvie Ricard-Blum^b, Sergey A. Samsonov^{a,*}

^a Faculty of Chemistry, University of Gdańsk, ul. Wita Stwosza 63, Gdańsk 80-308, Poland

^b Univ Lyon, University Claude Bernard Lyon 1, CNRS, Institute of Molecular and Supramolecular Chemistry and Biochemistry, UMR 5246, Villeurbanne CEDEX F-69622, France

ARTICLE INFO

Keywords:

Heparin
Calcium ions
Molecular dynamics
Conformational analysis
Glycosaminoglycans

ABSTRACT

Heparin (HP) belongs to glycosaminoglycans (GAGs), anionic linear polysaccharides composed of repetitive disaccharide units. They are key players in many biological processes occurring in the extracellular matrix and at the cell surface. GAGs are challenging molecules for computational research due to their high chemical heterogeneity, flexibility, periodicity, pseudosymmetry, predominantly electrostatics-driven nature of interactions with their protein partners and potential multipose binding. The molecular mechanisms underlying GAG interactions mediated by divalent ions, which are important for GAG binding to several proteins, are not well understood. The goal of this study was to characterize the binding of Ca^{2+} to two HP oligosaccharides of different lengths (dp10 and dp18, dp: degree of polymerization) and their impact on HP conformational space and their dynamic behavior with the use of molecular dynamics (MD)-based approaches with two Ca^{2+} parameter sets. MD data suggested that the flexibility of the monosaccharides, the glycosidic linkages and ring puckering were not affected by the presence of Ca^{2+} , in contrast to H-bond propensities and the calculated R_g for a fraction of the oligosaccharide populations in both dp10 and dp18. Moreover, the essential differences in the data obtained by using two Ca^{2+} parameter sets were reported.

1. Introduction

Glycosaminoglycans (GAGs) are linear long negatively charged polysaccharides made of disaccharide units comprised of amino sugar and uronic acid, except for keratan sulfate (Lindahl et al., 2015–2017). Moreover, GAG monosaccharide units have different sulfation patterns forming a GAG sulfation code (Habuchi et al., 2004). GAGs are found in the extracellular matrix and at the cell surface, where they play key roles as part of proteoglycans, besides the non-covalently bound hyaluronan, in numerous cellular (e.g., cell proliferation, adhesion and signalling) and biological processes such as development, angiogenesis, coagulation and diseases (e.g., cancer and infectious and neurodegenerative diseases) (Karamanos et al., 2018) which are mediated by their interactions with numerous protein partners (Vallet et al., 2021). Protein-GAG interactions may be electrostatically-driven, defined by the charge density distribution on the protein surface, or may involve specific GAG-structural features (Kjell  n and Lindahl, 2018). Heparin (HP) is a GAG that consists of a 2-O-sulfated iduronic acid (IdoA2S) and

6-O-sulfated and N-sulfated glucosamine (GlcNS6S) with a -4 net charge per disaccharide unit. These disaccharides are also the major building blocks of highly heterogenous heparan sulfate (Shriver et al., 2012), which can consist of up to 48 types of theoretically possible disaccharide units (Lu et al., 2006).

Interactions between GAGs and cations (*i.e.*, Zn^{2+} , Mn^{2+} , Cu^{2+} , Ca^{2+} , Co^{2+} , Na^+ , Mg^{2+} , Fe^{3+} , Ni^{2+} , Al^{3+} , and Sr^{2+}) have been experimentally identified (Woodhead et al., 1986; Stevic et al., 2011). Their crucial role in protein-GAG interactions was shown experimentally for the amyloid precursor protein (APP) (Multhaup et al., 1994), HP cofactor II (Eckert and Raag, 2003), endostatin (Sasaki et al., 1999; Ricard-Blum et al., 2004; Han et al., 2007), Fibroblast Growth Factor-1 (FGF1), interleukin-7 (IL-7) (Zhang et al., 2014), and procollagen C-proteinase enhancer-1 (PCPE-1) (Weiss et al., 2010) and sclerostin (Zhang et al., 2020). However, few theoretical analyses were performed to explore those experimental findings and the formation of protein-ion-GAG complexes at the conformational and atomic levels. We have previously investigated the effect of a divalent cation (Ca^{2+}) on GAG binding

Abbreviations: GAG, Glycosaminoglycan; HP, Heparin; dp, degree of polymerization; MD, Molecular Dynamics; H-bond, Hydrogen bond.

* Corresponding author.

E-mail address: sergey.samsonov@ug.edu.pl (S.A. Samsonov).

<https://doi.org/10.1016/j.compbiolchem.2022.107727>

Received 12 May 2022; Received in revised form 24 June 2022; Accepted 7 July 2022

Available online 9 July 2022

1476-9271/  2022 Elsevier Ltd. All rights reserved.

to the Ca²⁺-binding protein PCPE-1 by molecular modelling and have shown that Ca²⁺ may either bind to GAGs before they interact with PCPE-1, or that they may stabilize the conformation of the full-length protein. Moreover, we analysed the crystal structures of annexin-Ca²⁺-HP complexes (Shao et al., 2006; Capila et al., 2001) using molecular dynamics (MD)-based approaches, and reported that the results of molecular docking, MD, and binding free energy calculations as well as the calculated ligand dissociation pathways were affected by accounting for the presence of Ca²⁺ in the modelling protocols (Kogut et al., 2021). While experimental techniques to study protein-ion-GAG ternary complexes or ion-GAG interactions are not always sufficient to gain insights into GAG-ion interactions at the atomic level, theoretical approaches can provide such specific details (Almond, 2018; Kogut et al., 2022). There are many excellent reviews that offer more background on GAG structural analysis by computational tools as well as challenges encountered in the MD studies of these molecules (Nagarajan et al., 2022, 2019; Fadda and Woods, 2010; Sankaranarayanan et al., 2018). Recent advances in the characterization and modelling of protein-ion-GAG complexes (Kogut et al., 2021; Potthoff et al., 2019; Guevench and Whitmore, 2021) have been made indeed. However, the investigation of binary and ternary protein complexes involving both GAGs and ions remains challenging due to the high negative charge of GAGs, which leads to the prominence of electrostatically driven interactions, abundant solvent-mediated interactions (Samsonov et al., 2011a) and ion-dependent mechanisms. In addition, only three X-ray structures (1G5N, 2HYV, 2HYU), which correspond to three different protein-Ca²⁺-HP systems, encompassing a protein bound to a GAG and containing an ion in the GAG binding interface, are available (Perez et al., 2020).

Another challenge in modelling protein-ion-GAG interactions is the choice of the appropriate parameters for ions and the limited knowledge on their influence on specific ion effects. Most of the current ion models in molecular mechanics (referred to as type I throughout this publication) (Case et al., 2017) are simple spheres, and their interactions are exclusively determined by Lennard-Jones parameters and a charge localized in a single point. Saxena and Sept proposed a fundamentally different model (referred to as type II throughout this publication, Supplementary Figure 1) where the total charge of the ion is distributed into *n*-dummy centres placed to appropriately account for the ion's coordinating features (Saxena and Sept, 2013). Although Ca²⁺ can dynamically switch from 6 to 8 coordinating atoms, the pentagonal bipyramid coordination of Ca²⁺ is the most common form in biological systems, and, therefore, is used in this parametrization. Furthermore, it has been recently highlighted that the accurate modelling of ion-containing molecular systems has a significant computational cost, and that further research is needed in this area (Li and Merz, 2017; Miranda-Quintana and Smiatek, 2021).

This study aimed to characterise the structural and dynamic properties of HP oligosaccharides of various lengths (dp10 and dp18; dp stands for the degree of polymerization) by performing an in-depth computational analysis of these HP molecules to investigate the impact of Ca²⁺, using both Ca²⁺ ion models, on the oligosaccharide conformation in a microsecond-timescale MD simulation. A decrease in the simulated radius of gyration (*R_g*) was observed in a fraction of HP oligosaccharides. While the flexibility of the monosaccharides, glycosidic linkages and ring puckering were not affected by the presence of Ca²⁺, H-bond propensities, *R_g* and HP thermodynamics properties were shown to be Ca²⁺-dependent. This dependence was found to be essentially affected by the ion model utilised in the simulations.

2. Materials and methods

2.1. Modelled systems

The structures of HP dp10 and dp18 were generated from the experimental structure of HP (dp12, PDB ID: 1HPN) with IdoA2S being

in ¹C₄ conformation. Two parameter sets were used for Ca²⁺ ions as mentioned above, namely Ca²⁺ (type I) and Ca²⁺ (type II). In total, six molecular systems were prepared for the simulations consisting of HP dp10 and HP dp18 without Ca²⁺ and with 10 and 18 of Ca²⁺ of both types, respectively.

2.2. Molecular dynamics

Molecular Dynamics (MD) simulations of the six systems were performed with the AMBER16 software package (Case et al., 2017). The residue libraries for HP monosaccharide units were used (Pichert et al., 2012), where the charges were obtained from the GLYCAM06 force field (Kirschner et al., 2008) and from the literature (Huige and Altona, 1995). Truncated octahedron TIP3P periodic box of 8 Å water layer from box border to the solute was used to solvate complexes. This corresponded to ~8500 and ~25000 water molecules in periodic boundary boxes for HP dp10 and HP dp18, respectively. The systems without Ca²⁺ were neutralized with 20 and 36 Na⁺ counterions for HP dp10 and HP dp18, respectively. Energy minimization was carried out in two steps: 500 steepest descent cycles and 10³ conjugate gradient cycles with 100 kcal/mol/Å² harmonic force restraints were followed by 3 × 10³ steepest descent cycles and 3 × 10³ conjugate gradient cycles without any restraints. The system was then heated from 0 to 300 K for 10 ps with harmonic force restraints of 100 kcal/mol/Å² and equilibrated at 300 K and 10⁵ Pa in isothermal isobaric ensemble for 100 ps. Afterwards, the MD production run was performed in the same isothermal isobaric ensemble for 1 μs. The MD run was performed five times and the simulations were combined yielding 5 μs for each system, in total. According to Genheden and Ryde, running several independent simulations and averaging the results before analysis is statistically more appropriate than performing a single simulation of the length equal to the sum of the lengths of the shorter simulations (Genheden and Ryde, 2010). Particle mesh Ewald method for treating electrostatics and SHAKE algorithm for all the covalent bonds containing hydrogen atoms were implemented in the MD simulations.

2.3. Free energy components and entropy analysis

Molecular Mechanics Generalized Born Surface Area (MM/GBSA) with igb=2 within AMBER16 software package (Onufriev et al., 2000) was used for the free energy calculations. Entropy calculations were performed using quasi-harmonic (QH) and normal mode (NM) approaches as implemented in AMBER16 within the default procedure. It should be kept in mind that entropy calculations for nanosecond-scale MD simulations may be very noisy and therefore should be interpreted meticulously for the systems with a high number of degrees of freedom (Homeyer and Gohlke, 2012; Gandhi and Mancera, 2009). QH involves calculations of the covariance matrix of atomic coordinates from MD simulations resulting from a harmonic energy surface, while neglecting anharmonic energy surface, and computes QH force constants, which are then used to calculate the entropy (Karplus and Kushick, 1981). In this way, the QH approach yields the conformational entropy of a molecule in an "effective" quadratic potential, implicitly including effects of solvent. Even though QH is essentially faster than NM, it is usually less accurate and has several limitations (Chang et al., 2005). NM method analyses the harmonic frequencies obtained from minimized snapshots of MD simulations (Srinivasan et al., 1998) and estimates the conformational entropy of a molecular system *in vacuo*. Despite being computationally expensive, this method has a solid capability to yield potentially valuable results in reproducing experimental binding affinities (Genheden et al., 2012). The conformational entropy calculated by both above-mentioned methods represents only a part of the total entropy.

In these calculations, all frames from the MD production run were considered.

2.4. Computational analysis of the HP properties

The obtained MD trajectories for the six systems were processed with the CPPTRAJ (Roe and Cheatham, 2013) tool to generate the following data:

- The radius of gyration (R_g) used in polymer physics to describe the dimensions of a polymer chain (here HP oligosaccharides). It is defined as:

$$R_g^2 \stackrel{\text{def}}{=} \frac{1}{N} \sum_{k=1}^N (r_k - r_{\text{mean}})^2$$

where r_{mean} is the mean position of N atoms.

- Sugar ring puckering, calculated by determining the fraction of frames where a particular conformation occurred. Ring puckering Cremer-Pople conformation definitions (Cremer and Pople, 1975) were based on experimentally available data (for details see Table S1).
- The distribution of dihedral angles of 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 n stands for the sequential number of a sugar monomeric unit.
- Per-residue root mean square (rms) fluctuations.
- Fractions and lifetimes of intramolecular hydrogen bond (H-bonds) for HP oligosaccharides. They were calculated defining a H-bond by the default values in CPPTRAJ (Roe and Cheatham, 2013) with a distance cut-off of 3.0 Å between the heavy atoms (H-bond acceptor and the one covalently bound to the hydrogen atom) and a hydrogen bond angle cut-off of 135°. The number of reported hydrogen bonds was averaged over the individual MD trajectories and across replicates, summarized and visualized using in-house scripts coded in Python 3.8.5 using the numpy 1.19.2 (Harris et al., 2020), pandas 1.1.3 (McKinney, 2010) and matplotlib 3.3.2 (Hunter, 2007) libraries.

Additionally, we performed analysis to monitor the coordination of Ca^{2+} to HP's groups capable of this coordination: 2 sulfate groups (NS and 6 S) from GlcNS6S and one sulfate group (S) plus one carboxylate group (C6) from IdoA2S sugar unit. The distances between each Ca^{2+} and NS, 6S, S and C6 were calculated with CPPTRAJ. Then for each type of Ca^{2+} cut-off distances were measured based on distance distributions for the coordinated ions: for Ca^{2+} (type I) was 4.24 Å for sulfate groups and 4.02 Å for carboxylate group whereas for Ca^{2+} (type II) it was 4.70 Å for sulfate groups and 4.55 Å for carboxylate groups. Finally, the probability of each Ca^{2+} to be coordinated by a particular HP's chemical group was calculated.

2.5. Data analysis and visualization

Statistical analysis and graphical presentation of the data were performed with R-package (Team, 2013) and gnuplot (Williams and Kelley, 2013). Each trajectory was visualized with VMD (Humphrey et al., 1996).

3. Results and discussion

The focus of our study was solely on HP- Ca^{2+} interactions. The role of Na^+ was not investigated in this work in the detail. Visualization of Na^+ ions in the obtained trajectories did not reveal either specific interaction between these ions and particular groups of the HP molecules or any events in which these interactions were established longer than several picoseconds. The occupancy of specific position of Ca^{2+} is challenging to analyse visually due to the high flexibility of HP: when the HP trajectory is self-aligned, the respective modification of Ca^{2+} trajectory positions in space yield noisy representation, which is not useful for drawing any qualitative or quantitative conclusions.

Therefore, we described HP- Ca^{2+} interactions in terms of the following structural, dynamic and energetic parameters: radius of gyration, ring puckering and glycosidic linkage conformational space, monosaccharide flexibility, intramolecular hydrogen bonding and MM-GBSA free energy components.

3.1. Radius of gyration

The flexibility of HP dp10 and HP dp18 was investigated by calculating their R_g in the absence and presence of Ca^{2+} . For HP dp10 a small peak with a shoulder (10.5 Å) and a major well-defined peak (13 Å) were observed in the absence of Ca^{2+} . The presence of Ca^{2+} (type I) induced a broadening of the peak at 13 Å whereas a well-defined peak with a lower R_g value (9.5 Å) was additionally observed. A decrease in the R_g values was also observed when the calculations were made with Ca^{2+} (type II) with the appearance of two peaks with R_g of 11 and 12 Å (Fig. 1). These data suggest that the significant part of the oligosaccharide population undergoes a decrease in R_g value when the calculations are made with Ca^{2+} (type II), whereas only a far smaller fraction of the population undergoes this shift when the parameters of Ca^{2+} (type I) are used. This is reflected in the appearance of the probability density peaks that are missing for the data obtained from the trajectories in the absence of Ca^{2+} (left-shifted peaks in the probability density are labeled with an asterisk * in Fig. 1). Ca^{2+} also essentially induced a decrease in R_g values in HP dp18 oligosaccharides for both Ca^{2+} types. Recently Lazar et al. (2022) reported that conformational collapse of oligosaccharides in the MD simulations with GLYCAM06j force field, which is used in our study, could occur for saccharide chains containing the deoxy sugar α -L-rhamnose after 400 ns (Lazar et al., 2022). They deduced that the irreversible collapse involved a formation of a hairpin structure, with a hinge in the middle of the saccharide and subsequent stabilization of the hairpin by multiple hydrogen bonds between opposing arms of the chain. Therefore, to refine our understanding of R_g changes in the presence of Ca^{2+} during the simulations, the R_g values were plotted as a function of time for HP dp10 (Fig. S2) and HP dp18 (Fig. S3). In the absence of Ca^{2+} in 3 out of 5 MD simulations we did not observe any significant changes in R_g for HP dp10 (Fig. S2A) and similar was found to be true for HP dp18 (Fig. S3A). Furthermore, regardless of the inspected oligosaccharide and the type of Ca^{2+} used, the drop in R_g values was detected and apart from one MD simulation for HP dp18 in the presence of Ca^{2+} (type I), the changes in R_g value were gradually decreasing with time. Additionally, we performed a visual inspection of the HP dp18 structure in the absence and the presence of Ca^{2+} in 1 μ s MD simulations. We chose the simulations for which the most pronounced decrease in R_g was observed (Fig. S3). The snapshots from the simulations were taken at 0, 250, 500, 750 and 1000 ns and are depicted in Fig. 2. Although some bends were established in the HP structures (mainly at the chain's ends) towards the end of simulations, no conformational collapse was observed in any of the investigated systems. Moreover, GLYCAM06j recently proved to be capable of yielding good agreement with the NMR data for the conformational ensemble of Arixta, which is a heparan sulfate specific pentasaccharide (GlcNS6D-GlcA-GlcNS3S6S-IdoA2S-GlcNS6S) (Janke et al., 2022). Based on our findings, we claim that R_g of oligosaccharides of various sizes is affected by the presence of Ca^{2+} in the MD simulations. The MD data are consistent with previous MD-based findings showing that Ca^{2+} and sulfation favour compact conformations of chondroitin sulfate (Guevench and Whitmore, 2021).

3.2. Ring puckering conformational space

The ability to accurately model HP and other oligosaccharides is hindered by the ring flexing (or puckering) motions that occur in individual monosaccharides. Puckering of six-membered pyranose rings is proposed to be a 3D microsecond-timescale signature of numerous chemically distinct monosaccharides (Sattelle et al., 2010, 2012; Sattelle

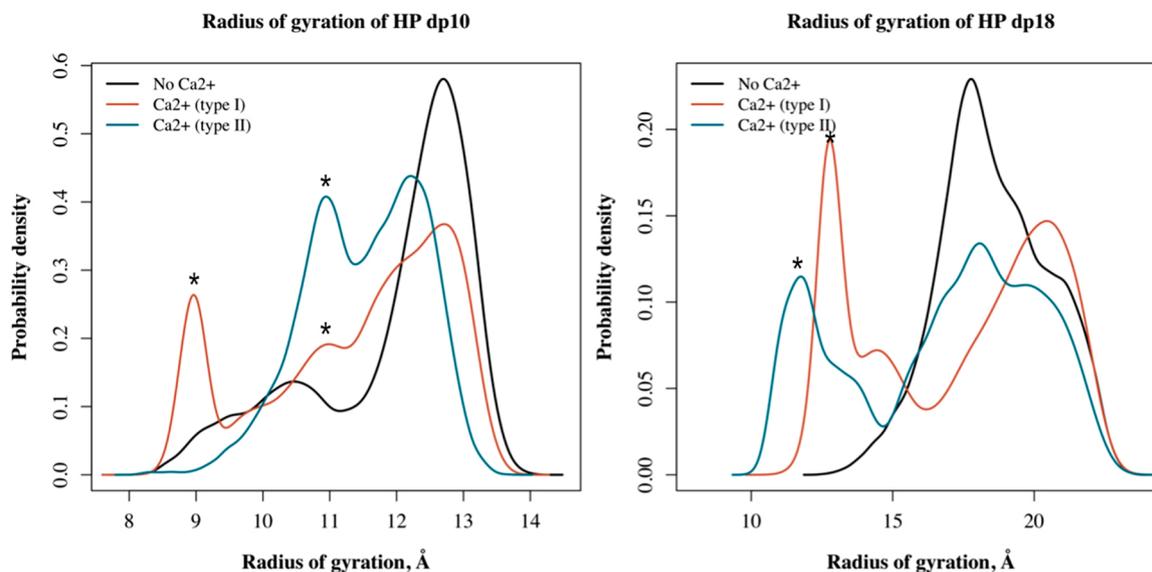


Fig. 1. The radius of gyration of HP dp10 (left) and HP dp18 (right) in the absence and presence of each Ca^{2+} type. All frames from $5 \times 1 \mu\text{s}$ MD simulations were analysed to calculate the probability. The peaks in the probability density caused by the presence of Ca^{2+} are labeled with *.

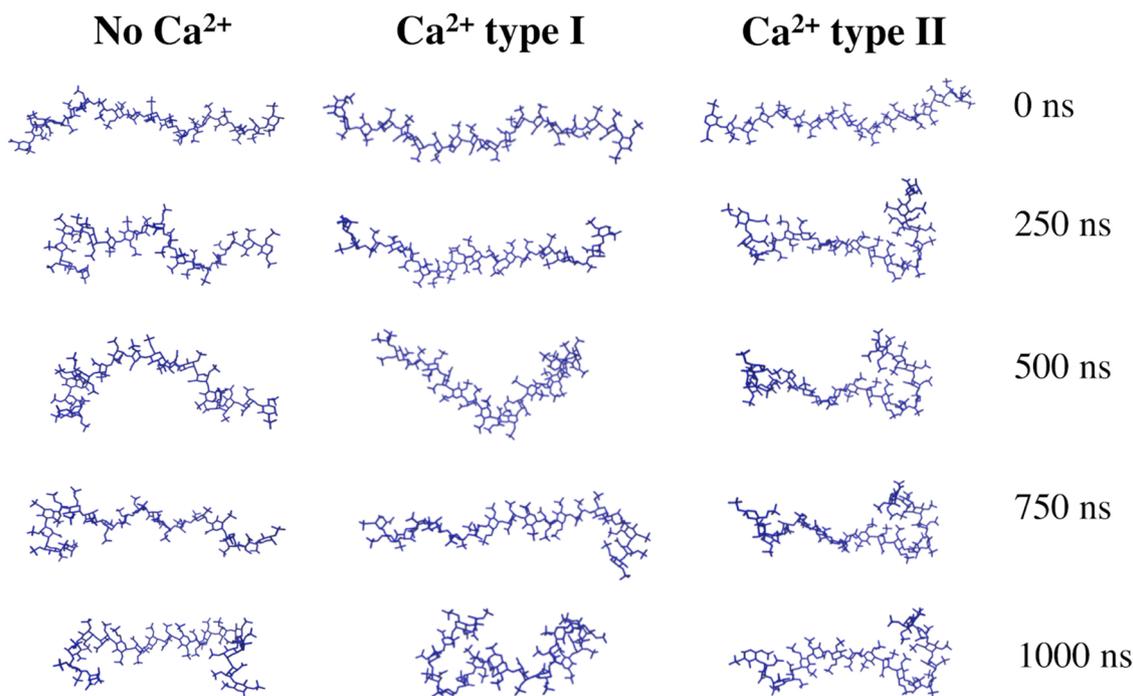


Fig. 2. Behaviour of HP dp18 during the $1 \mu\text{s}$ MD simulations in the absence and presence of Ca^{2+} . The snapshots were taken from all 5 MD simulations at 0, 250, 500, 750 and 1000 ns, respectively. For the sake of clarity hydrogens and Ca^{2+} are not shown.

and Almond, 2011, 2012) and is mechanistically critical for GAG molecular recognition and bioactivity (e.g., in heparin-mediated anticoagulation) (Hricovini et al., 2001; Das et al., 2001; Samsonov et al., 2011b). Nevertheless, the puckering is difficult to quantify experimentally since only the average values from the complete conformational ensembles are accessible in experiments. Furthermore, few computational studies have performed rigorous microsecond simulations of oligosaccharides to investigate this phenomenon in biologically relevant carbohydrate sequences (Sattelle et al., 2015). We report here for the first time the impact of Ca^{2+} on HP monosaccharide unit ring puckering throughout the MD simulations and visualise it as bar plots reflecting the percentage of the conformations of each ring (i.e., ${}^1\text{C}_4$, ${}^4\text{C}_1$, ${}^2\text{S}_0$, ${}^1\text{S}_3$,

Fig. 3). The GlcNS6S residues of both HP dp10 and HP dp18 predominantly adopted ${}^4\text{C}_1$ ring conformation except for GlcNS6S (1) that also adopted ${}^1\text{C}_4$ conformation in the presence of Ca^{2+} . This conformation accounted for 20 % of the HP dp10 population for Ca^{2+} (type I) and 60 % for Ca^{2+} (type II), and a bit less than 10 % and 20 % in HP dp18 oligosaccharides, respectively (Fig. 3). Furthermore, for IdoA2S residues three conformations ${}^1\text{C}_4$, ${}^4\text{C}_1$, and ${}^2\text{S}_0$ were observed. The ${}^2\text{S}_0$ conformation is usually the less favourable one and was observed in HP dp10 and HP dp18 in the presence of Ca^{2+} (type II), and in trace amounts in HP dp10 without Ca^{2+} . To sum up, for HP dp10, the presence of Ca^{2+} affected ring conformations of residues 1 and 4 regardless of the Ca^{2+} type. For HP dp18, the conformational changes were observed for the

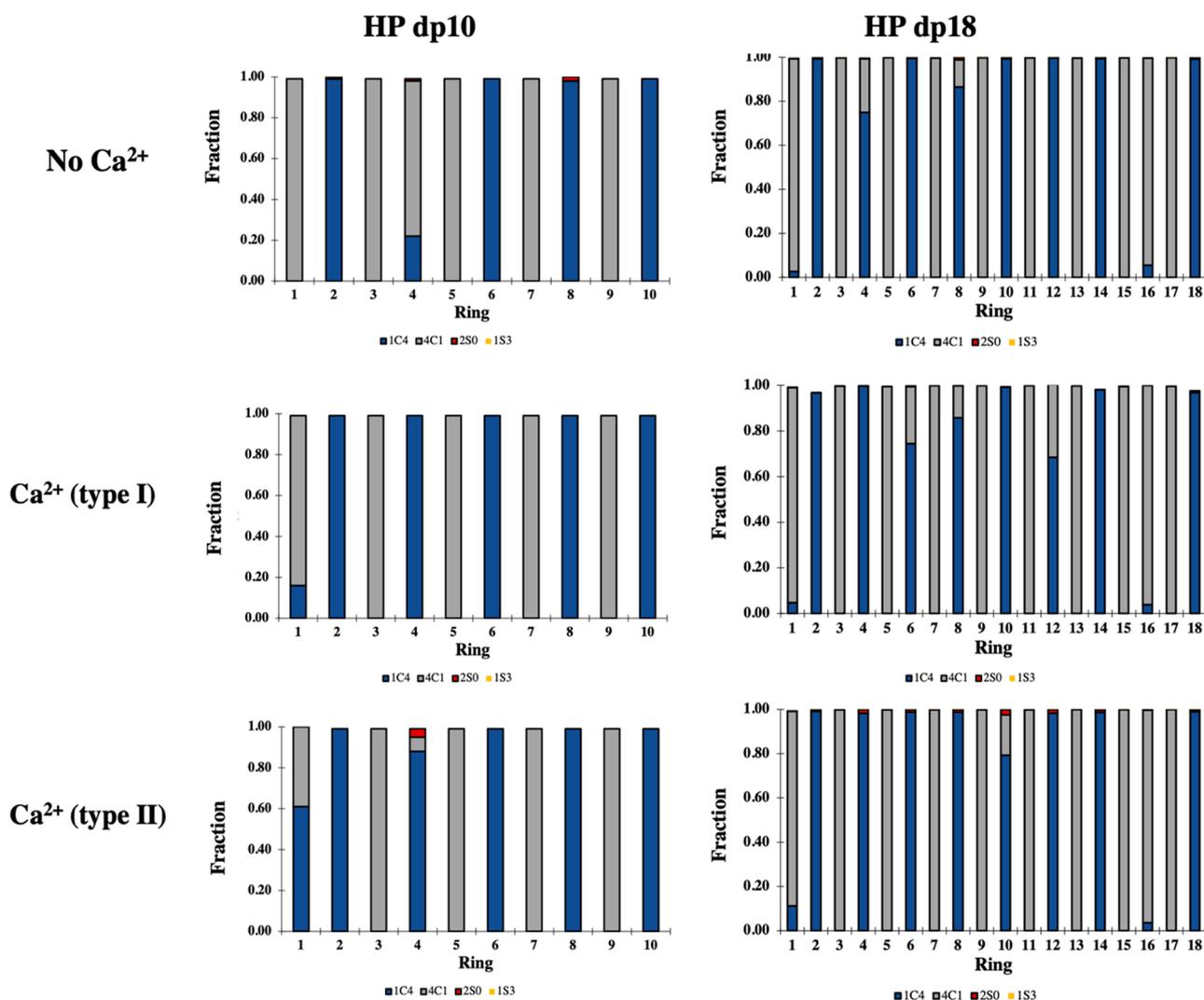


Fig. 3. Fractions of the HP residues ring puckering conformations 1C_4 , 4C_1 , 2S_0 , 1S_3 . For clarity, name and number of each residue was substituted with its number (GlcNS6S: odd numbers; IdoA2S: even numbers).

rings in residues 1, 4, 10 and 12 with differences for rings 4, 10 and 12 depending on the Ca^{2+} type. These results are consistent with the data obtained by [Sattelle et al. \(2013\)](#) from 5 μs MD simulations for HP and its fragments of different lengths ([Sattelle et al., 2013](#)) as well as with our data from 10 μs MD simulations for HP dp6 ([Bojarski et al., 2019](#)), showing the energetic preference of the 1C_4 conformation for the IdoA2S ring independently of Ca^{2+} presence. However, taking into account that our simulations started with IdoA2S in 1C_4 conformation and that their total duration (5 μs) was similar to the characteristic times for the ring puckering transitions reported by [Sattelle et al. \(2010\)](#), (2013), we cannot rule out that effects of Ca^{2+} on HP ring puckering could be observed in longer MD simulations.

3.3. Glycosidic linkage conformational space

Additionally, the conformational space of HP glycosidic linkages was analysed for HP dp10 and HP dp18 in the absence and presence of Ca^{2+} . The conformations of glycosidic linkages were almost identical in all cases ([Figs. 4 and 5](#)). The heatmaps were obtained by summing up all glycosidic linkage populations corresponding to GlcNS6S-IdoA2S (referred to as even) and IdoA2S-GlcNS6S (referred to as odd) for each HP oligosaccharide. For GlcNS6S-IdoA2S, the first free energy minimum corresponds to the dihedral angles $\varphi \sim -100^\circ$ and $\psi \sim 100^\circ$ and the second one to $\varphi \sim -100^\circ$ and $\psi \sim -50^\circ$. For IdoA2S-GlcNS6S

only one minimum was observed for $\varphi \sim 100^\circ$ and $\psi \sim 100^\circ$. The results agree with those previously reported by [Sattelle et al. \(2013\)](#) and [Bojarski et al. \(2019\)](#). Besides, for each heatmap, the number of points near the minimum (within either ± 2 or ± 6 kcal/mol) was calculated, presented as a fraction, and summarized in [Tables S2 and S3](#). For both HP dp10 and HP dp18 the lowest values were obtained when Ca^{2+} (type II) parameters were used in the MD simulations. Despite these small differences, no overall trend in the heatmaps can be proposed suggesting that either Ca^{2+} does not significantly influence the conformation of glycosidic linkages or that the force field we used prevented the observation of differences.

3.4. Monosaccharide flexibility

HP oligosaccharide behavior was also investigated in terms of per-residue root mean square fluctuations. Apart from the characteristic increase in the fluctuation of the first ring in each case, the subsequent rings looked rather similar, and no significant changes were observed in the presence of Ca^{2+} whatever the model used ([Fig. 6](#)).

3.5. Intramolecular hydrogen bonds

Another parameter that was considered in our computational analysis was the intramolecular hydrogen bonds. The total number of

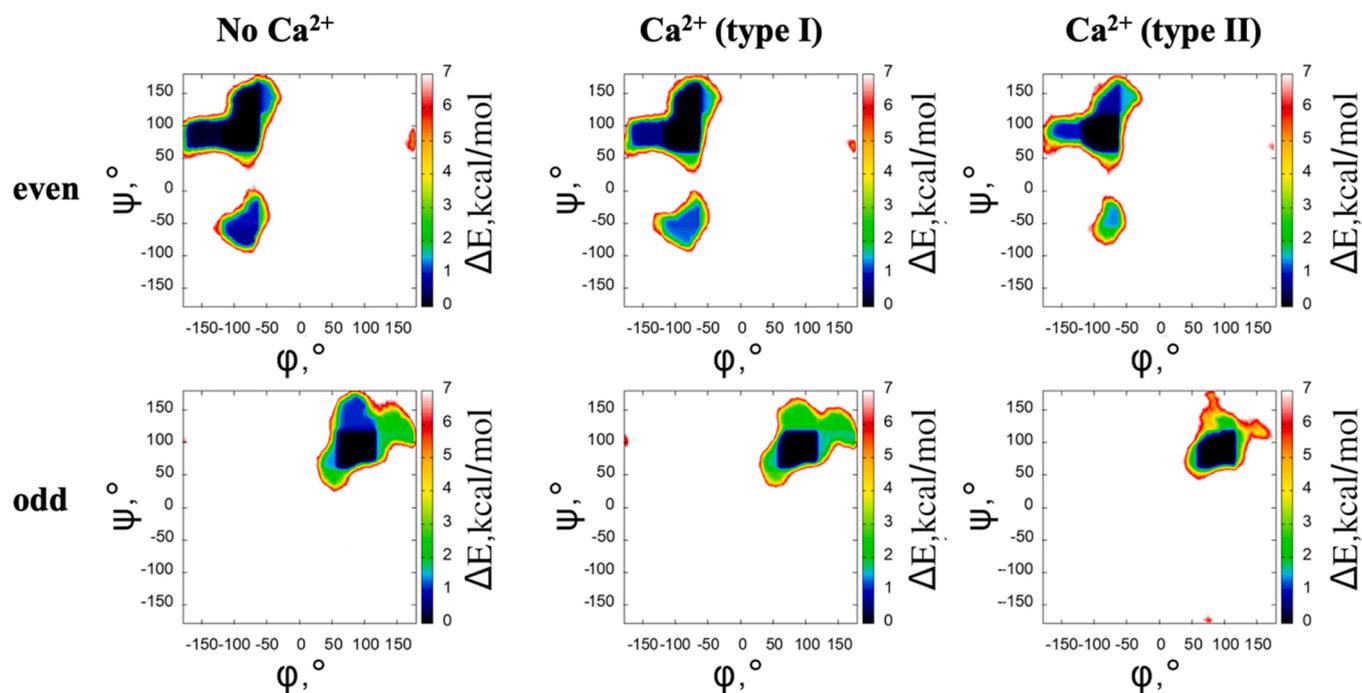


Fig. 4. Glycosidic linkage heatmaps for ϕ and ψ dihedral angles of HP dp10. The data have been summed up for all GlcNS6S-IdoA2S (even) and all IdoA2S-GlcNS6S (odd) glycosidic linkages.

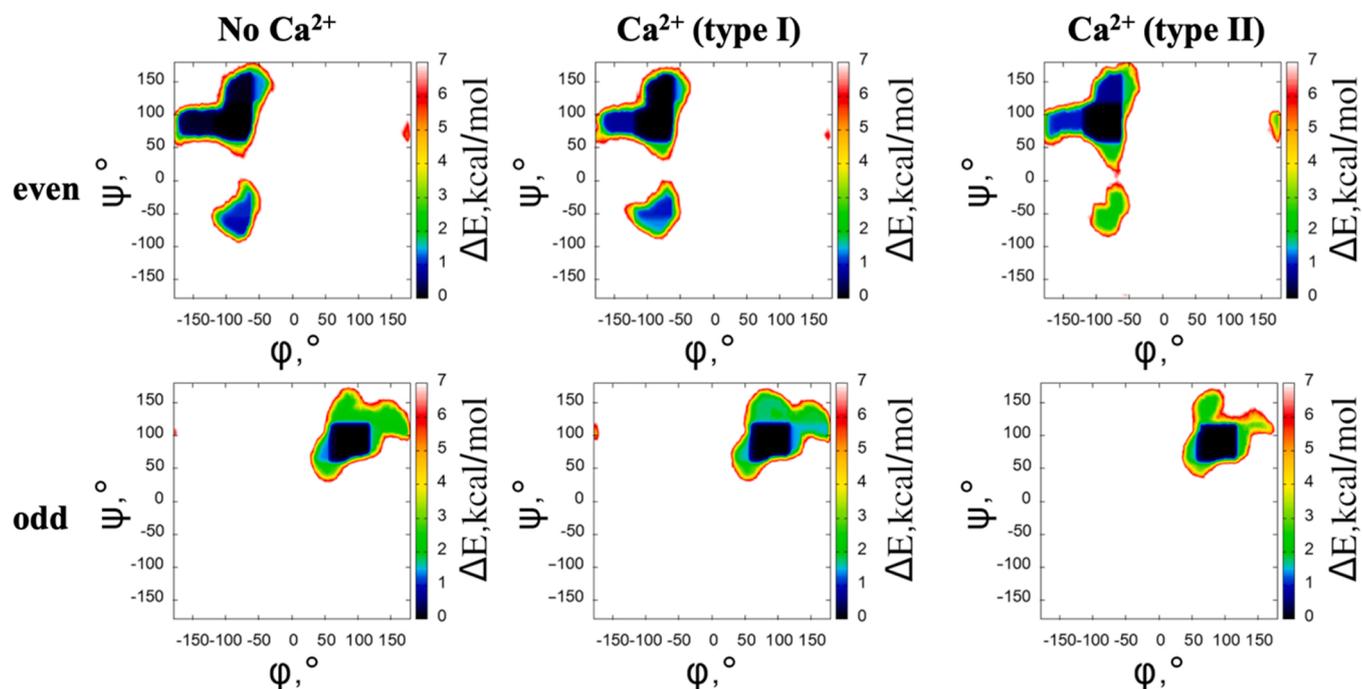


Fig. 5. Glycosidic linkage heatmaps for ϕ and ψ dihedral angles of HP dp18. The data have been summed up for all GlcNS6S-IdoA2S (even) and all IdoA2S-GlcNS6S (odd) glycosidic linkages.

hydrogen bonds was represented as a fraction throughout the MD simulation (Fig. 7). Both HP dp10 and HP dp18 have a similar distribution in the absence or presence of Ca^{2+} (type I). However, there was a marked change in the pattern when Ca^{2+} (type II) was used. Our results show that Ca^{2+} (type II) contributed to the formation of a higher number of intramolecular hydrogen bonds, shifting its distribution maximum from 5 to 6 hydrogen bonds formed for HP dp10 and from 9 to 11 for HP dp18 in comparison to Ca^{2+} (type I), respectively. The higher fraction of

intramolecular hydrogen bonds formed suggests that HP oligosaccharides adopt a more compact conformation as it was similarly observed by MD in chondroitin sulfate (Fadda and Woods, 2010). The increase in the number of established intramolecular hydrogen bonds could also reflect a decrease in HP oligosaccharide flexibility in the presence of Ca^{2+} . Furthermore, the formation of hydrogen bonds between pairs of HP monosaccharide residues was normalized by the length of each simulation and summarized in Fig. 8. These data also reveal clear differences

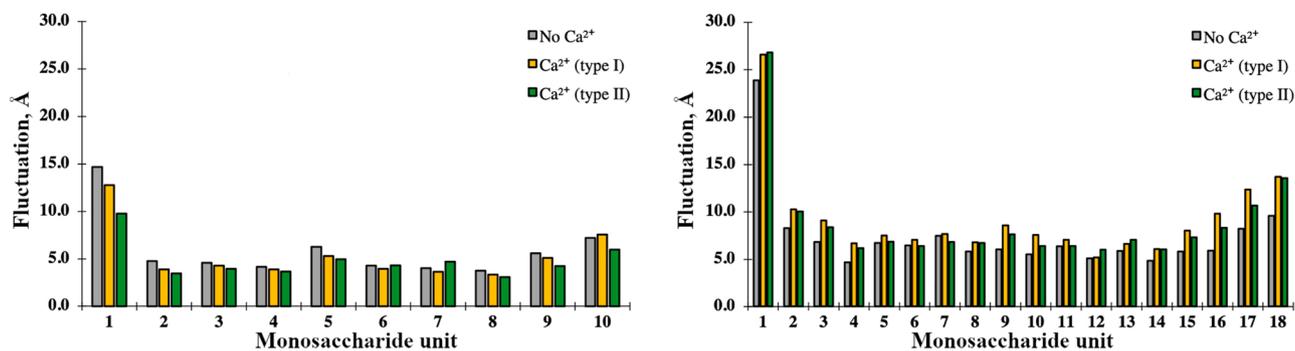


Fig. 6. Root mean squared fluctuations of monosaccharide unit in HP dp10 (left) and HP dp18 (right) in the absence and presence of each Ca^{2+} type.

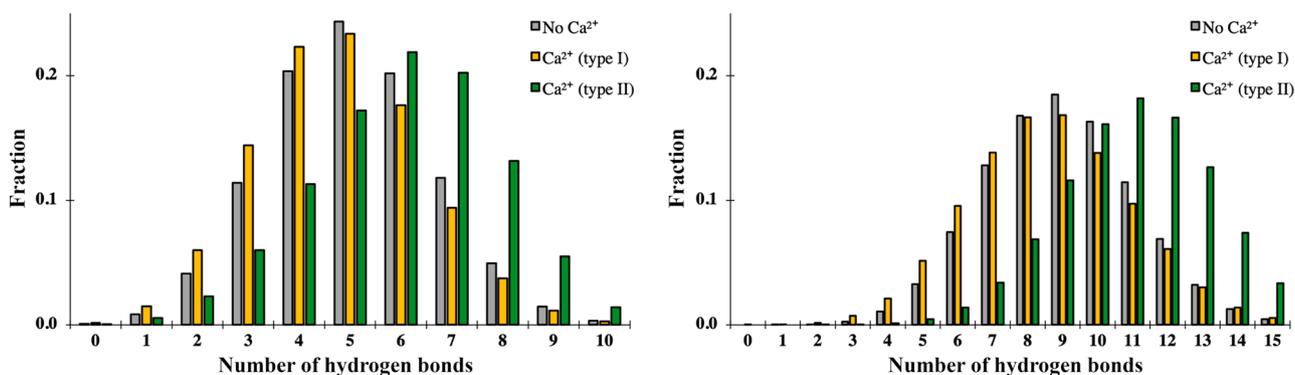


Fig. 7. Number of hydrogen bonds present in HP dp10 (left) and HP dp18 (right) in the absence and presence of each Ca^{2+} type as the fraction of MD trajectory frames in which this number of hydrogen bonds was formed.

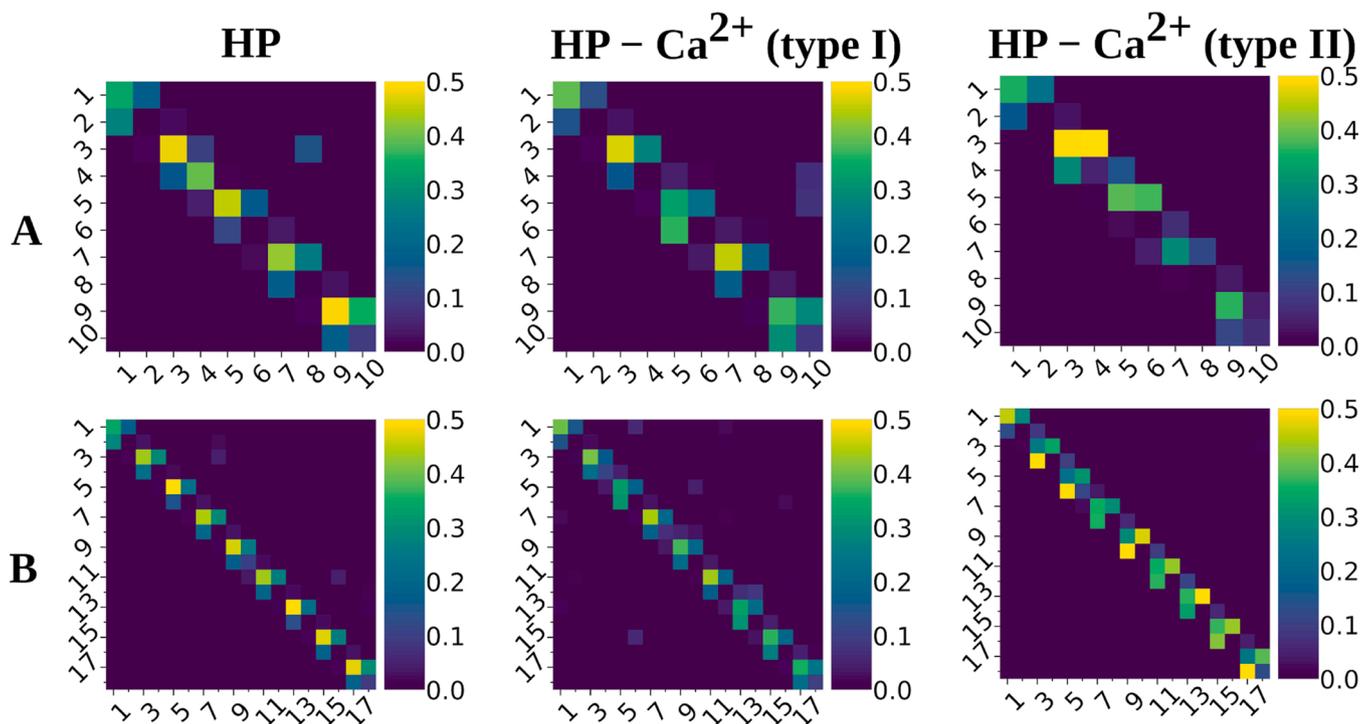


Fig. 8. The mean occupancy of intramolecular hydrogen bonds formed during MD simulations for HP dp10 (A) and HP dp18 (B) in the absence and presence of Ca^{2+} . The occupancy (increasing from blue to yellow) is computed as the fraction of MD trajectory frames in which a hydrogen bond was formed between the given residues as calculated by CPPTRAJ.

for the Ca^{2+} types used.

3.6. Free energy components

Thermodynamic properties of the HP molecules were analysed in the presence of Ca^{2+} of both types to estimate the impact of Ca^{2+} model on the free energy components and the conformational ensemble of the HP molecules. The values for free energies and entropy obtained at 300 K are summarized in Table S4. Total free energy and QH entropy values differed in the presence of Ca^{2+} types I and II for both HP dp10 and HP dp18. Differences between Ca^{2+} (type I) and Ca^{2+} (type II) were observed for ΔG_{vdw} , ΔG_{elec} , ΔG_{GB} , and ΔG_{SURF} in both HP dp10 and HP dp18. The absolute values of ΔG_{elec} and ΔG_{GB} were significantly lower and higher for Ca^{2+} (type I) in comparison to Ca^{2+} (type II), respectively, suggesting fundamentally distinct electrostatic screening effects within the Generalized Born implicit solvent model originated from the ion's parameters as charge density distribution and ion geometry/size. In the Generalized Born model, the electrostatic potential is calculated in grid points around the solute aiming to reproduce the effect of the solvent implicitly. Therefore, both the changes in the spacial charge distribution of the solute as well as of the space occupied by the ion and, therefore, inaccessible for the implicit solvent (corresponding grid points) essentially affect the electrostatic potential calculated in the proximity of the solute. These data suggest that the type of Ca^{2+} used in the simulation also substantially affects MM-GBSA free energy and entropy calculation results.

3.7. Coordination of Ca^{2+} to HP

We monitored the distance between Ca^{2+} and potential coordinating groups of HP: 2 sulfate groups (NS and 6S) from GlcNS6S and one sulfate group (S) and one carboxylate group (C6) from IdoA2S sugar unit during the MD simulations. The findings are summarized in Fig. 9. There was a clear difference in coordination depending on which type of Ca^{2+} was used. In case of Ca^{2+} (type I) the coordination by a carboxylate group was preferred as the overall probability of Ca^{2+} binding to C6 is twice as much as to the other HP coordinating groups. In contrast, for Ca^{2+} (type II) we did not observe any preferred coordination suggesting that Ca^{2+} (type II) had equal probability to be coordinated by one of HP negatively charged groups. In case of Ca^{2+} (type II), ions were coordinated substantially longer than when Ca^{2+} (type I) was used.

4. Conclusion

This work aimed at analysing the impact of Ca^{2+} on conformational and dynamic properties of two HP oligosaccharides of two different lengths using MD-based approaches. HP sequences represent major

building blocks of long heparan sulfate molecules, which have high biological relevance in many processes critical for the proper functioning of the extracellular matrix. The in-depth computational analysis using two parameter sets for Ca^{2+} focused on the R_g , ring puckering, glycosidic linkages, intramolecular hydrogen bonding and thermodynamic parameters of the HP molecules was performed for the trajectories obtained from μs -scale MD simulations. MD-based analysis allowed to conclude that the flexibility of the monosaccharides, the glycosidic linkages and ring puckering were not significantly affected by the presence of Ca^{2+} , in contrast to H-bond propensities and the theoretically determined R_g for a fraction of the oligosaccharide populations in both HP dp10 and dp18. Our findings also suggest the crucial importance of the Ca^{2+} parameters for modelling protein-HP interactions involving Ca^{2+} and shows the limitation of the currently available parameters to model binary (GAG-ion) and ternary (protein-ion-GAG) complexes. We did not dissect highly specific interactions with either type of Ca^{2+} which could be partially attributed to the persisting fundamental challenges in the ion parameterization. At the same time, we observed that thanks to their divalent nature, Ca^{2+} are able to decrease the length of HP molecules. This increased compactness in the presence of Ca^{2+} is caused by the bridging between the negatively charged groups of the HP and the ion, which is a dynamic process. In contrast, Na^+ and other potentially other monovalent ions are not charged enough to bring the negative groups together and stabilize them essentially. Moreover, our work clearly indicates the need for more advanced and physically more appropriate models for the analysis of these complexes, where electrostatic effects are crucial: in particular, the use of polarizable water models could be highly beneficial for achieving an improved description of these molecular systems. Increased amount and accuracy of structural data as well as further computational and experimental analyses of such molecular systems are needed to draw firm conclusions about the most appropriate divalent ion parameters for modelling ion-dependent protein-GAG interactions.

CRedit authorship contribution statement

Małgorzata M. Kogut: Conceptualization, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Annemarie Danielsson:** Formal analysis, Investigation, Methodology, Visualization. **Sylvie Ricard-Blum:** Conceptualization, Writing – original draft, Writing – review & editing. **Sergey A. Samsonov:** Conceptualization, Investigation, Funding acquisition, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing.

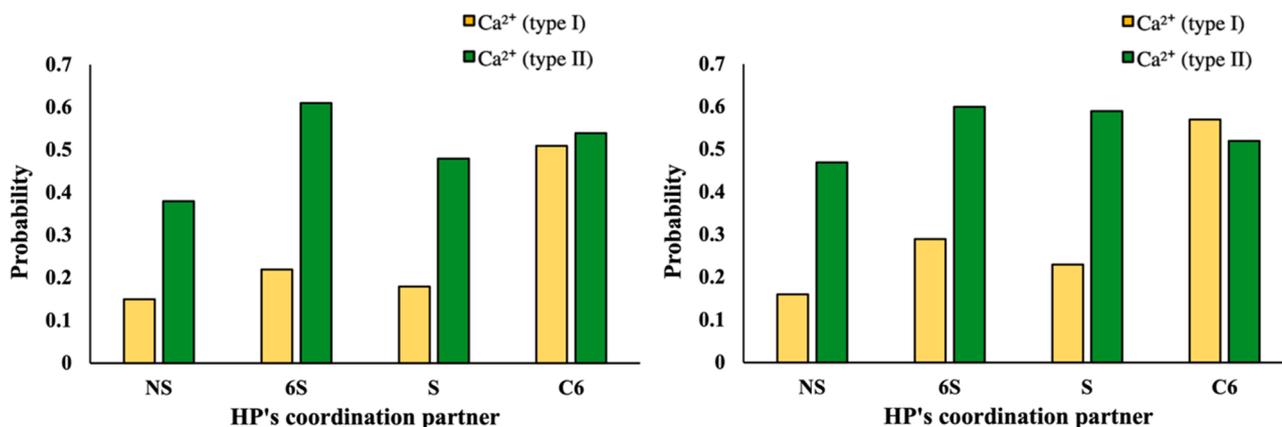


Fig. 9. Probability of Ca^{2+} (type I) and (type II) to be coordinated by negatively charged chemical groups of HP dp10 (left) and HP dp18 (right).

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.

Acknowledgements

This research was funded by National Science Centre of Poland, grant number UMO-2018/30/E/ST4/00037. The computational resources were provided by the Polish Grid Infrastructure (PL-GRID, grant plggagstr2gpu) and the cluster at the Faculty of Chemistry, University of Gdansk.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.compbiolchem.2022.107727.

References

- Almond, A., 2018. Multiscale modeling of glycosaminoglycan structure and dynamics: current methods and challenges. *Curr. Opin. Struct. Biol.* 50, 58–64.
- Bojarski, K.K., Sieradzka, A.K., Samsonov, S.A., 2019. Molecular dynamics insights into protein-glycosaminoglycan systems from microsecond-scale simulations. *Biopolymers* 110 (7), e23252.
- Capila, I., Hernaiz, M.J., Mo, Y.D., Mealy, T.R., Campos, B., Deadman, J.R., Linhardt, R. J., Seaton, B.A., 2001. Annexin V-heparin oligosaccharide complex suggests heparan sulphate-mediated assembly on cell surfaces. *Structure* 9, 57–64.
- Case, D.A.; Cerutti, D.S.; III Cheatham, T.E.; Darden, T.A.; Duke, R.E.; Giese, T.J.; Gohlke, H.; Goetz, A.W.; Greene, D.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Lee, T.S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D.; Merz, K.M.; Monard, G.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D.R.; Roitberg, A.; Sagui, C.; Simmerling, C.L.; Botello-Smith, W.M.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R.M.; Wu, R.X.; Xiao, L.; York, D.M.; Kollman, P.A.; AMBER 2017; University of California: San Francisco, 2017.
- Chang, C.E., Chen, W., Gilson, M.K., 2005. Evaluating the accuracy of the quasiharmonic approximation. *J. Chem. Theory Comput.* 1, 1017–1028.
- Cremer, D., Pople, J.A., 1975. General definition of ring puckering coordinates. *J. Am. Chem. Soc.* 97, 1354–1358.
- Das, S.K., Mallet, J.M., Esnault, J., Driguez, P.A., Duchaussoy, P., Sizun, P., Herault, J.P., Herbert, J.M., Petitou, M., Sinay, P., 2001. Synthesis of conformationally locked carbohydrates: a skew-boat conformation of L-iduronic acid governs the antithrombotic activity of heparin. *Angew. Chem. Int. Ed.* 40 (9), 1670–1673.
- Eckert, R., Raag, H., 2003. Zinc ions promote the interaction between heparin and heparin cofactor II. *FEBS Lett.* 541, 121–125.
- Fadda, E., Woods, R.J., 2010. Molecular simulations of carbohydrates and protein-carbohydrate interactions: motivation, issues and prospects. *Drug Discov. Today* 15, 596–609.
- Gandhi, N.S., Mancera, R.L., 2009. Free energy calculations of glycosaminoglycan-protein interactions. *Glycobiology* 19 (10), 1103–1115.
- Genheden, S., Ryde, U., 2010. How to obtain statistically converged MM/GBSA results. *J. Comput. Chem.* 31 (4), 837–846.
- Genheden, S., Kuhn, O., Mikulskis, P., Hoffmann, D., Ryde, U., 2012. The normal-mode entropy in the MM/GBSA method: effect of system truncation, buffer region, and dielectric constant. *J. Chem. Inf. Model.* 52, 2079–2088.
- Guevench, O., Whitmore, E.K., 2021. Sulfation and calcium favor compact conformations of chondroitin in aqueous solutions. *ACS Omega* 6 (20), 13204–13217.
- Habuchi, H., Habuchi, O., Kimata, K., 2004. Sulfation pattern in glycosaminoglycan: does it have a code? *Glycoconj. J.* 21, 47–52.
- Han, Q., Fu, Y., Zhou, H., He, Y., Lou, Y., 2007. Contributions of Zn (II)-binding to the structural stability of endostatin. *FEBS Lett.* 581, 3027–3032.
- 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 Río, J.F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., Oliphant, T.E., 2020. Array programming with NumPy. *Nature* 585 (7825), 357–362.
- Homeyer, N., Gohlke, H., 2012. Free energy calculations by the molecular mechanics Poisson-Boltzmann surface area method. *Mol. Inform.* 31 (2), 114–122.
- Hricovini, M., Guerrini, M., Bisio, A., Torri, G., Petitou, M., Casu, B., 2001. Conformation of heparin pentasaccharide bound to antithrombin III. *Biochem. J.* 359 (2), 265–272.
- Huige, C.J.M., Altona, C., 1995. Force field parameters for sulphates and sulfamates based on ab initio calculations: extensions of AMBER and CHARMM fields. *J. Comput. Chem.* 16, 56–79.
- Humphrey, W., Dalke, A., Schulten, K., 1996. VMD: visual molecular dynamics. *J. Mol. Graph.* 14, 33–38.
- Hunter, J.D., 2007. Matplotlib: A 2D graphics environment IEEE. *Ann. Hist. Comput.* 9, 90–95.
- Janke, J.J., Yu, Y., Pomin, V.H., Zhao, J., Wang, C., Linhardt, R.J., García, A.E., 2022. Characterization of heparin's conformational ensemble by molecular dynamics simulations and nuclear magnetic resonance spectroscopy. *Chem. Theory Comput.* <https://doi.org/10.1021/acs.jctc.1c00760> (Article ASAP).
- Karamanos, N.K., Piperigkou, Z., Theocharis, A.D., Watanabe, H., Franchi, M., Baud, S., Brézillon, S., Götte, M., Passi, A., Vignetti, D., Ricard-Blum, S., Sanderson, R.D., Neill, T., Iozzo, R.V., 2018. Proteoglycan chemical diversity drives multifunctional cell regulation and therapeutics. *Chem. Rev.* 118 (18), 9152–9232.
- Karplus, M., Kushick, J.N., 1981. Method for estimating the configurational entropy of macromolecules. *Macromolecules* 14, 325–332.
- Kirschner, K.N., Yongye, A.B., Tschampel, S.M., González-Outeiriño, J., Daniels, C.R., Foley, B.L., Woods, R.J., 2008. GLYCAM06: a generalizable biomolecular force field. *Carbohydr. J. Comput. Chem.* 29, 622–655.
- Kjellén, L., Lindahl, U., 2018. Specificity of glycosaminoglycan-protein interactions. *Curr. Opin. Struct. Biol.* 50, 101–108.
- 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.* 23, 3519–3530.
- Kogut, M.M., Marcisz, M., Samsonov, S.A., 2022. Modeling glycosaminoglycan-protein complexes. *Curr. Opin. Struct. Biol.* 73, 1023–1032.
- Lazar, R.D., Akher, F.B., Ravenscroft, N., Kuttel, M.M., 2022. Carbohydrate force fields: the role of small partial atomic charges in preventing conformational collapse. *J. Chem. Theory Comput.* 18 (2), 1156–1172.
- Li, P., Merz, K.M., 2017. Metal ion modeling using classical mechanics. *Chem. Rev.* 117 (3), 1564–1686.
- Lindahl, U., Couchman, J., Kimata, K., Esko, J.D., 2015–2017. Chapter 17. In: Varki, A., Cummings, R.D., Esko, J.D. (Eds.), *Essentials of Glycobiology*, third ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.
- Lu, L.D., Shie, C.R., Kulkarni, S.S., Pan, G.R., Lu, X.A., Hung, S.C., 2006. Synthesis of 48 disaccharide building blocks for the assembly of a heparin and heparan sulphate oligosaccharide library. *Org. Lett.* 8 (26), 5995–5998.
- McKinney, W. Data structures for statistical computing in python, in: *Proc. 9th Python Sci. Conf.*, Austin, TX, 2010, 51–56.
- Miranda-Quintana, R.A., Smiatek, J., 2021. Specific ion effects in different media: current status and future challenges. *J. Phys. Chem. B.* 125 (51), 13840–13849.
- Multhaup, G., Bush, A.I., Pollwein, P., Masters, C.L., 1994. Interaction between the zinc (II) and the heparin binding site of the Alzheimer's disease beta A4 amyloid precursor protein (APP). *FEBS Lett.* 335, 151–154.
- Nagarajan, B., Sankaranarayanan, N.V., Desai, U.R., 2019. Perspective on computational simulations of glycosaminoglycans. *WIREs Comp. Mol. Sci.* 9, e1388.
- Nagarajan, B., Holmes, S.G., Sankaranarayanan, N.V., Desai, U.R., 2022. Protein-carbohydrate complexes and glycosylation. *Curr. Opin. Struct. Biol.* 74, 1023–1056.
- Onufriev, A., Bashford, D., Case, D.A., 2000. Modification of the generalized born model suitable for macromolecules. *J. Phys. Chem. B* 104 (15), 3712–3720.
- Perez, S., Bonnardel, F., Lisacek, F., Imberty, F.A., Ricard-Blum, S., Makashova, O., 2020. GAG-DB, the new interface of the three-dimensional landscape of glycosaminoglycans. *Biomolecules* 10 (12), 1660.
- Pichert, A., Samsonov, S.A., Theissen, S., Thomas, L., Baumann, L., Schiller, J., Beck-Sickinge, 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 (1), 134–145.
- Pothoff, J., Bojarski, K.K., Kohut, G., Lipska, A.G., Liwo, A., Kessler, E., Ricard-Blum, S., Samsonov, S.A., 2019. Analysis of procollagen C-Proteinase enhancer-1/ glycosaminoglycan binding sites and of the potential role of calcium ions in the interaction. *Int. J. Mol. Sci.* 20 (20), 5021.
- Ricard-Blum, S., Féraud, O., Lortat-Jacob, H., Rencurosi, A., Fukai, N., Dkhissi, F., Vittet, D., Imberty, A., Olsen, B.R., van der Rest, M., 2004. Characterization of endostatin binding to heparin and heparan sulphate by surface plasmon resonance and molecular modeling: role of divalent cations. *J. Biol. Chem.* 279 (4), 2927–2936.
- Roe, D.R., Cheatham III, T.E., 2013. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* 9, 3084–3095.
- Samsonov, S.A., Teyra, J., Pisabarro, M.T., 2011a. Computational analysis of solvent inclusion in docking studies of protein-glycosaminoglycan systems. *J. Comput. Aided Mol. Des.* 25, 477.
- Samsonov, S.A., Teyra, J., Pisabarro, M.T., 2011b. Docking glycosaminoglycans to proteins: analysis of solvent inclusion. *J. Comput. Aided Mol. Des.* 25 (5), 477–489.
- Sankaranarayanan, N.V., Nagarajan, B., Desai, U.R., 2018. So you think computational approaches to understanding glycosaminoglycan-protein interactions are too dry and too rigid? Think again! *Curr. Opin. Struct. Biol.* 50, 91–100.
- Sasaki, T., Larsson, H., Kreuger, J., Salmivirta, M., Claesson-Welsh, L., Lindahl, U., Hohenester, E., Timpl, R., 1999. Structural basis and potential role of heparin/heparan sulphate binding to the angiogenesis inhibitor endostatin. *EMBO J.* 18 (22), 6240–6248.
- Sattelle, B.M., Almond, A., 2011. Is N-acetyl-D-glucosamine a rigid 4C1 chair? *Glycobiology* 21 (12), 1651–1662.
- Sattelle, B.M., Almond, A., 2012. Assigning kinetic 3D-signatures to glycocones. *Phys. Chem. Chem. Phys.* 16, 5843–5848.
- Sattelle, B.M., Hansen, S.U., Gardiner, J., Almond, A., 2010. Free energy landscapes of iduronic acid and related monosaccharides. *J. Am. Chem. Soc.* 132 (38), 13132–13134.
- Sattelle, B.M., Bose-Basu, B., Tessier, M., Woods, R.J., Serianni, A.S., Almond, A., 2012. Dependence of pyranose ring puckering on anomeric configuration: methyl idopyranosides. *J. Phys. Chem. B* 116 (112), 6380–6386.

- Sattelle, B.M., Shakeri, J., Almond, A., 2013. Does microsecond sugar ring flexing encode 3D-shape and bioactivity in the heparanome? *Biomacromolecules* 14 (4), 1149–1159.
- Sattelle, B.M., Shakeri, J., Cliff, M.J., Almond, A., 2015. Proteoglycans and their heterogenous glycosaminoglycans at the atomic scale. *Biomacromolecules* 16 (3), 951–961.
- Saxena, A., Sept, D., 2013. Multisite ion models that improve coordination and free energy calculations in molecular dynamics simulations. *J. Chem. Theory Comput.* 9 (8), 3538–3542.
- Shao, C., Zhang, F., Kemp, M.M., Linhardt, R.J., Waisman, D., Head, J.F., Seaton, B.A., 2006. Crystallographic analysis of calcium-dependent heparin binding to annexin A2. *J. Biol. Chem.* 281, 31689–31695.
- Shriver, Z., Capila, I., Venkataraman, G., Sasisekharan, R., 2012. Heparin and heparan sulphate: analyzing structure and microheterogeneity. *Handb. Exp. Pharmacol.* 207, 159–176.
- Srinivasan, J., Cheatham III, T.E., Cieplak, P., Kollman, P.A., Case, D.A., 1998. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate–DNA helices. *J. Am. Chem. Soc.* 9401–9409.
- Stevic, I., Parmar, N., Paredes, N., Berry, L.R., Chan, A.K.C., 2011. Binding of heparin to metals. *Cell Biochem Biophys.* 59, 171–178.
- Team, R.C.R.: *A language and environment for statistical computing*, (2013).
- Vallet, S.D., Clerc, O., Ricard-Blum, S., 2021. Glycosaminoglycan-protein interactions: the first draft of the glycosaminoglycan interactome. *J. Histochem Cytochem.* 69 (2), 93–104.
- Weiss, T., Ricard-Blum, S., Moschovich, L., Wineman, E., Mesilaty, S., Kessler, E., 2010. Binding of procollagen C-Proteinase Enhancer-1 (PCPE-1) to heparin/heparan sulfate. *J. Biol. Chem.* 285 (44), 33867–33874.
- Williams, T.; Kelley, C. *Gnuplot 4.6: An Interactive Plotting Program*, (<http://gnuplot.sourceforge.net/>) (2013). Accessed date: December 2019.
- Woodhead, N.E., Long, W.F., Williamson, F.B., 1986. Binding of zinc ions to heparin. Analysis by equilibrium dialysis suggests the occurrence of two, entropy-driven, processes. *Biochem. J.* 237, 281–284.
- Zhang, F., Liang, X., Beaudet, J.M., Lee, Y., Linhardt, R.J., 2014. The effects of metal ions on heparin/heparin sulphate-protein interactions. *J. Biom. Tech. Res* 1, 1–7.
- Zhang, F., Zhao, J., Liu, X., 2020. Interactions between sclerostin and glycosaminoglycans. *Glycoconj. J.* 37, 119–128.

PUBLICATION D3

**AFFINITY AND PUTATIVE ENTRANCE MECHANISMS OF
ALKYL SULFATES INTO THE β -CD CAVITY**



Affinity and putative entrance mechanisms of alkyl sulfates into the β -CD cavity



Małgorzata M. Kogut¹, Ola Grabowska¹, Dariusz Wyrzykowski^{*}, Sergey A. Samsonov^{*}

Faculty of Chemistry, University of Gdańsk, Wita Stwosza 63, 80-308 Gdańsk, Poland

ARTICLE INFO

Article history:

Received 13 June 2022

Revised 26 July 2022

Accepted 28 July 2022

Available online 1 August 2022

ABSTRACT

Cyclodextrins (CDs) are novel pharmaceutical excipients and are widely used in the pharmaceutical industry and drug delivery due to their ability to form host–guest complexes. There are countless structural features of guest molecules responsible for the efficiency of guest–CD interactions. Among these factors, the hydrophobic moieties of a guest and their size are commonly considered to govern its binding affinity to CD molecules. Experimentally, it is often challenging to extract structural and dynamic information at the atomic scale about CD inclusion complexes, especially in the aqueous phase. Therefore, this study consisted of experimental (isothermal titration calorimetry and conductometric titration) and *in silico* analysis to rigorously characterize the interactions of β -CD with three alkyl sulfates of different hydrophobic chain lengths: sodium dodecyl sulfate (S12S), sodium decyl sulfate (S10S) and sodium octyl sulfate (S8S) in aqueous solutions. We find that the hydrophobic interactions are not always a key factor leading to the formation of stable inclusion complexes. Our results demonstrate that van der Waals interactions contribute more to complexation efficiency for sulfates with a longer hydrophobic tail, whereas electrostatic interactions are more pronounced for sulfates with shorter ones. Furthermore, we propose and discuss for the first time the putative mechanisms of the guest entering the host cavity conditioned by the lengths of hydrocarbon chains of alkyl sulfates. Finally, we prove that different types of inclusion complexes can exist in a conformational equilibrium depending on the size of a guest tail.

© 2022 Elsevier B.V. All rights reserved.

1. Introduction

Cyclodextrins (CDs) are well-known excipients with applications ranging from life sciences to innovative technologies [1–10], and recently published reviews reflect their importance [11–17]. CDs belong to a class of cyclic oligosaccharides, the most common ones being α , β and γ -CDs formed by six, seven and eight α -(1–4) D-glucopyranoses units, respectively. The number of these units in the cyclic oligosaccharides determines the size of the cavity. The specific coupling regarding the glycosidic linkage type of these glucose monomers is responsible for a unique molecular structure of CDs with a hydrophilic exterior and hollow hydrophobic interior [3,18]. This distinctive structure of CDs is at the origin of their peculiar physicochemical properties, particularly their ability to form host–guest complexes. Due to the formation of these inclusion complexes, the guest molecules' biological, chemical, and physical properties can be altered. CDs have been used for pharmaceutical applications, mainly as solubilizers and stabilizers for hydrophobic

or biological drugs [19,20]. Therefore, investigations of the molecular recognition between CD (the host) and the rich diversity of guest species, namely low-molecular-weight compounds such as potential drugs, natural metabolites etc., are the subject of the ongoing studies. Of great interest are the structural factors governing the affinity of a guest to hosts.

Considering the specific structural features of CDs, the size and specific properties of the guests' hydrophobic and/or hydrophilic moiety play a crucial role in the stability of the resulting inclusion CD-complexes. Among guest molecules, surfactants are an important group of compounds for both a fundamental understanding of these systems and their practical applications [16]. In addition, they can be used as models for studying the relationship between the structure of a guest molecule and the stability of the resulting CD-complexes. To get better insights into the host–guest interactions numerous experimental and theoretical methods are employed. Much attention has been paid to the synthesis and structural characterization of CD complexes. Among these, there are 32, 112 and 6 inclusion complexes for α -CD, β -CD and γ -CD, respectively, which were identified by the X-ray diffraction technique (R-factor ≤ 0.075) [21]. According to the CCDC, the number of deposited β -CD-(guest) structures is much larger in comparison to the complexes with α -CD and γ -CD hosts due to the relatively low cost of the CD

^{*} Corresponding authors.

E-mail addresses: dariusz.wyrzykowski@ug.edu.pl (D. Wyrzykowski), sergey.samsonov@ug.edu.pl (S.A. Samsonov).

¹ Authors contributed equally.

production and the size of the β -CD cavity into which many guest fit. However, there are no reports on the crystal structures of CD-surfactant complexes. The examination of the nature and magnitude of the driving forces in a solution will provide useful information regarding the strategy of designing new, appropriate guests of desired properties. The characterization of CD inclusion complexes formation is not always a simple task and often requires different analytical techniques depending on the physical and chemical properties of the guest. The review by Mura portrays the most conventional analytical approaches for studying the CD-guest interactions in an aqueous solution, emphasizing their potential merits as well as disadvantages and limitations [22]. In the case of ionic surfactants as guest molecules, there are two main analytical tools to characterize the CD-surfactant interactions in solution [23]. The first one is conductometric titration successfully used to evaluate a binding constant and the critical micelles concentration (cmc) [24–26]. The second one is isothermal titration calorimetry (ITC) more accurate quantitative technique than electric conductivity. ITC allows the direct determination of stoichiometry, a binding constant and a binding enthalpy, and is considered a powerful tool.

Molecular dynamics (MD) simulations are becoming invaluable tools for studying biomolecular structure and dynamics [27,28]. They enable the determination of structural, energetic, and thermodynamical properties of molecular systems [29]. Since the early 2010 s, there has been a rapid increase in the number of works published concerning the substantial advantages of deploying MD simulations in the CD host-guest complexes, followed by post-MD calculations. Furthermore, the latest applications of umbrella sampling (US) proved that state-of-the-art methods of MD could be beneficial in studying and designing CD complexes. Mazurek et al. provided a comprehensive review of the current developments in the applications of MD simulations in the analysis of CDs [30]. Recently, we have demonstrated that MD simulations followed by binding free energy calculations are an effective way to predict, identify and rank the potential binding sites in protein-small molecule complexes [31,32].

In this paper, we report the results of calorimetric (ITC) and conductometric studies supported by *in silico* analysis on the interactions of β -CD with three anionic surfactants of different alkyl chain lengths, namely sodium dodecyl sulfate (S12S), sodium decyl sulfate (S10S) and sodium octyl sulfate (S8S) (Fig. 1.). The selected surfactants enabled to assess the effect of the length of the hydrophobic carbon chain of the guest on the stability of inclusion complexes and the binding mode of the guest molecules. The presented findings can broaden our knowledge of how the size of hydrophobic guest (drug) moiety influences the efficiency of complex formation, and thus the drug delivery potential of different lipophilicity as β -CD complexes.

2. Materials and methods

2.1. Materials

β -Cyclodextrin (β -CD, $\geq 97\%$, CAS: 7585-39-9), sodium dodecyl sulfate (S12S, CAS: 151-21-3, $\geq 99\%$), sodium decyl sulfate (S10S, CAS: 142-87-0, 98%), sodium octyl sulfate (S8S, CAS: 142-31-4, $\geq 95\%$) were purchased from Sigma Aldrich (Poland) and used without further purification. Double-distilled water (purified by Hydrolab-system) with conductivity not exceeding $0.18 \mu\text{S cm}^{-1}$ was used for preparations of aqueous solutions.

2.2. Conductometric measurements

Conductometric titration experiments were performed in water at 298.15 ± 0.01 K, in a 30-mL thermostated cell, using the Cerko

Lab System microtitration unit fitted with the 5-mL Hamilton's syringe and the conductometric sensor (CD-201, Hydromet, Poland). The sensor was calibrated with conductivity standards (aqueous KCl solutions) of conductivity $84 \mu\text{S/cm}$ and $200 \mu\text{S/cm}$, purchased from Hamilton Company. The titrant (4 mM β -Cyclodextrin solution) was added to the titrand (0.2 mM surfactant solution: S12S, S10S or S8S) in increments of 0.02 mL and a pose 60 sec. Each titration was repeated a least twice in order to check the reproducibility of the data.

2.3. Isothermal titration calorimetry

All ITC experiments were performed at 298.15 K using the Auto-ITC isothermal titration calorimeter (MicroCal Inc. GE Healthcare, Northampton, USA). The details of the measuring devices and experimental setup were described previously [33]. The reagents were dissolved directly into H_2O . The experiment consisted of injecting $10.02 \mu\text{L}$ (29 injections, $2 \mu\text{L}$ for the first injection only, injection duration: 20 sec., injection interval: 240 sec.) of 4 mM β -cyclodextrin solution into the reaction cell which initially contained 0.2 mM surfactant solution (S12S, S10S or S8S). For each experiment, a blank was performed by injecting the titrant solution (β -CD) into the cell filled with water only. This blank was subtracted from the corresponding titration to account for the heat of a dilution.

2.4. Structures

The structures of β -CD-host with S12S, S10S and S8S guests were built in Xleap module of AMBER16 software package [34]. The structure of S12S was used from our previous work [31] and based on it S10S and S8S structures were prepared by removing $-\text{C}_2\text{H}_5$ and $-\text{C}_4\text{H}_9$ from S12S, respectively. The initial complexes were built by placing a guest molecule close but in a random orientation to the host molecule.

2.5. Electrostatic potential calculations

To calculate the electrostatic potential isosurfaces of β -CD, a Poisson Boltzmann surface area (PBSA) program from AmberTools [34] was used with a 1 \AA grid spacing step. The obtained electrostatic potential maps were visualized with VMD software [35].

2.6. Molecular dynamics simulations

All-atom Molecular Dynamics (MD) simulation for each host-guest system was performed with the use of AMBER16 software package [34]. Truncated octahedron TIP3P periodic box of 10 \AA water layer from box's border to solute was used to solvate complexes. Charge was neutralized with Na^+ counterions. GLYCAM06 force field for the β -CD-host [36] and gaff force field parameters [37] with RESP charges [38] obtained in antechamber module of AMBER16 [34] for the guests S8S, S10S and S12S were used, respectively. Energy minimization was carried out in two steps: beginning with 500 steepest descent cycles and 10^3 conjugate gradient cycles with $100 \text{ kcal/mol/\AA}^2$ harmonic force restraints, continued with 3×10^3 steepest descent cycles and 3×10^3 conjugate gradient cycles without any restraints. Following minimization steps, the system was heated up from 0 to 300 K for 10 ps with harmonic force restraints of $100 \text{ kcal/mol/\AA}^2$. Subsequently, the system was equilibrated at 300 K and 10^5 Pa in the isothermal isobaric ensemble for 500 ps. Afterwards, the productive MD run was carried out in the same isothermal isobaric ensemble for $1 \mu\text{s}$. The particle mesh Ewald method for treating electrostatics and SHAKE algorithm for all the covalent bonds containing hydrogen atoms were implemented in the MD simulations.

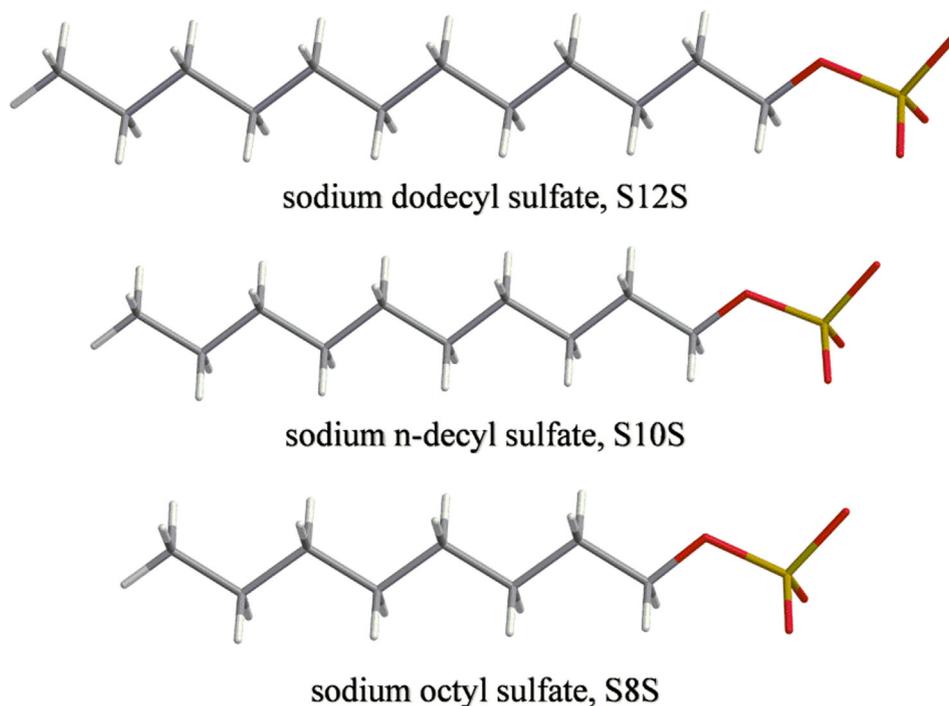


Fig. 1. Chemical structure of the investigated guest anions.

2.7. Binding free energy calculations

Energetic postprocessing of the trajectories and per-residue energy decomposition were performed for all systems with the use of Linear Interaction Energy (LIE) with a dielectric constant of 80 and Molecular Mechanics Generalized Born Surface Area (MM-GBSA) [39] using a model with surface area and Born radii default parameters as implemented in the $igb = 2$ model in AMBER16 [34]. MD frames corresponding to the complex in which structure reached convergence in terms of the RMSD were analysed.

2.8. Umbrella sampling and the potential of mean force

2.8.1. Preparation of US simulations

The initial structure of each complex for the US simulations was obtained from the MD trajectory frame, in which the lowest RMSD value to the average one of the guests throughout the simulation was observed. The distance between the C-end (carbon that is furthest from the S-atom, see Fig. 1) and the centre of mass of β -CD was measured and defined as a reaction coordinate. By using this bound conformation as a starting window, we gradually moved the guest along the positive (C-pull) and negative (S-pull) Y-axis with a 1.0 Å step for 50 Å in both directions, until the guest was unbound (was beyond the cut-off distance for non-bonded interactions). This way we defined the paths C-pull and S-pull for guest dissociation along the primary and secondary cavity of β -CD (Fig. 1).

2.8.2. Running US simulations

US was performed to compute the potential of mean force (PMF) along the dissociation pathway in AMBER 16 [40]. In US, a series of windows was evenly located along the reaction coordinate, and conformational sampling in these windows was achieved by enforcing an external biasing potential for 10 ns in each window. The sampling in each window overlapped with the ones in adjacent windows so that the unbiased PMF could be reproduced

by removing the biasing potential. The size of the periodic box was essentially increased in comparison to the previously described simulations to avoid potential artifacts in PMF calculations that could be caused by insufficient periodic box size in the course of the guest movement. Once all the US MD simulations were completed, the data collected from different simulation windows were combined along the reaction coordinate to calculate the PMF profile for the whole structural transition process with the Weighted Histogram Analysis Method (WHAM) [41] using the code developed by Alan Grossfield [42]. For each host-guest complex two US simulations were performed, along C-pull and S-pull pathways, making, in total, six US MD simulations. In WHAM, the tolerance of iteration was set to 0.001, and the temperature to 300 K. Finally, LIE was calculated for each window for each system for the energetically most favourable path.

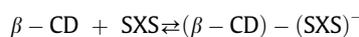
2.9. Data analysis and visualisation

Data analysis and its visualization were performed with the use of the R-package [43] and VMD [35].

3. Results and discussion

3.1. Equilibrium constants of complex formation – Conductometric studies

Equilibrium constants of the investigated (β -CD) – (SXS, X = 12, 10 and 8) complexes formation were calculated based on conductometric titration data. On the assumption a stoichiometry of the complexes under study equals 1:1 (host:guest), the binding constant K is expressed by the following equation:



$$K = \frac{[(\beta - CD - SXS)^{-}]}{[\beta - CD][SXS^{-}]}$$

where $[(\beta\text{-CD})\text{-(SXS)}]^-$, $[\beta\text{-CD}]$, $[\text{SXS}]^-$ are the concentrations of the complex, host and guest molecules in a solution in the equilibrium state, respectively. The binding constants K were refined from the equation [44]:

$$\Lambda_{\text{obs}} = [K \cdot (c_{\text{SXS}} - c_{\text{CD}}) - 1 + \sqrt{K^2 (c_{\beta} - c_{\text{SXS}})^2 + 2K(c_{\text{SXS}} + c_{\text{CD}}) + 1}] \left[\frac{\Lambda_f - \Lambda_c}{2K \cdot c_{\text{SXS}}} \right] + \Lambda_c$$

(where Λ_{obs} is the observed molar conductance of the titrated solution, Λ_f is the molar conductance of SXS in a pure solvent, Λ_c is the molar conductance of the $(\beta\text{-CD})\text{-(SXS)}$ complex, c_{CD} and c_{SXS} are the total concentration of the host and the guest, respectively) by least-squares calculations from conductometric data using the Origin 8.5 program (ver. 8.5). The adjustable parameters, namely K and Λ_c were collected in Table 1. The free energy of binding (ΔG) was calculated using the standard thermodynamic relationship: $\Delta G = -RT \ln K$ (R is the gas constant, and T , in degrees Kelvin, is the absolute temperature at which the interaction takes place). The changes of molar conductance (Λ_{obs}) vs the $\beta\text{-CD}$ to SXS molar ratio ($[\beta\text{-CD}]/[\text{SXS}]$) together with the theoretical curves calculated based on the parameters (K and Λ_c) collected in Table 1 are presented in Fig. 2. The molar conductance decreases slightly with the increase of $\beta\text{-CD}$ in the reaction mixture on account of the formation of the $\beta\text{-CD}\text{-(SXS)}$ complexes which are less effective as a charge carrier than the free (uncomplexed) SXS anions.

The conductometric data revealed that the thermodynamic stability of the resulting $\beta\text{-CD}\text{-(SXS)}$ complexes increases with the increase of the hydrophobic chain length of the SXS surfactant (Table 1):

$$\log K_{\text{S12S}} > \log K_{\text{S10S}} > \log K_{\text{S8S}}$$

$$|\Delta G_{\text{S12S}}| > |\Delta G_{\text{S10S}}| > |\Delta G_{\text{S8S}}|$$

A similar phenomenon has already been observed for $\beta\text{-CD}$ complexes with cationic single chain surfactants and bolaforms (the surfactants with two water-soluble heads connected by a hydrophobic spacer) [23]. The changes in the binding parameter ν defined as the moles of bound SXS by each mole of $\beta\text{-CD}$ ($\nu = \frac{[(\beta\text{-CD})\text{-(SXS)}]}{[\beta\text{-CD}] + [(\beta\text{-CD})\text{-(SXS)}]}$) as a function of the free SXS ($[\text{SXS}]$) were compared for the investigated $\beta\text{-CD}/\text{SXS}$ systems. Graphic representations of the simulated data for the $\beta\text{-CD}/(\text{SXS})$ binding equilibrium with the experimentally determined binding constants (Table 1) are presented in Fig. 3.

The binding constant (K) and calculated indirectly the free energy of binding ($\Delta G = f(K)$) are well-known parameters describing the affinity of a macromolecule to a ligand. However, these parameters provide only general information about the stability of the complex. The complex formation is accompanied by a variety of processes, such as conformational changes of the reactants, breaking and formation of hydrogen bonds, electrostatic and hydrophobic interactions as well as fluctuations in the reaction environment. For these reasons, the direct determination of binding parameters (ΔH , and indirectly ΔS) becomes necessary for describing the energetic aspects of the interactions as well as the nature and magnitude of the forces responsible for the mutual affinity of the host to the guest.

3.2. Thermodynamic parameters of the interactions – Calorimetric studies

Binding parameters K and ΔG can be determined with a variety of methods [22]. However, the key to quantifying the binding thermodynamics is measuring the binding enthalpy (ΔH). Among the diverse analytical methodologies, isothermal titration calorimetry (ITC) is considered one of the most useful techniques enabling direct

Table 1

Logarithms of the equilibrium constants ($\log K$), the molar conductance (Λ_c) of the complex and free energies of binding (ΔG) of the $\beta\text{-CD}$ complexes with sodium dodecyl sulfate (S12S), sodium decyl sulfate (S10S) and sodium octyl sulfate (S8S) in aqueous solutions, at 298.15 K (standard deviation values in parentheses).

Parameter	S12S	S10S	S8S
$\log K$	4.11 (± 0.03)	3.69 (± 0.05)	3.38 (± 0.02)
Λ_c [$\text{S cm}^2 \text{ mol}^{-1}$]	65.7 (± 0.2)	67.9 (± 0.6)	77.6 (± 0.2)
ΔG [kcal mol^{-1}]	-5.61 (± 0.05)	-5.04 (± 0.07)	-4.61 (± 0.03)

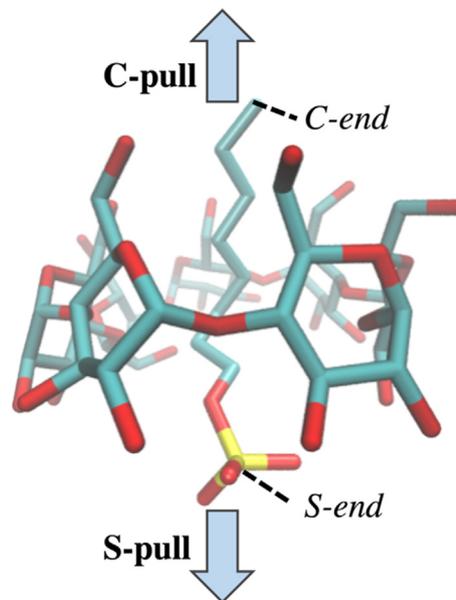


Fig. 2. Dissociation pathways of S8S from the $\beta\text{-CD}$ complex. The C-pull pathway is when the C-end of the guest is pulled and the S-pull pathway is when the S-end is pulled from the $\beta\text{-CD}$. For clarity, hydrogens are not shown.

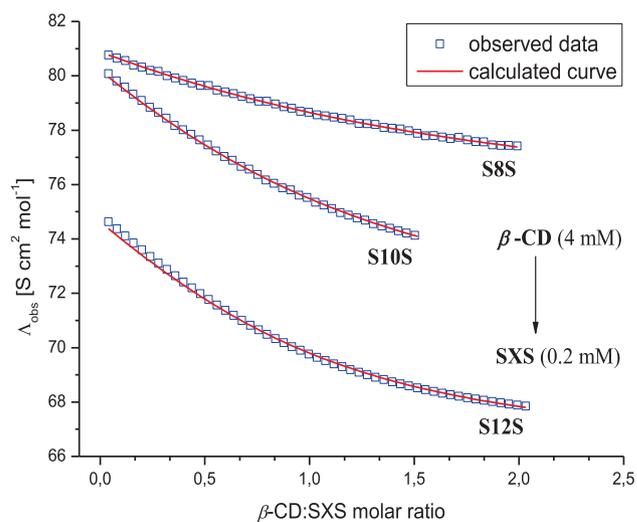


Fig. 3. Molar conductance (experimental data – blue squares) and (theoretical data – solid red lines) vs molar ratio $[\beta\text{-CD}]/[\text{SXS}]$ in an aqueous solution at 298.15 K.

calculation of ΔH_{ITC} based on the measurement of the heat released or absorbed upon a complex formation [45]. The knowledge of thermodynamic parameters of the reactions provides a complete thermodynamic characterization of the interactions under study, and thus enables a better understanding of the investigated processes than that of simple binding constants. Furthermore, the dependence of thermodynamic properties on structural detail is essential to comprehend how the interactions occur [46,47].

Representative binding isotherms for β -CD – SXS interactions in an aqueous solution are shown in Fig. 4, whereas thermodynamic parameters of the complex formation are summarized in Table 2. The stoichiometry of the inclusion complexes, the binding constant K_{ITC} and the binding enthalpy ΔH_{ITC} were obtained directly from the ITC experiments by fitting binding isotherms to the experimental data, using the non-linear fitting least-squares procedure and the equilibrium model that assumes a single set of identical binding sites. The free energy of binding ΔG_{ITC} and the entropy ΔS_{ITC} of the interaction were calculated using the relationships: $\Delta G_{ITC} = -RT \ln K_{ITC} = \Delta H_{ITC} - T \cdot \Delta S_{ITC}$.

The stoichiometry (N) of β -CD:SXS complexes determined directly from ITC data (Table 2) justifies the assumption taken during the analysis of conductometric data about the formation of 1:1 (the host to the guest) complexes. Furthermore, the binding constant K of the resulting β -CD complexes and thus the free energy of binding ΔG obtained based on conductometric and calorimetric data are equal, within the range of the experimental errors (Tables 1 and 2). The formation of the complexes is accompanied by the release of the heat. The binding enthalpy adopts slightly more negative values with the increase of the length of the hydrophobic tail of the guest. Simultaneously, the increase in the entropy change is observed. Among

Table 2

The thermodynamic parameters of sodium dodecyl sulfate (S12S), sodium decyl sulfate (S10S) and sodium octyl sulfate (S8S) binding to β -CD (standard deviation values in parentheses) in aqueous solutions, at 298.15 K.

ITC	S12S	S10S	S8S
N (the stoichiometry)	1.10 (± 0.03)	1.17 (± 0.07)	0.84 (± 0.06)
$\log K_{ITC}$	4.12 (± 0.03)	3.61 (± 0.04)	3.23 (± 0.03)
ΔG_{ITC} [kcal mol ⁻¹]	-5.62 (± 0.04)	-4.92 (± 0.05)	-4.41 (± 0.03)
ΔH_{ITC} [kcal mol ⁻¹]	-2.30 (± 0.07)	-2.13 (± 0.17)	-1.76 (± 0.16)
$T\Delta S_{ITC}$ [kcal mol ⁻¹]	3.32 (± 0.08)	2.79 (± 0.17)	2.65 (± 0.04)

different types of non-covalent forces, electrostatic and van der Waals interactions, as well as hydrophobic interactions, are involved in the interactions of β -CD with SXS. The former are enthalpy driven whereas the latter are entropy-driven. The inner volume of the β -CD cavity (270 Å³) fits best to accommodate the S12S chain with the methylene (-CH₂-) volume equal to 27 Å³. Consequently, due to the established essential contacts of the S12S tail with the CD cavity, van der Waals interactions seem to contribute significantly to the exothermic effect of the complex formation. Furthermore, a more rigid structure is expected for the shortest tail guest (S8S) on account of electrostatic interactions of a negatively charged head group (-OSO₃⁻) with primary or the secondary hydroxyl groups located on rims. It manifests itself in the decrease of translation and conformation freedom of SXS upon complexation that is reflected in less positive values of the entropic factors ($T\Delta S$) (Table 2). To get some insight into the factors underlying the stability of the investigated inclusion complexes and for a better understanding of the

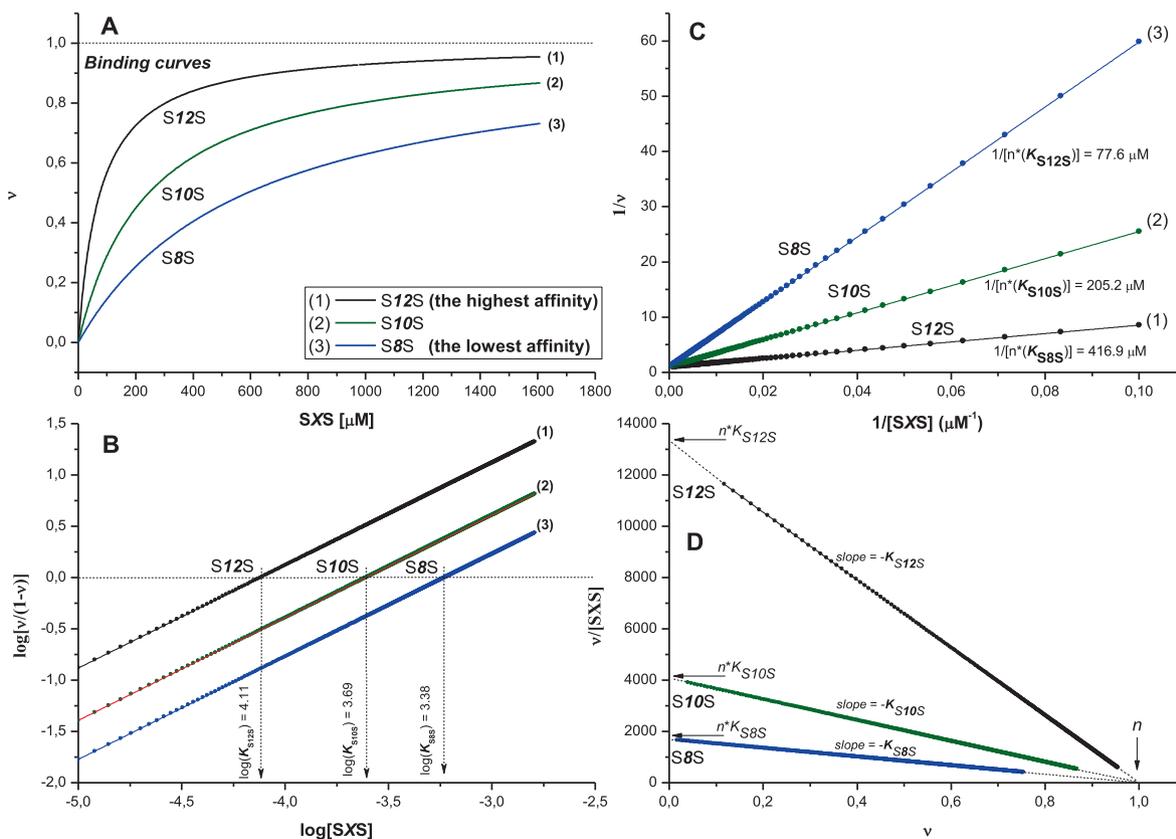


Fig. 4. Binding plots of the investigated systems: (1) β -CD/(S12S), (2) β -CD/(S10S) and β -CD/(S8S) simulated based on binding constants K_{SXS} (Table 1). (A) Binding curves: $v = \frac{K_{SXS}[SXS]}{1 + K_{SXS}[SXS]}$; (B) Double reciprocal representation: $\frac{1}{v} = 1 + \frac{1}{K_{SXS}[SXS]}$; (C) Hill representation: $\log\left(\frac{v}{1-v}\right) = \log(K_{SXS}[SXS])$; (D) Scatchard representation: $\frac{v}{[SXS]} = K_{SXS} - K_{SXS}v$.

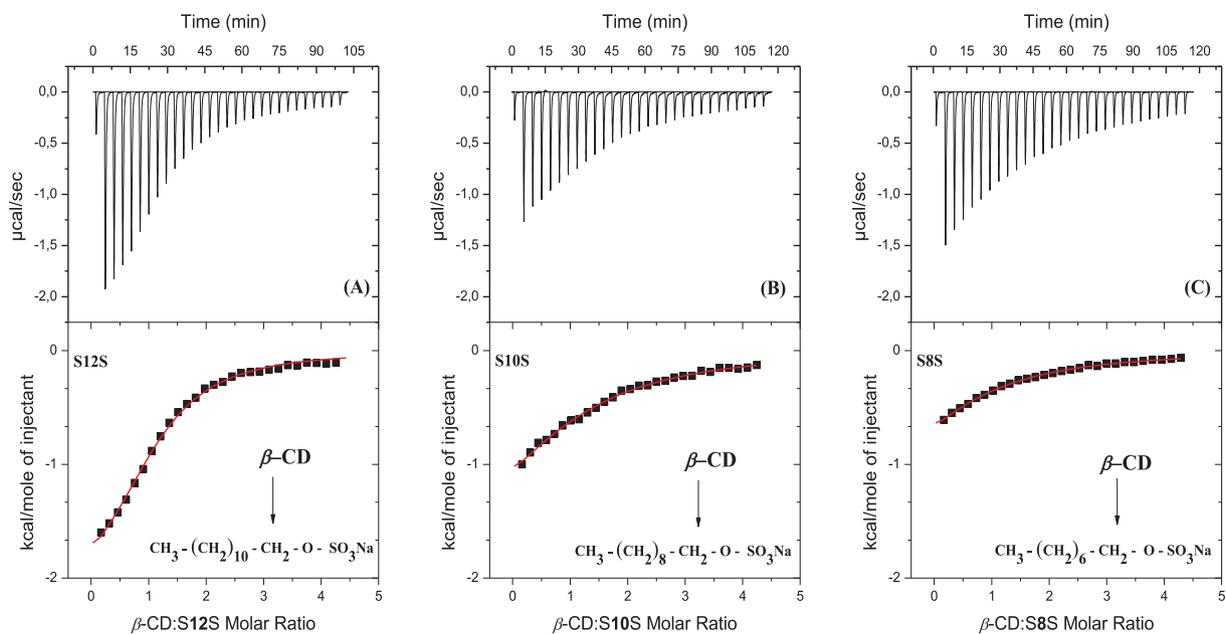


Fig. 5. Calorimetric titration isotherms of the binding interactions between β -CD and (A) sodium dodecyl sulfate (S12S), (B) sodium decyl sulfate (S10S) and (C) sodium octyl sulfate (S8S) in aqueous solutions, at 298.15 K.

mechanisms responsible for β -CD/SXS interactions computational methods have been employed.

3.3. Assessment of the electrostatic potential of β -CD cavities

The PBSA program from the AmberTools package was used to calculate the electrostatic potential isosurfaces corresponding to the primary and secondary cavity of β -CD. The obtained electrostatic potential maps are depicted in Fig. 5. The results revealed substantial differences between the analyzed cavities. It is clearly

visible that the potential binding site for the negatively charged $-\text{OSO}_3$ group is extended over a greater area in the case of the secondary cavity.

3.4. Putative entrance mechanisms - molecular dynamics (MD) simulations

Three 1 μs MD simulations were carried out to examine how various lengths of hydrocarbon chains of the guests might influence a host-guest complex formation. Mazurek et al. reported that

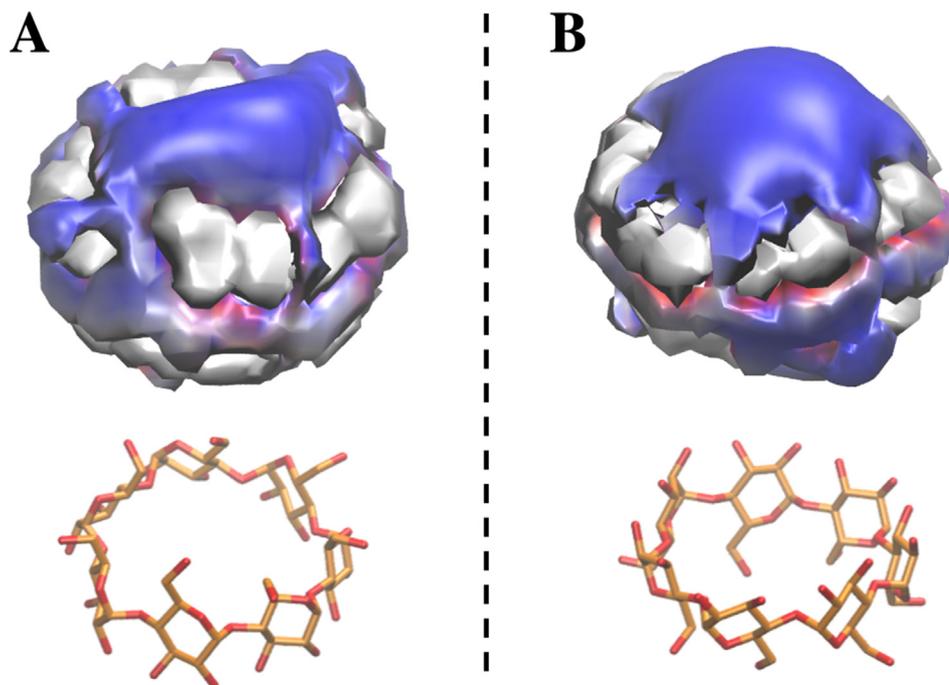


Fig. 6. Electrostatic potential isosurfaces for the primary (A) and secondary (B) cavities of β -CD (blue, $4 \text{ kcal mol}^{-1} \text{ e}^{-1}$).

Table 3

For each system five independent simulations were performed to inspect for the preferred mechanism of the entrance. The table summarizes whether the guest entered the host at any point of the simulation.

Guest	Mechanism of entrance							
	C1		C2		S1		S2	
S12S	1	No	1	No	1	Yes	1	No
	2	No	2	Yes	2	Yes	2	No
	3	No	3	Yes	3	Yes	3	No
	4	Yes	4	Yes	4	Yes	4	No
	5	No	5	Yes	5	Yes	5	No
S10S	1	Yes	1	No	1	Yes	1	No
	2	Yes	2	Yes	2	No	2	No
	3	Yes	3	Yes	3	No	3	No
	4	Yes	4	Yes	4	No	4	No
	5	Yes	5	Yes	5	No	5	No
S8S	1	No	1	Yes	1	Yes	1	No
	2	Yes	2	Yes	2	Yes	2	No
	3	Yes	3	Yes	3	Yes	3	No
	4	Yes	4	Yes	4	Yes	4	No
	5	Yes	5	Yes	5	Yes	5	No

Table 4

MMGBSA and LIE free energy decomposition values for dodecyl sulfate (S12S), decyl sulfate (S10S) and octyl sulfate (S8S) guests in the guest–host complexes (ΔG_{ele} = electrostatic component in vacuum, ΔG_{vdw} = van der Waals component, ΔG_{GB} = Generalized Born reaction field component, ΔG_{SURF} = non-polar solvation component, ΔG_{total} = total free energy of binding).

Guest	LIE [kcal/mol]			MMGBSA [kcal/mol]				
	ΔG_{ele}	ΔG_{vdw}	ΔG_{total}	ΔG_{ele}	ΔG_{vdw}	ΔG_{GB}	ΔG_{SURF}	ΔG_{total}
S12S	-0.2	-23.8	-23.9 ± 2.7	-12.5	-23.8	26.5	-3.5	-13.2 ± 2.3
S10S	-0.2	-22.4	-22.6 ± 2.7	-15.3	-22.4	28.9	-3.3	-12.1 ± 2.4
S8S	-0.2	-20.6	-20.8 ± 2.7	-16.7	-20.6	29.8	-3.1	-10.6 ± 2.3

Table 5

Free energy binding ($\Delta G_{\text{binding}}$) for dodecyl sulfate (S12S), decyl sulfate (S10S) and octyl sulfate (S8S) guests obtained from C-pull and S-pull dissociation pathways for Complex 1 and Complex 2.

Guest	$\Delta G_{\text{binding}}$ [kcal/mol]			
	Complex 1		Complex 2	
	C-pull path	S-pull path	C-pull path	S-pull path
S12S	-3.4	-1.3	-3.3	-2.0
S10S	-2.7	-1.8	-1.4	-2.2
S8S	-1.3	-1.5	-1.4	-3.4

there had not been any published attempt to simulate such systems other than through the molecular docking procedure [30]. In our work, we relied exclusively on the MD findings without performing a molecular docking procedure as a first step. Initially in each simulation, we placed the guest in proximity to β -CD. Subsequently, we performed a visual inspection of the generated trajectories. Although the guests successfully entered the cavity of the host, it did not happen in the same manner. We observed that S12S accessed the host through the primary cavity entering via S-end first, whereas S10S and S8S entered through the secondary cavity with the C-end first. Since the initial positions of guests were random, we decided to check whether any mechanism of how the guest enters the host would be preferable. We analyzed four ways of how it could happen: S-end first via the primary cavity (S1), C-end first via the secondary cavity (C2), C-end first via the primary cavity (C1) and S-end first via the secondary cavity (S2). In each

case, a guest was placed in proximity to either a primary or secondary cavity pointing its either C- or S-end at the center of β -CD (as shown in Fig. 3) and five short independent MD simulations were carried out. Subsequently, different entrance mode leads to different inclusion complex formation, if S1 or C2 were preferred Complex 1 was formed if C1 or S2 then Complex 2 (Fig. 6). Based on our findings that are summarized in Table 3 we see that:

- the mechanism S2 was never observed;
- for S12S, C2 and S1 were preferred, and Complex 2 was formed only once;
- for S10S, C1 and C2 were observed;
- S8S was the most flexible in terms of entering β -CD because C1, C2 and S1 entrance mechanisms were detected. To sum up, Complex 1 forms more preferentially.

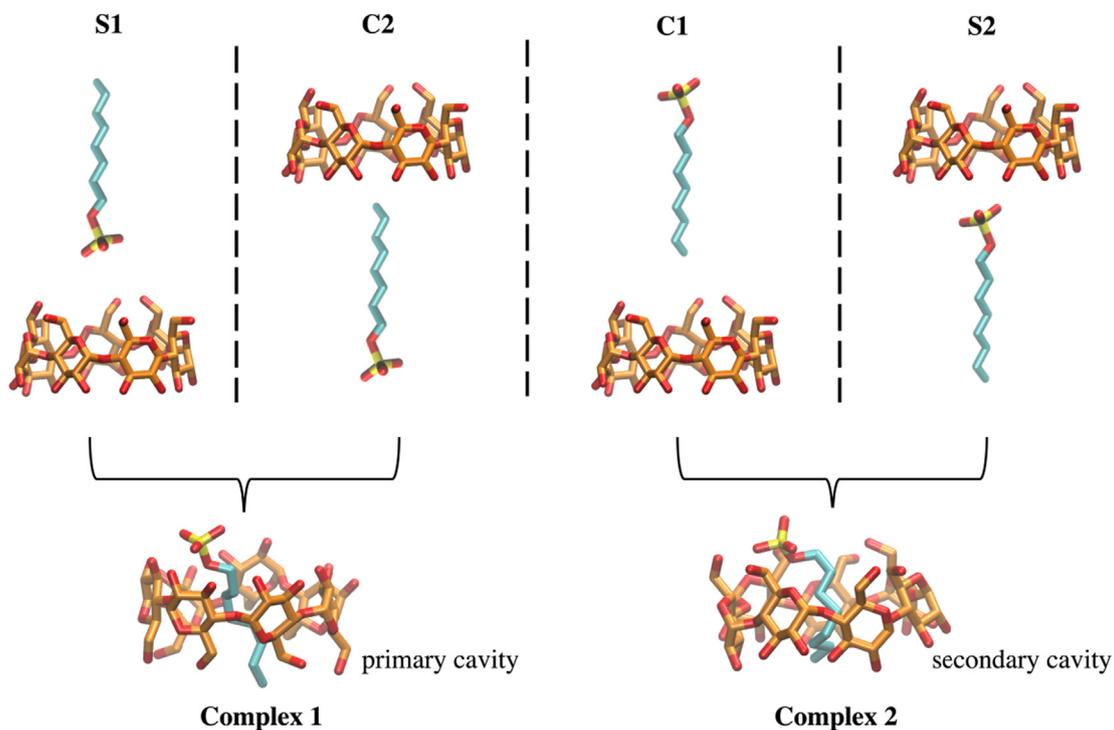


Fig. 7. Four potential mechanisms for how the guest could enter β -CD: S1 = S-end first via primary cavity, C2 = C-end first via secondary cavity, C1 = C-end first via primary cavity and S2 = S-end first via secondary cavity. For the clarity the colour of the guest molecule in Complex 1 and Complex 2 was changed in comparison to the unbound guest, and hydrogen atoms were omitted.

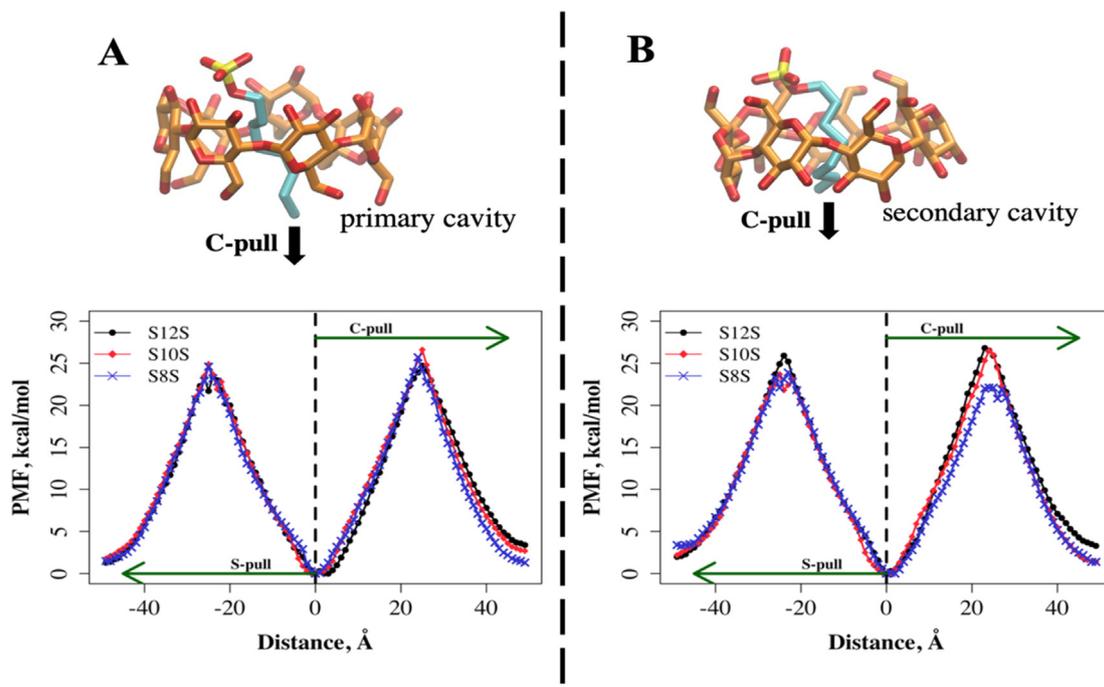


Fig. 8. PMF of dodecyl sulfate (S12S, black), decyl sulfate (S10S, red) and octyl sulfate (S8S, blue) dissociation from β -CD for Complex 1 (A) and Complex 2. Pathways C-pull and S-pull are combined on the same plot and the direction of each is marked by a green arrow.

3.5. Electrostatic vs Van der Waals interactions - binding free energy analysis

To estimate the free energy of binding in the β -CD-guest inclusion complexes, we used two popular endpoints approaches: LIE and MMGBSA (Table 4). Based on LIE results, we did not observe the change in the electrostatic component of the binding, which is calculated with the use of the isotropic dielectric constant. On the other hand, the MMGBSA technique showed that the shorter the hydrophobic chain, the stronger the electrostatic interactions in the host-guest complex (both *in vacuo* and with the consideration of GB implicit solvent model). At the same time, both methods revealed the same change in the van der Waals component, suggesting that the longer guests establish more contacts with the host. Based on both free energy calculation approaches, the total free energy of binding (ΔG_{total}) is the most favourable for S12S and least favourable for S8S, which agrees with the experimental

findings. Since we suggest that Complex 1 is more probable than Complex 2 this analysis was performed only for Complex 1.

3.6. Guest dissociation pathway analysis by umbrella sampling

To gather more detailed information about the binding interactions we conducted umbrella sampling simulations (US). The histograms used for WHAM analysis are presented in Supplementary Material Figs. S1 and S2. Deploying this method to build free energy profiles has already proved to be very effective in the case of β -CD and other small molecules [48–50]. As illustrated in Fig. 1, the S12S, S10S, and S8S ligands could potentially dissociate from the primary cavity (path C-pull) or the second cavity of β -CD in the case of Complex 1. Therefore, we constructed the PMF for both dissociation paths. According to the electrostatic potential analysis, the primary and secondary cavities of β -CD are not symmetric, and since the guest molecules are neither

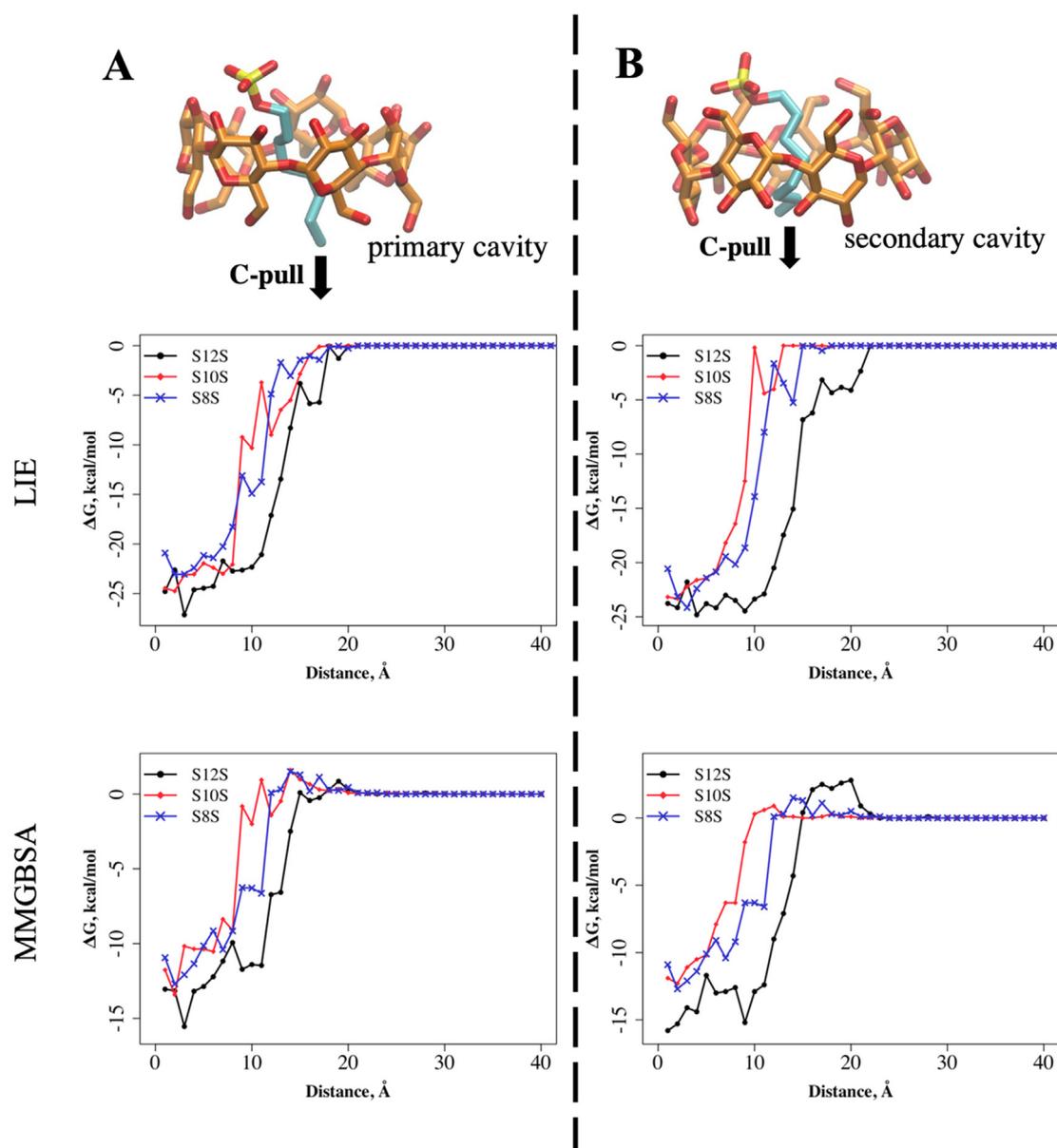


Fig. 9. Calculated LIE and MMGBSA in each US window for dodecyl sulfate (S12S), decyl sulfate (S10S) and octyl sulfate (S8S) for Complex 1 (A) and Complex 2 (B). Both LIE and MMGBSA were calculated for the C-pull pathway only.

symmetric, paths C-pull and S-pull may disclose the differences at least qualitatively. US calculations were performed for both Complex 1 and Complex 2 to check whether the different orientations of the guest could influence the dissociation pathway and resulting binding free energies. The $\Delta G_{\text{binding}}$ values are summarized in Table 5 and the constructed PMFs are depicted in Fig. 7. Small local minima visible on PMF plots for the S12S (Fig. 7A, S-pull) and S10S, S8S (Fig. 7B, S-pull) could be potentially explained in terms of the internal conformational preferences of the guest molecules appearing stochastically in the corresponding US windows.

3.6.1. Complex 1

For all guests, the energetic barrier is slightly higher in the case of the C-pull path, therefore this event could be less probable (Fig. 7). The data suggest that whereas the C-pull path shows distinct differences for $\Delta G_{\text{binding}}$ between guests, that is not the case for the S-pull path. This could be explained by the observation that during the S-pull path a part of the hydrocarbon chain is already outside the β -CD cavity together with the $-\text{OSO}_3$ group, therefore the remaining interactions are not that strong anymore. In the case of the C-pull path, the residual hydrocarbon chain needs to be pulled through the β -CD together with the $-\text{OSO}_3$ group.

3.6.2. Complex 2

Similarly to Complex 1, the energetic barrier is slightly higher for the guest in the case of the C-pull path except for S8S. The resulting $\Delta G_{\text{binding}}$ allows to rank the guest with the lowest binding free energy for S12S ($-3.3 \text{ kcal mol}^{-1}$) and then S10S and S8S (each $-1.4 \text{ kcal mol}^{-1}$). Interestingly, in the case of the S-pull path, S8S shows the lowest $\Delta G_{\text{binding}}$ value ($-3.4 \text{ kcal mol}^{-1}$), this might be due to the strong interactions between the hydrophobic interior of β -CD and shorter, compared to other guests, hydrocarbon tail of S8S.

In addition to the PMF profiles, we calculated $\Delta G_{\text{binding}}$ in each window, using LIE and MMGBSA, for the C-pull path only to check if these fundamentally different free energy calculation approaches would yield qualitatively the same results. In particular, the differences between the lowest and the highest binding free energy values calculated for trajectories in all US windows by LIE and MMGBSA define the probability of the dissociation of the guest (Fig. 8). Both methods suggest that the strongest binding between host and guest was in the β -CD/S12S complex, which agrees both with the PMF data and the experimental findings (see Fig. 9).

4. Conclusions

In this work, we characterized the structural features of three alkyl sulfates (S12S, S10S and S8S) governing the efficiency of the formation of inclusion complexes with β -cyclodextrin. We showed that the thermodynamic stability of the investigated complexes increased in proportion to the elongation of the guest alkyl chain. Both experimental and computational studies uncovered that besides the hydrophobicity, electrostatic and van der Waals interactions should be considered for investigating a structure-based host-guest complexes design. In fact, the obtained results revealed that the contribution of electrostatic interactions is more pronounced for the shorter alkyl chain sulfates. In contrast, van der Waals interactions were crucial for the complexation efficiency of the long tail guest. Additionally, four putative mechanisms of alkyl sulfates' entrance into the β -CD cavity were inspected for the first time by molecular dynamics-based computational approaches. It is worth noticing that a different entrance mode leads to the formation of different types of inclusion complexes. It can be supposed that the resulting complexes exist in a conformation equilibrium. Our results indicate that the length of guest

tails was vital for the complexation mode of β -CD. Among investigated mechanisms of the entrance of a guest through the secondary cavity with a hydrophilic head was never observed. S12S preferred to access the host predominantly through the primary cavity via the negatively charged $-\text{OSO}_3$ group (the S1 mechanism), whereas for S10S the C1 and C2 entrance mechanisms predominated. Oppositely, S8S demonstrated the highest variability in terms of entering the β -CD cavity. Finally, the observed phenomena were subsequently confirmed by a guest dissociation pathway analysis. To sum up, our work sheds more light on how the size of hydrophobic guest moiety impacts the host-guest complex formation, suggests potential entrance mechanisms and stresses the predictive power of computational approaches. All above could contribute to successful and faster drug design involving β -CD.

CRedit authorship contribution statement

Małgorzata M. Kogut: Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Ola Grabowska:** Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Dariusz Wyrzykowski:** Conceptualization, Investigation, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing. **Sergey A. Samsonov:** Conceptualization, Investigation, Funding acquisition, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing.

Data availability

No data was used for the research described in the article.

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.

Acknowledgements

This research was funded by National Science Centre of Poland, grant number UMO-2018/30/E/ST4/00037. Computational resources were provided by the Polish Grid Infrastructure (PL-GRID, grants plggagstr3, plgionsgpu), and local cluster at the Faculty of Chemistry, University of Gdansk.

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.molliq.2022.119978>.

References

- [1] E.V.R. Campos, P.L.F. Proença, J.L. Oliveira, C.C. Melville, J.F. Della Vecchia, D.J. Andrade, L.F. Fraceto, Chitosan nanoparticles functionalized with β -cyclodextrin: a promising carrier for botanical pesticides, *Sci. Rep.* 8 (1) (2018) 1–15.
- [2] H.M.C. Marques, A review on cyclodextrin encapsulation of essential oils and volatiles, *Flavour Fragr. J.* 25 (5) (2010) 313–326.
- [3] E.M.M. Del Valle, Cyclodextrins and their uses: a review, *Process Biochem.* 39 (9) (2004) 1033–1046.
- [4] P. Mahalapbutr, B. Nutho, P. Wolschann, W. Chavasiri, N. Kungwan, T. Rungrotmongkol, Molecular insights into inclusion complexes of mansonone E and H enantiomers with various β -cyclodextrins, *J. Mol. Graph. Model.* 79 (2018) 72–80.
- [5] P. Mahalapbutr, K. Thitinanthavet, T. Kedkham, H. Nguyen, L.T.h. Theu, S. Dokmaisrijan, L. Huynh, N. Kungwan, T. Rungrotmongkol, A theoretical study

- on the molecular encapsulation of luteolin and pinocembrin with various derivatized beta-cyclodextrins, *J. Mol. Struct.* 1180 (2019) 480–490.
- [6] P. Wongpituk, B. Nutho, W. Panman, N. Kungwan, P. Wolschann, T. Rungrotmongkol, N. Nunthaboot, Structural dynamics and binding free energy of neral-cyclodextrins inclusion complexes: Molecular dynamics simulation, *Mol. Simul.* 43 (13–16) (2017) 1356–1363.
- [7] G. Astray, C. Gonzalez-Barreiro, J.C. Mejuto, R. Rial-Otero, J. Simal-Gándara, A review on the use of cyclodextrins in foods, *Food Hydrocoll.* 23 (7) (2009) 1631–1640.
- [8] G. Crini, E. Fenyvesi, L. Sente, Outstanding contribution of Professor József Szejtli to cyclodextrin applications in foods, cosmetics, drugs, chromatography and biotechnology: a review, *Environ. Chem. Lett.* 19 (3) (2021) 2619–2641.
- [9] L.B.d. Carvalho, K.K. Buruso, C. Jaime, T. Venâncio, A.F.S.d. Carvalho, L.D.S. Murgas, L.d.M.A. Pinto, Complexes between methyltestosterone and β -cyclodextrin for application in aquaculture production, *Carbohydr. Polym.* 179 (2018) 386–393.
- [10] Z. Xiao, Y. Zhang, Y. Niu, Q. Ke, X. Kou, Cyclodextrins as carriers for volatile aroma compounds: A review, *Carbohydr. Polym.* 269 (2021) 118292, <https://doi.org/10.1016/j.carbpol.2021.118292>.
- [11] H.H. Wu, P. Garidel, B. Michaela, HP- β -CD for the formulation of IgG and Ig-based biotherapeutics, *Int. J. Pharm.* 601 (2021) 120531.
- [12] B. Liu, W. Li, J. Zhao, Y. Liu, X. Zhu, G. Liang, Physicochemical characterisation of the supramolecular structure of luteolin/cyclodextrin inclusion complex, *Food Chem.* 141 (2) (2013) 900–906.
- [13] S.-Y. Zhou, S.-X. Ma, H.-L. Cheng, L.-J. Yang, W. Chen, Y.-Q. Yin, Y.-M. Shi, X.-D. Yang, Host-guest interaction between pinocembrin and cyclodextrins: Characterization, solubilization and stability, *J. Mol. Struct.* 1058 (2014) 181–188.
- [14] C. Jullian, C. Cifuentes, M. Alfaro, S. Miranda, G. Barriga, C. Olea-Azar, Spectroscopic characterization of the inclusion complexes of luteolin with native and derivatized beta-cyclodextrin, *Bioorg. Med. Chem.* 18 (14) (2010) 5025–5031.
- [15] E.T. Hwang, S. Lee, J.S. Kim, J. Jeong, B.S. Jeon, J.W. Lee, J.H. Kim, J. Kim, Highly Stable and Fine-Textured Hybrid Microspheres for Entrapment of Cosmetic Active Ingredients, *ACS Omega* 5 (45) (2020) 29577–29584.
- [16] L. dos Santos Silva Araújo, G. Lazzara, L. Chiappisi, Cyclodextrin/surfactant inclusion complexes: an integrated view of their thermodynamic and structural properties, *Adv. Colloid Interface Sci.* 289 (2021) 102375, <https://doi.org/10.1016/j.cis.2021.102375>.
- [17] H. Cui, M. Bai, L. Lin, Plasma-treated poly (ethylene oxide) nanofibers containing tea tree oil/ β -cyclodextrin inclusion complex for antibacterial packaging, *Carbohydr. Polym.* 179 (2018) 360–369.
- [18] J. Szejtli, Introduction and general overview of cyclodextrin chemistry, *Chem. Rev.* 98 (5) (1998) 1743–1754.
- [19] D.J. Cram, J.M. Cram, Host-guest chemistry, *Science* 183 (4127) (1974) 803–809.
- [20] K. Uekama, F. Hirayama, T. Irie, Cyclodextrin drug carrier systems, *Chem. Rev.* 98 (5) (1998) 2045–2076.
- [21] C.R. Groom, I.J. Bruno, M.P. Lightfoot, S.C. Ward, The Cambridge Structural Database, *Acta Crystallogr. B* 72 (2) (2016) 171–179.
- [22] P. Mura, Analytical techniques for characterization of cyclodextrin complexes in aqueous solution: a review, *J. Pharm. Biomed. Anal.* 101 (2014) 238–250.
- [23] A.J.M. Valente, O. Söderman, The formation of host-guest complexes between surfactants and cyclodextrins, *Adv. Colloid Interface Sci.* 205 (2014) 156–176.
- [24] R. Palepu, V.C. Reinsborough, Surfactant-cyclodextrin interactions by conductance measurements, *Can. J. Chem.* 66 (2) (1988) 325–328.
- [25] R. Palepu, J.E. Richardson, V.C. Reinsborough, Binding constants of beta-cyclodextrin/surfactant inclusion by conductivity measurements, *Langmuir* 5 (1) (1989) 218–221.
- [26] A.A. Rafati, A. Bagheri, H. Iloukhani, M. Zarinehad, Study of inclusion complex formation between a homologous series of n-alkyltrimethylammonium bromides and β -cyclodextrin, using conductometric technique, *J. Mol. Liq.* 116 (1) (2005) 37–41.
- [27] A. Hospital, J.R. Goñi, M. Orozco, J.L. Gelpi, Molecular dynamics simulations: Advances and applications, *Adv. Appl. Bioinform. Chem.* 8 (2015) 37–47.
- [28] J.R. Perilla, B.C. Goh, C.K. Cassidy, B.o. Liu, R.C. Bernardi, T. Rudack, H. Yu, Z. Wu, K. Schulten, Molecular dynamics simulations of large macromolecular complexes, *Curr. Opin. Struct. Biol.* 31 (2015) 64–74.
- [29] J. Feng, J. Chen, B. Selvam, D. Shukla, Computational microscopy: Revealing molecular mechanisms in plants using molecular dynamics simulations, *The Plant Cell* 31 (2019) 1319.
- [30] A.H. Mazurek, L. Szeleszczuk, T. Gubica, Application of Molecular Dynamics Simulations in the Analysis of Cyclodextrin Complexes, *Int. J. Mol. Sci.* 22 (2021) 9422.
- [31] A. Tesmar, M.M. Kogut, K. Żamojć, O. Grabowska, K. Chmur, S.A. Samsonov, J. Makowska, D. Wyrzykowski, L. Chmurzyński, Physicochemical nature of sodium dodecyl sulfate interactions with bovine serum albumin revealed by interdisciplinary approaches, *J. Mol. Liq.* 340 (2021) 117185.
- [32] O. Grabowska, M.M. Kogut, K. Żamojć, S.A. Samsonov, J. Makowska, A. Tesmar, K. Chmur, D. Wyrzykowski, L. Chmurzyński, Effect of Tetraphenylborate on Physicochemical Properties of Bovine Serum Albumin, *Molecules* 26 (2021) 6565.
- [33] K. Hemine, A. Skwierawska, C. Kleist, M. Olewniczak, K. Szwarz-Karabyka, D. Wyrzykowski, A. Mieszkowska, J. Chojnacki, J. Czub, L. Nierzwicki, Effect of chemical structure on complexation efficiency of aromatic drugs with cyclodextrins: the example of dibenzazepine derivatives, *Carbohydr. Polym.* 250 (2020) 116957, <https://doi.org/10.1016/j.carbpol.2020.116957>.
- [34] D.A. Case, D.S. Cerutti, T.E. III Cheatham, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, D. Greene, N. Homeyer, S. Izadi, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith, J. Swails, R.C. Walker, J. Wang, R. M. Wolf, R.X. Wu, L. Xiao, D.M. York, P.A. Kollman, AMBER 2017, University of California: San Francisco (2017).
- [35] W. Humphrey, A. Dalke, K. Schulten, VMD—Visual Molecular Dynamics, *J. Mol. Graph.* 14 (1) (1996) 33–38.
- [36] K.N. Kirschner, A.B. Yongye, S.M. Tschampel, J. González-Outeiriño, C.R. Daniels, B.L. Foley, R.J. Woods, GLYCAM06: a generalizable biomolecular force field. Carbohydrates, *J. Comput. Chem.* 29 (4) (2008) 622–655.
- [37] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing of a general Amber force field, *J. Comput. Chem.* 25 (9) (2004) 1157–1174.
- [38] C.I. Bayly, P. Cieplak, W.D. Cornell, P.A. Kollman, A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model, *J. Phys. Chem.* 97 (40) (1993) 10269–10280.
- [39] V. Tsui, D.A. Case, Theory and applications of the generalized Born solvation model in macromolecular simulations, *Biopolymers* 56 (4) (2000) 275–291.
- [40] A. Onufriev, D. Bashford, D.A. Case, Modification of the generalized Born model suitable for macromolecules, *J. Phys. Chem. B* 104 (15) (2000) 3712–3720.
- [41] S. Kumar, J.M. Rosenberg, D. Bouzida, R.H. Swendsen, P.A. Kollman, The weighted histogram analysis method for free-energy calculations on biomolecules. I: The method, *J. Comput. Chem.* 13 (8) (1992) 1011–1021.
- [42] A. Grossfield, “WHAM: the weighted histogram analysis method”, version 2.0.10.2, http://membrane.urmc.rochester.edu/wordpress/?page_id=126.
- [43] R Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2020 <https://www.R-project.org/>.
- [44] M. Józwiak, L. Madej, Complex Formation of Crown Ethers and Cations in Water-Organic Solvent Mixtures: The Thermodynamic Functions of Complex Formation of Benzo-15-crown-5 with Na⁺ in Water + Ethanol at 298.15 K, *J. Chem. Eng. Data* 55 (2010) 1965–1970.
- [45] K. Bouchemal, S. Mazzaferro, How to conduct and interpret ITC experiments accurately for cyclodextrin-guest interactions, *Drug Discov. Today* 17 (11-12) (2012) 623–629.
- [46] A. Woźniowska, G. Gołuski, D. Wyrzykowski, R. Kaźmierkiewicz, J. Piosik, Caffeine and other methylxanthines as interceptors of food-borne aromatic mutagens: inhibition of Trp-P-1 and Trp-P-2 mutagenic activity, *Chem. Res. Toxicol.* 26 (11) (2013) 1660–1673.
- [47] D.P. Voronin, A.S. Buchelnikov, V.V. Kostjukov, S.V. Khrapaty, D. Wyrzykowski, J. Piosik, Y.I. Prylutskiy, U. Ritter, M.P. Evstigneev, Evidence of entropically driven C60 fullerene aggregation in aqueous solution, *J. Chem. Phys.* 140 (10) (2014) 104909, <https://doi.org/10.1063/1.4867902>.
- [48] H. Zhang, T. Tan, C. Hetenyi, Y. Lv, D. van der Spoel, Cooperative Binding of Cyclodextrin Dimers to Isoflavone Analogues Elucidated by Free Energy Calculations, *J. Phys. Chem C* 118 (2014) 7163–7173.
- [49] M.I. Sancho, S. Andujar, R.D. Porasso, R.D. Enriz, Theoretical and experimental study of inclusion complexes of β -cyclodextrins with chalcone and 2',4'-dihydroxychalcone, *J. Phys. Chem. B* 120 (12) (2016) 3000–3011.
- [50] W. You, Z. Tang, C.A. Chang, Potential mean force from umbrella sampling simulations: What can we learn and what is missed?, *J. Chem. Theory and Comput.* 15 (4) (2019) 2433–2443.

PUBLICATION D4

**PHYSICOCHEMICAL NATURE OF SODIUM DODECYL
SULFATE INTERACTIONS WITH BOVINE SERUM
ALBUMIN REVEALED BY INTERDISCIPLINARY
APPROACHES**



Physicochemical nature of sodium dodecyl sulfate interactions with bovine serum albumin revealed by interdisciplinary approaches



Aleksandra Tesmar¹, Małgorzata M. Kogut¹, Krzysztof Żamojć, Ola Grabowska, Katarzyna Chmur, Sergey A. Samsonov, Joanna Makowska, Dariusz Wyrzykowski*, Lech Chmurzyński

Faculty of Chemistry, University of Gdańsk, Wita Stwosza 63, 80-308 Gdańsk, Poland

ARTICLE INFO

Article history:

Received 4 May 2021

Revised 1 July 2021

Accepted 3 August 2021

Available online 5 August 2021

Keywords:

Sodium dodecyl sulfate

Bovine serum albumin

Thermodynamic parameters

Binding site

Molecular dynamics

ABSTRACT

To rigorously characterize the interactions of sodium dodecyl sulfate (SDS) with bovine serum albumin (BSA) a set of experimental methods, namely isothermal titration calorimetry, conductometric titration, steady-state fluorescence spectroscopy, differential scanning calorimetry and circular dichroism spectroscopy, supported by *in silico* analysis have been applied. The influence of pH and temperature on the binding mode has been revealed. At a low molar ratio of SDS to BSA up to ca. 16:1, there are at least two structurally distinct binding sites in BSA. The formation of SDS-BSA complexes is an enthalpy-driven process in which the van der Waals interactions play a crucial role. The first binding site, located close to the Trp-134 residue within the sub-domain IA, is pH-independent and binds two molecules of SDS per one molecule of BSA whereas the total number of SDS molecules bound to the second site of albumin is affected by temperature and pH. The saturation of the first binding site of BSA (ca. 0.009 mg of SDS per 1 mg of BSA) is sufficient to thermally stabilize the helical conformation of BSA. The presented results have important structural and thermodynamic implications to understand the influence of a widely used anionic surfactant on globular protein functionality in modern branches of chemistry.

© 2021 Elsevier B.V. All rights reserved.

1. Introduction

Surfactants are the subject of interest to many research groups since it has been found that they can be utilized in several industrial and technical applications, including detergent industry, food chemistry, drug delivery, and cosmetic preparation as well as developments in life sciences [1–4]. Much attention has been focused on studies of surfactant interactions with biologically relevant macromolecules such as proteins [5–11].

An excellent review on the protein - surfactant interactions provided by Otzen reveals the most important techniques for analysing these interactions and highlights the different issues related to this field, namely the impact of surfactant on protein denaturation, binding affinity and unfolding processes [12]. The macromolecule - surfactant interactions depend on many factors, including the topology of the investigated surfactant species (monomeric and micelle forms), a chemical structure of a surfactant (anionic, cationic, amphoteric or non-ionic) as well as experimental conditions (pH, temperature). A remarkable degree of specificity and high

affinity of surfactants towards proteins results in changes in their physicochemical properties. Consequently the conformational changes and the saturation of the binding sites of a macromolecule affect its biological activity upon a complex formation.

The understanding of the thermodynamic and thermal stability of the binding process and the structural changes of a protein in the presence of ubiquitous surfactants may be helpful for an in-depth interpretation of a surfactant role underlying biologically relevant protein's functions. Moreover, it is fundamental from the viewpoint of the surfactants application in environmental processes and the pharmaceutical industry [13]. In this paper, we report the influence of pH and temperature on the interactions of the sodium dodecyl sulfate (SDS), a common, anionic surfactant, with the protein bovine serum albumin (BSA). As far as we are concerned, there are few reports on the SDS - BSA interactions [14–19]. However, in contrast to these previous studies, we have focused our attention on a low SDS to BSA molar ratio range. Furthermore, the experimental and *in silico* data related to the structural impact of pH and temperature on the potential binding sites and poses of BSA have been discussed.

* Corresponding author.

E-mail address: dariusz.wyrzykowski@ug.edu.pl (D. Wyrzykowski).

¹ Authors contributed equally.

2. Experimental part

2.1. Materials

Bovine serum albumin (BSA, lyophilized powder, $\geq 96\%$, Sigma-Aldrich, Poland), sodium dodecyl sulfate (SDS, BioUltra, for molecular biology, $\geq 99\%$, Sigma-Aldrich, Poland), sodium cacodylate trihydrate (Caco, $\geq 99\%$, Sigma-Aldrich, Poland) were used as obtained without further purification. Double-distilled water with conductivity not exceeding $0.18 \mu\text{S cm}^{-1}$ was used for preparations of aqueous solutions.

2.2. Conductometric analysis

A microtitration unit (Cerko Lab System, Poland) fitted with a 5 mL syringe (Hamilton, Poland) and a CD-201 conductometric cell (Hydromet, Poland) were used for conductometric measurements. A weight calibration method was used for the syringe. The conductometric electrode was standardized with conductivity standards (aqueous KCl solutions) of a conductivity of 84 and $200 \mu\text{S cm}^{-1}$ (Hamilton, Poland). The measurements were carried out at $298.15 \pm 0.10 \text{ K}$, controlled by the Lauda E100 circulation thermostat.

2.3. Isothermal titration calorimetry

All ITC experiments were performed at 288.15 K and 298.15 K using an AutoITC isothermal titration calorimeter (MicroCal Inc. GE Healthcare, Northampton, USA). The details of the measuring devices and experimental setup have been described previously [20]. The reagents (BSA, SDS) were dissolved directly in 10 mM Caco buffers of pH 5 and 7. The experiment consisted of injecting $10.02 \mu\text{L}$ (29 injections, $2 \mu\text{L}$ for the first injection only) of 1 mM buffered solution of SDS into the reaction cell which initially contained the 0.0125 mM buffered solution of BSA. A background titration, consisting of an identical titrant solution but with the buffer solution in the reaction cell only, was subtracted from each experimental titration on account of the heat of dilution. The titrant (SDS) was injected at 5 min intervals. Each injection lasted 20 s. The stirrer speed was kept constant at 300 rpm. The CaCl_2 - EDTA titration was performed to check the apparatus and the results (stoichiometry, K , ΔH) were compared with those obtained for the same samples (a test kit) at MicroCal Inc. GE Healthcare.

2.4. Steady-state fluorescence spectroscopy and UV spectrophotometry

Steady-state fluorescence experiments were carried out at 288.15 K and 298.15 K using a Cary Eclipse Varian (Agilent, Santa Clara, CA, USA) spectrofluorometer equipped with a temperature controller and a 1.0 cm quartz multicell holder. The initial concentration of the stock solution of BSA ($2 \mu\text{M}$) was confirmed spectrophotometrically ($\epsilon_{\text{BSA}}^{280} = 41180 \text{ M}^{-1} \text{ cm}^{-1}$) based on the absorbance and the values of the molar extinction coefficients of tryptophan, tyrosine, and cysteine determined at 280 nm ($\epsilon_{\text{W}}^{280} = 5690 \text{ M}^{-1} \text{ cm}^{-1}$; $\epsilon_{\text{Y}}^{280} = 1280 \text{ M}^{-1} \text{ cm}^{-1}$; $\epsilon_{\text{C}}^{280} = 120 \text{ M}^{-1} \text{ cm}^{-1}$) [21]. The absorption spectra were recorded at 288.15 K and 298.15 K on a Perkin Elmer Lambda 650 (Waltham, MA, USA) UV/Vis spectrophotometer. The fluorescence emission spectra ($\lambda_{\text{ex}} = 275 \text{ nm}$) of BSA ($2 \mu\text{M}$) were recorded from 280 to 500 nm in the absence and presence of increasing concentrations of SDS, up to $29.3 \mu\text{M}$. In the performed fluorescence titration experiments, 2 mL of BSA at $2 \mu\text{M}$ was titrated with thirty $1 \mu\text{L}$ aliquots of SDS solution at 1 mM and then with fifteen $1 \mu\text{L}$ aliquots of SDS solution at 2 mM . The intensity of the band at 348 nm (corresponding to the maximum emission of BSA) was used to calculate the binding constants (K) and other parameters.

2.5. Differential scanning calorimetry

Calorimetric (DSC) measurements were made with a VP-DSC microcalorimeter (MicroCal Inc. GE Healthcare, Northampton, USA) at a scanning rate 90 K h^{-1} . All scans were run at two different pH values: 5 and 7 in the Caco buffer, in a temperature range from 298.15 K to 363.15 K . At first, scans were obtained for pure BSA at a concentration of 0.015 mM . In the second step, scans were obtained for BSA/SDS mixtures in stoichiometric ratios: 1:4, 1:8, and 1:14, respectively. The cell volume was always 0.5 mL . The reversibility of the transition was checked by cooling and reheating the same sample. These measurements were recorded two times. Results from DSC measurements were analyzed using the software Origin 7.0 from MicroCal, employing routines included with the instrument [22]. The quantity measured by DSC is the difference between the heat capacity of the Caco buffer-protein BSA solution and that of a pure Caco buffer. To perform the DSC measurement of protein unfolding, the reference cell was filled with buffer and the sample cell with the appropriate protein solution. They were then heated at a constant scan rate (90 K h^{-1}). When a protein unfolds, during DSC measurements, the occurring absorption of heat causes a temperature difference (ΔT) between the cells. The reference (Caco buffer) and sample solutions (protein BSA) were properly equilibrated with dissolved air before being introduced into the cells. Five minutes of vacuum treatment was required to degas all samples. The first step of the calibration was to carry out the buffer/buffer scans with run parameters exactly the same as for the comparative scans when the protein is in the sample cell. A pre-scan thermostat period was 15 min as recommended.

2.6. Circular dichroism spectroscopy

Circular dichroism (CD) spectra were recorded in water on a Jasco-715 automatic recording spectropolarimeter (Jasco Inc., USA) for following systems: pure BSA and two BSA/SDS mixtures in stoichiometric ratios: 1:4 and 1:8, respectively. The experiments were performed in the 10 mM Caco buffer at two different pH: 5 and 7 in a temperature range from 298.15 K to 368.15 K , every 10 degrees. Measurements were made in 1 mm quartz cuvettes. The spectra were recorded in the 200 – 260 nm wavelength range, using a sensitivity of five millidegrees and a scan speed of 50 nm min^{-1} . The results were plotted as the mean residue ellipticity Θ [$\text{degree cm}^2 \text{ dmol}^{-1}$]. The CD measurements were made at the 0.015 mM protein concentration. The volume of samples was 0.3 cm^3 . The secondary structure content was estimated from all CD spectra using the CONTIN/LL program, a variant of the CONTIN method developed by Provencher and Glockner [23] provided within the CDPPro software package.

2.7. Theoretical studies

Structures. The structure of BSA was obtained from the Protein Data Bank (PDB ID: 1F5S, 2.47 \AA). The structure of SDS was built in xleap of AMBER 16 package [24].

Molecular docking. The Autodock3 [25] software was used to perform molecular docking. A maximum gridbox size was used ($126 \text{ \AA} \times 126 \text{ \AA} \times 126 \text{ \AA}$) that covered the entire protein surface, including the expected/predicted SDS "first" binding site with the default grid step of 0.375 \AA . The Lamarckian genetic algorithm was deployed for 1000 independent runs. The size of 300 for the initial population and 10^5 generations for termination conditions were chosen and 9995×10^5 energy evaluations were performed. For different protonation states, His residues with and without a proton covalently bound to the nitrogen atom were used.

Molecular dynamics simulations. All-atom molecular dynamics (MD) simulations were performed with the use of a AMBER16

software package [24]. The initial structures of the complexes were created by a random placement of 15 SDS molecules around the BSA protein. The truncated octahedron TIP3P periodic box of 8 Å water layer from box's border to solute was used to solvate complexes. The ff14SBonlysc force field for the protein [26] and gaff force field [27] with RESP charges [28] obtained in antechamber module of AMBER16 [24] for SDS molecule were used. Charge was neutralized with either Na⁺ or Cl⁻ counterions. An energy minimization was carried out in two steps: beginning with 500 steepest descent cycles and 10³ conjugate gradient cycles with 100 kcal/mol/Å² harmonic force restraints, continued with 3x10³ steepest descent cycles and 3x10³ conjugate gradient cycles without any restraints. Following minimization steps the system was heated up from 0 to 300 K for 10 ps with harmonic force restraints of 100 kcal/mol/Å². Then, the system was equilibrated at 300 K and 10⁵ Pa in an isothermal isobaric ensemble for 100 ps. Afterwards, the two independent prediction MD runs were carried out in the same isothermal isobaric ensemble for 100 ns. The particle mesh Ewald method for treating electrostatics and SHAKE algorithm for all the covalent bonds containing hydrogen atoms were implemented in the MD simulations.

Binding free energy calculations. Energetic post-processing of the trajectories and per-residue energy decomposition were performed for all systems with the use of the Linear Interaction Energy (LIE) with dielectric constant of 80 and Molecular Mechanics Generalized Born Surface Area (MM-GBSA) [29] using a model with the surface area and Born radii default parameters as implemented in the igb = 2 model in AMBER16 [30]. All frames from MD simulations were analysed.

3. Results

3.1. Critical micelle concentration (CMC) of SDS in cacodylate buffer

Surfactant molecules reveal tendency to self-association and formation of ordered structures (micelles). Thus, the investigations of systems in which surfactants are involved require determination of the critical micelle concentration (CMC) [31,32]. The presence of monomeric surfactants (below the CMC) and micelles modifies solution properties as well as has a considerable impact on their interaction's mode with various proteins [33–35]. The CMC value depends on many factors, among which the ionic strength of an aqueous solution as well as a type of electrolytes play the crucial role [36–38].

The CMC values of SDS under the experimental conditions (the 10 mM Caco buffers of pH 5 and 7, at 298.15 K) were determined from conductometric titration data and ITC measurements (Fig. 1). The specific conductivity–surfactant concentration plots comprise two straight lines with different slopes corresponding to the SDS concentration range below and above CMC, respectively. The CMC values were calculated from the intersection of these lines [36]. In turn, the CMC values from ITC data were determined from the maximum peak of the fitted curve (the Gauss model from Origin 8.5) to the heats associated with injections of the SDS buffered solution into the buffer [39].

The calculated CMC values (5.4–5.6 mM) are lower than that reported for water (CMC ca. 8 mM) [40]. It is worth noticing that the pH of the Caco buffers (5 and 7) does not affect the CMC of SDS. Thus, the ionic strength of the solution maintained by buffer components is a key factor which favours the SDS micelle formation. The experimentally determined CMC values (CT and ITC measurements) correlate very well with a theoretical one (5.5 mM) calculated through the empirical expression developed by Corrin and Harkin [41]:

$$\log(\text{CMC}) = a + b \log(c_1)$$

where c_1 denotes the total counterion concentration (mol dm⁻³). The literature values of a and b variables for SDS are -3.23 and -0.486 , respectively [36].

3.2. Isothermal titration calorimetry (ITC)

Representative binding isotherms for the SDS-BSA interactions in the 10 mM Caco buffers of pH 5 and 7 at 298.15 K and 288.15 K are shown in Fig. 2, whereas conditional parameters of the interactions are summarized in Table 1. The thermodynamic parameters (K_{ITC} and ΔH_{ITC}) were obtained directly from ITC measurements by fitting isotherms (using nonlinear least-squares procedures) to a model that assumes two sets of binding sites. The assumed model has given the best fitting of calculated data to the experimental ones for all systems under study except for the SDS-BSA interactions at pH 5, at 288.15 K. Thus, for the latter system only a stoichiometry (expressed with N denoting SDS:BSA molar ratio) of the interactions was calculated based on the first derivative of the fitting curve obtained by the polynomial fit (OriginPro 8.5). The standard thermodynamic relationships: $\Delta G_{\text{ITC}} = -RT \ln K_{\text{ITC}} = \Delta H_{\text{ITC}} - T\Delta S_{\text{ITC}}$ were used to calculate the free energy of binding, ΔG_{ITC} , and the entropy change, ΔS_{ITC} (Table 1).

According to the ITC and conductometric titration measurements, under the experimental conditions ($c_{\text{SDS}} < \text{CMC}$), monomer SDS species take part in the interactions with the protein. Calorimetric data revealed that there are at least two binding sites of BSA capable to bind SDS. The first site has a higher affinity towards SDS than the second one ($\log K_{\text{ITC}(1)} > \log K_{\text{ITC}(2)}$, Table 1) and binds two SDS molecules followed by the saturation of the second site. The second binding site is saturated, depending on the pH of a buffer solution, by ca. 4–5 (at pH 7) and 9 (at pH 5) SDS particles. On the other hand, it is worth noticing that the reduction of the negative net charge of the protein by lowering pH from 7 to 5 does not affect the number of SDS molecules bound to the first (high affinity) site but has an impact on the number of SDS molecules inserted into the second (low affinity) BSA site.

It is interesting to note that an additional energy effect appears in the ITC curve at pH 5 at 288.15 K which suggests the presence of a third binding site in the protein at lower temperature and pH (see Fig. 2, 10 mM Caco, pH 5, 288.15 K). Unfortunately, in this case the theoretical curve generated based on equilibrium models implemented in Origin 7 from MicroCal cannot be fitted to the experimental points. Under these conditions ca. 2.8, 3.5 and 3.5 molecules of SDS are bound to the first, the second and the third binding sites of BSA, respectively. However, a total number of SDS molecules bound to BSA at 288.15 K is equal (in the range of the experimental error) to the number of SDS molecules interacting with the albumin under the same conditions (pH 5) at 298.15 K. This phenomenon indicates that lowering the temperature (288.15 K) favours the formation of a kinetically stable complex of the stoichiometry SDS:BSA equal to ca. 6.5:1, at pH 5.

The stability of the investigated SDS-BSA complexes, both in the first and in the second protein binding sites, is governed by the enthalpy factor (Table 1). The negative values of ΔH_{ITC} point to the fact that mainly the charge–charge electrostatic interactions established between negatively charged SDS anions and positively charged functional groups of amino acids such as lysine, arginine, histidine as well as van der Waals interactions and/or hydrogen bonds [42,43] involved in SDS–protein binding events, play a pivotal role in the stabilization of the resulting complexes. The release of a greater amount of energy on account of SDS-BSA interactions in the first binding site can be explained by the participation of a larger number of bonds or the presence of stronger interactions compared to the second binding site. Moreover, the presence of a cavity in the BSA protein to which SDS molecules fit spatially (sterically) could be taken into considerations to explain the differences

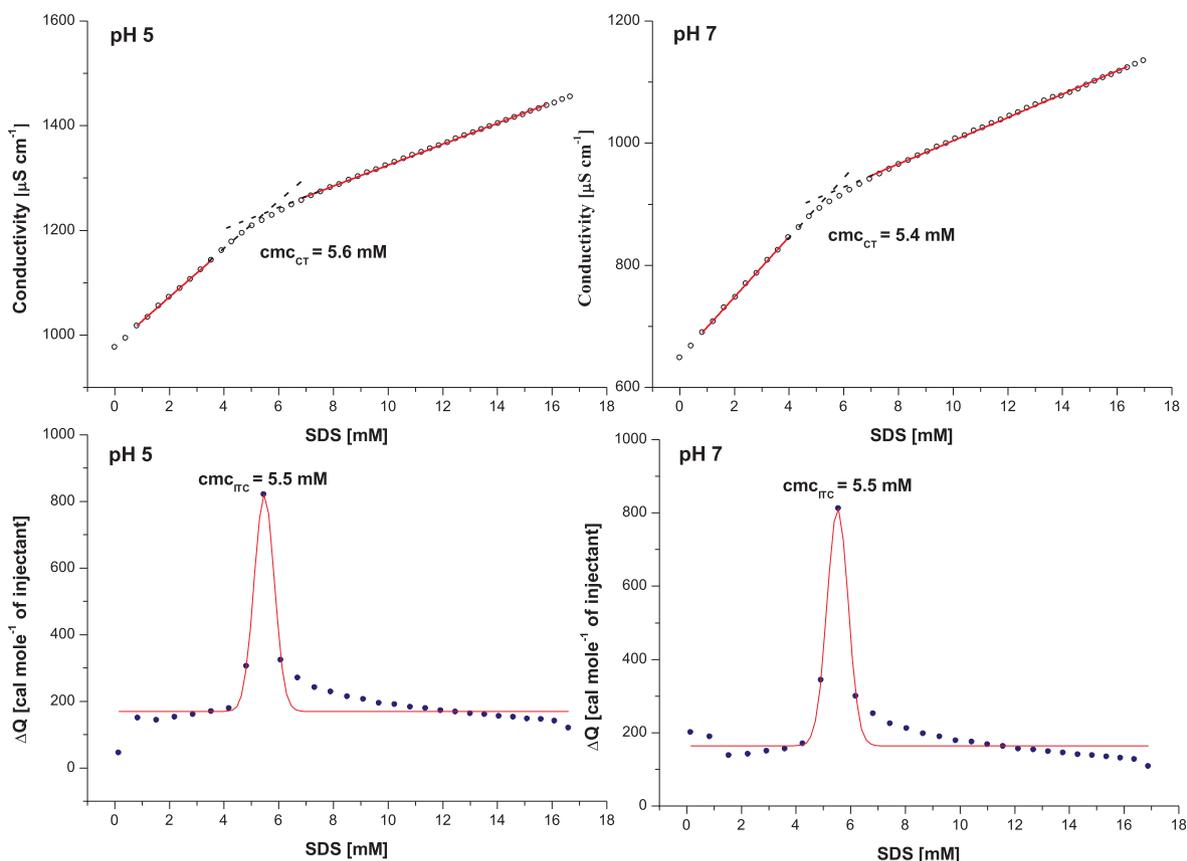


Fig. 1. The electrical conductivity (χ) and enthalpy change (ΔQ) per mole of SDS as a function of SDS concentration recorded in the 10 mM Caco buffers of pH 5 and 7, at 298.15 K. The CMC_{CT} and CMC_{ITC} denote CMC determined from conductometric titrations and ITC measurements, respectively.

in the enthalpy changes corresponding to the binding interactions in the first and second binding sites. Furthermore, the enthalpy – entropy compensation phenomenon is observed in the SDS-BSA interactions. The stronger interactions (the more negative value of ΔH) the lower mobility of the interacting molecules which is reflected by the low value of the entropic factor (Table 1) [44–46].

3.3. Steady-state fluorescence spectroscopy

The presence of two tryptophan residues in the BSA structure, namely Trp-134 and Trp-212 enables to exploit fluorescence spectroscopy for studying the BSA – ligand interactions [47–50]. The fluorescence emission spectra of BSA recorded in the presence of various SDS concentrations, using protein excitation at 275 nm, revealed that with the increasing concentration of the surfactant the fluorescence intensity decreases steadily with a blue shift of the maximum of emission wavelength. It indicates the presence of several different species in solution and is consistent with the fact that the change in the environment of tryptophan residues takes place and an increase of hydrophobicity in the vicinity of this residue occurs. The changes in the Trp fluorescence intensity are pronounced in the SDS concentration range corresponding to the binding of the substrate molecules in the first site of albumin (Figs. 3 and 4). After the first binding site is saturated only slight changes in the fluorescence intensity are observed with the increase of the SDS concentration. Thus, it can be assumed that the higher affinity BSA site is located close to the Trp-134 or Trp-212 residue which is sensitive to SDS binding to this site. Considering that Trp-134 (located in sub-domain IA) is exposed on the surface of albumin whereas a hydrophobic binding pocket com-

prises Trp-212 (located in sub-domain IIA), the Trp-134 containing binding site seems to be a potential first binding site.

The binding constants of the SDS-BSA complexes in the first binding site were calculated from spectrofluorimetric data using the Scatchard equation [51]:

$$\log\left(\frac{F_0-F}{F}\right) = \log K + n \log [Q]$$

where F_0 and F are the fluorescence intensities in the absence and presence of a quencher, n denotes the number of binding sites, while $[Q]$ is the quencher (SDS) concentration. The representative plots of $\log\left(\frac{F_0-F}{F}\right)$ vs. $\log [Q]$ for SDS/BSA systems under experimental conditions and the calculated parameters of the interactions ($\log K$ and n) are presented in Fig. 5. The binding constants obtained at the same temperature, but different pH values are comparable to one other (within the range of the experimental error). This proves that the change in pH of the system does not affect the affinity of SDS towards the first site of the protein. On the other hand, binding constants determined at 288.15 K are higher than those at 298.15 K. The increase of temperature favors the dissociation of weakly bound substrates leading to the reduction of the thermodynamic stability of SDS-BSA complexes. This, in turns, results in lower binding constant values [52].

3.4. Differential scanning calorimetry (DSC)

The BSA thermal stability in the presence and absence of SDS was assessed based on the transition midpoint T_m (Table 2). DSC measurements were carried out for the systems containing buffered solutions of pure BSA (control) and the SDS:BSA mixtures with the surfactant/protein molar ratios corresponding to the sat-

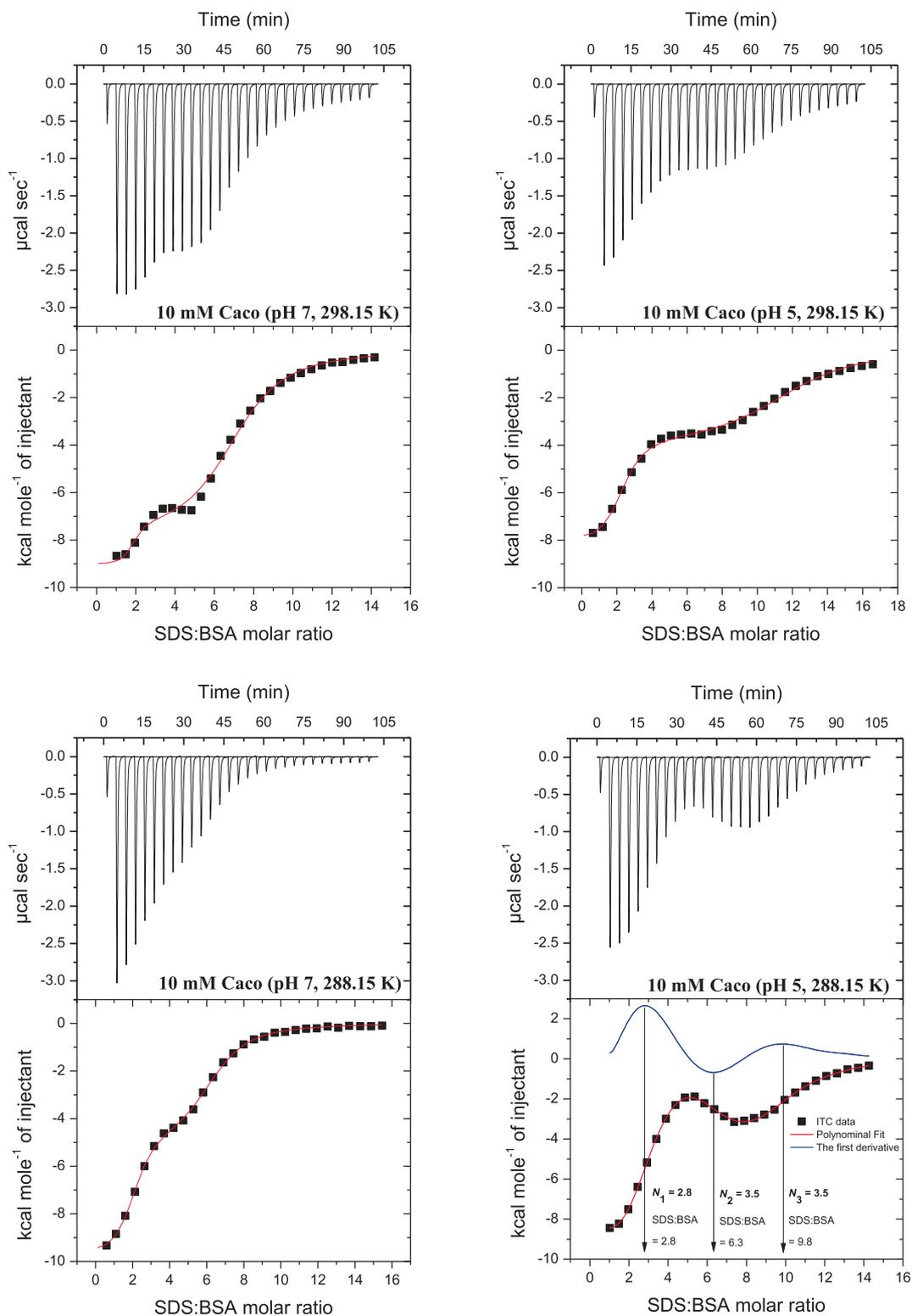


Fig. 2. Calorimetric titration isotherms of the binding interactions between SDS and BSA in the 10 mM Caco buffers of pH 5 and 7, at 288.15 K and 298.15 K.

uration of the first (4:1) and the second (8:1 and 14:1) binding sites of the albumin. The T_m values of BSA under experimental conditions were estimated by fitting the experimental C_p vs. T curves to a non-two state model using Marquardt non-linear least squares method.

For all investigated systems the irreversible thermal unfolding was observed. Furthermore, the lack of negative peaks in the

heat-capacity curves has indicated that all tested samples do not aggregate over the experimental temperature range [53]. The thermal stability of BSA at pH 5 which is close to the isoelectric point ($pI = 4.7$) [54] is slightly higher than at pH 7 (Table 2). The shift of T_m towards lower temperature with the increase of pH is due to the increase of the negative net charge of the protein on account of proton ionizations. In consequence,

Table 1

The conditional thermodynamic parameters of SDS monomers binding to BSA (standard deviation values in parentheses) in the 10 mM Caco buffers of pH 5 and 7, at 288.15 K and 298.15 K.

Parameter	Caco buffer (pH 7)		Caco buffer (pH 5)
	288.15 K	298.15 K	298.15 K
N_1	2.02 (± 0.04)	1.75 (± 0.14)	2.16 (± 0.07)
$\log K_{(\text{ITC})1}$	6.87 (± 0.06)	7.61 (± 0.38)	6.96 (± 0.11)
$\Delta G_{(\text{ITC})1}$ [kcal mol ⁻¹]	-9.08 (± 0.08)	-10.38 (± 0.67)	-9.49 (± 0.15)
$\Delta H_{(\text{ITC})1}$ [kcal mol ⁻¹]	-9.75 (± 0.12)	-9.11 (± 0.26)	-8.17 (± 0.21)
$T\Delta S_{(\text{ITC})1}$ [kcal mol ⁻¹]	-0.67	1.27	1.32
N_2	4.24 (± 0.05)	5.33 (± 0.29)	9.32 (± 0.18)
$\log K_2$	5.36 (± 0.03)	5.29 (± 0.06)	5.21 (± 0.07)
$\Delta G_{(\text{ITC})2}$ [kcal mol ⁻¹]	-7.07 (± 0.04)	-7.22 (± 0.09)	-7.11 (± 0.09)
$\Delta H_{(\text{ITC})2}$ [kcal mol ⁻¹]	-5.11 (± 0.16)	-7.98 (± 0.27)	-3.95 (± 0.13)
$T\Delta S_{(\text{ITC})2}$ [kcal mol ⁻¹]	1.96	-0.76	3.16

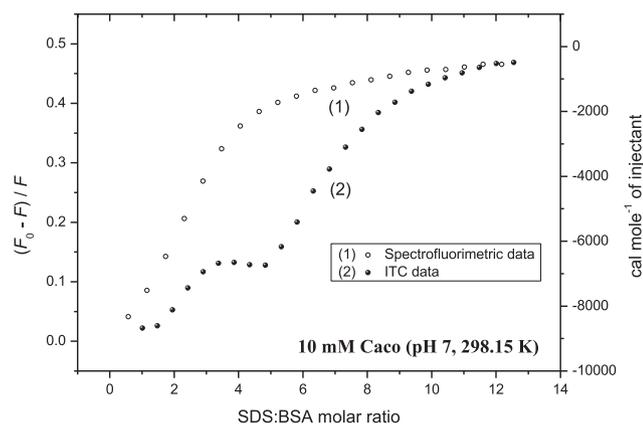


Fig. 3. The plots of $\frac{F_0-F}{F}$ (1) and integrated heats (2) vs. the SDS:BSA molar ratio in 10 mM Caco (pH 7), at 298.15 K.

the increase in the repulsive electrostatic interactions between the negatively charged functional groups of the side chains favours the BSA unfolding. The saturation of the first binding site (SDS:BSA = 4:1 M ratio system) improves the thermal stability of the system by ca. 10 K in comparison with the control (BSA buffered solution with no ligands). It should be highlighted that regardless of a pH of a system, the effect of SDS on a thermal unfolding of BSA after the saturation of the second binding site (SDS:BSA = 8:1 and 14:1 M ratio system) is not as pronounced as for the saturation of the first one. Thus, it can be supposed that the first binding site

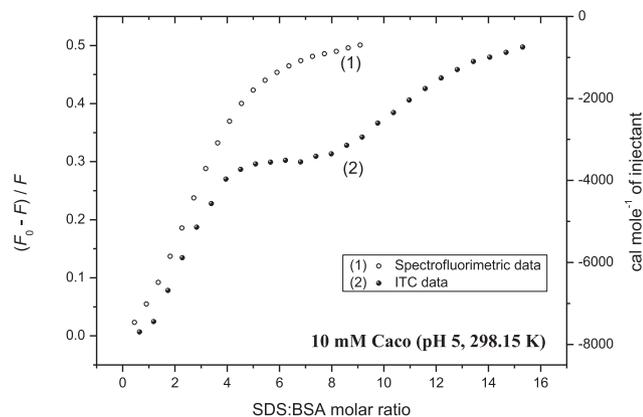


Fig. 4. The plots of $\frac{F_0-F}{F}$ (1) and integrated heats (2) vs. the SDS:BSA molar ratio in 10 mM Caco (pH 5), at 298.15 K.

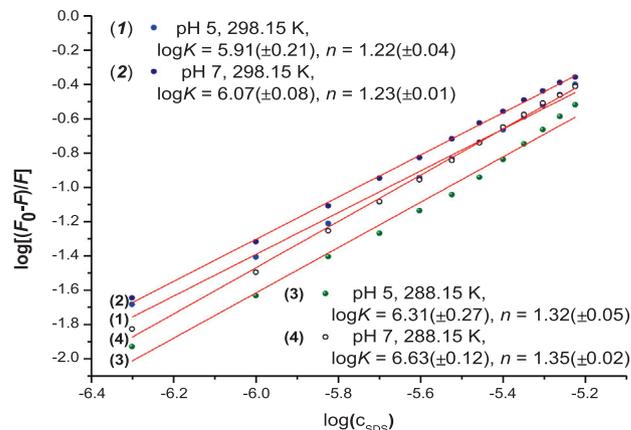


Fig. 5. The plots of $\log\left(\frac{F_0-F}{F}\right)$ vs. $\log(c_{\text{SDS}})$ for the BSA/SDS systems in the 10 mM Caco buffers of pH 5 and 7, at 288.15 K and 298.15 K (standard deviation values of $\log K$ and n in parentheses).

is more susceptible for the conformational changes upon the temperature increase.

3.5. Circular dichroism (CD)

CD spectra of the investigated protein in the absence and presence of SDS enabled to get some insight into the thermal stability of the BSA secondary structure (Supplementary Information Figs. S1 – S7). The percentage contents of α -helix, β -structure, β -turns and random coil were collected in Table 3.

It has been found that in the presence of SDS (the SDS:BSA molar ratio = 4:1, 8:1 and 14:1) the protein remains predominantly an α -helix structure at 298.15 K. Thus, under the experimental conditions the interactions of SDS with BSA do not alter the intermolecular forces maintaining the BSA structure. In CD spectra some conformational changes in the structure of BSA are observed due to the temperature increase. It is reflected in the reduction of the negative bands at 208 nm (the $\pi \rightarrow \pi^*$ transition) and 222 nm (the $n \rightarrow \pi^*$ transition) characteristic for the α -helix structure [55–57]. For all systems studied the content of α -helix decreases gradually with the increase of temperature. However, in the presence of SDS the thermal stability of the helical conformation is higher (Table 3). Above the temperature corresponding to the transition midpoint T_m (DSC results) a more prominent decrease in the α -helix content is observed, regardless the SDS:BSA molar ratio (Table 3). This is in lines with the results obtained from DSC measurements and confirms the fact that the saturation of the first binding site of the protein by SDS molecules is sufficient to noticeably increase its thermal stability.

3.6. Molecular modelling

To gain a better understanding of the experimental results at the atomistic level, docking simulations were performed. As a first step Autodock3 (AD3) was employed to localize potential binding

Table 2

The transition midpoints T_m for BSA (0.015 mM) in the presence of different concentrations of SDS in the 10 mM Caco buffers of pH 5 and 7.

SDS:BSA molar ratio	T_m [K] 10 mM Caco, pH 7	T_m [K] 10 mM Caco, pH 5
Control – no SDS	338.15	~342.15
4:1	349.15	351.15
8:1	352.15	353.15
14:1	–	~353.15

Table 3

The secondary structure content (in %) of BSA at different SDS:BSA molar ratios and different pH in 10 mM Caco buffer in the temperature range 298.15–368.15 K revealed from CD measurements.

Temperature	pH 7			pH 5			
	SDS:BSA 0:1	SDS:BSA 4:1	SDS:BSA 8:1	SDS:BSA 0:1	SDS:BSA 4:1	SDS:BSA 8:1	SDS:BSA 14:1
	The percentage content of α -helix/ β -structure/ β -turn/random coil [%]						
298.15 K	62.3/3.5/ 11.1/20.4	62.2/7.2/ 10.3/19.2	62.6/8.2/ 10.3/18.9	57.3/8.0/ 12.1/21.9	60.8/8.3/ 11.9/19.0	62.5/7.4/ 10.0/21.5	61.8/6.9/ 10.4/21.0
308.15 K	61.1/4.6/ 12.4/22.2	62.1/8.6/ 9.8/19.7	61.8/6.9/ 12.0/19.7	57.9/9.4/ 11.2/22.0	61.8/6.8/ 10.7/21.5	61.9/5.3/ 11.0/21.9	62.1/5.8/ 10.6/21.4
318.15 K	58.7/5.7/ 14.0/24.1	61.6/7.3/ 10.2/20.7	61.2/7.4/ 10.9/23.0	55.8/9.6/ 13.1/21.9	59.0/14.0/ 10.9/15.4	61.0/5.7/ 11.5/22.1	61.4/5.1/ 11.4/21.9
328.15 K	52.1/7.9/ 16.6/26.2	59.2/8.9/ 10.4/21.9	60.6/7.5/ 10.8/21.5	52.3/10.3/ 14.1/23.5	58.8/10.2/ 11.9/19.9	61.0/5.6/ 11.5/22.6	61.1/5.4/ 11.8/22.0
338.15 K	42.6/12.6/ 17.9/27.9	53.8/7.4/ 13.6/25.5	58.4/6.5/ 12.1/23.7	46.2/10.8/ 15.8/27.6	55.1/16.2/ 12.4/16.5	58.8/6.6/ 12.0/23.2	60.5/5.4/ 11.5/22.8
348.15 K	37.1/14.4/ 19.4/30.4	40.8/12.1/ 17.7/29.5	40.9/13.0/ 17.3/29.2	22.7/24.5/ 22.4/30.5	37.7/17.5/ 17.7/27.3	45.3/11.2/ 16.8/26.9	50.7/8.9/ 14.8/25.5
358.15 K	27.0/18.8/ 22.5/33.4	36.5/14.7/ 20.5/28.0	34.9/15.5/ 19.5/29.6	9.5/34.1/ 24.8/31.5	15.2/30.4/ 22.7/30.9	10.2/34.5/ 24.1/30.0	7.9/36.7/ 23.5/30.9
368.15 K	23.7/23.2/ 22.9/31.4	27.3/21.5/ 21.6/28.7	31.8/17.3/ 22.3/28.3	6.8/35.2/ 25.2/31.5	7.9/32.5/ 25.5/33.2	9.0/34.8/ 23.2/32.0	6.7/36.5/ 24.6/31.6

sites at two different pH values (5 and 7). In each case, the top 50 structures were visualized together on the BSA protein (Fig. 6). AD3 did not reveal any differences between the two pH values and so did not help to localize any pH-specific binding site. For this reason, the structures obtained by molecular docking were not used for further analysis.

Therefore, molecular dynamics (MD) simulations were conducted for four systems (pH 5 and 7, 290 K and 300 K). This approach allowed us to identify the potential “second” binding site, which was most pronounced at lower pH and temperature (see Fig. 7).

The “first” binding site was not found in any of these computational experiments suggesting that the MD simulations were too short for the SDS molecule to enter any of the cavities close to the Trp residues. Therefore, based on steady-state fluorescence spectroscopy results of the SDS-BSA interactions as well as taking into account the known structure of HSA with myristic acid (PDB ID: 2XVW, 2.65 Å), the SDS molecule was inserted in a similar binding mode close to Trp-134. Then, a 100 ns MD simulation with additional 15 SDS molecules was run twice at two different pH values. Subsequently, binding free energy ΔG was calculated with the use of two commonly deployed methods: linear interaction energy

(LIE) and molecular mechanism Generalized Born surface area (MM-GBSA). Binding free energy (ΔG) computation can play a key role in prioritizing compounds to be evaluated experimentally on their affinity for target proteins.

The data showed that the “inserted” SDS molecule and a neighbouring one approaching it during the MD simulation had the lowest energies. Therefore, they were the most stable in terms of binding to the BSA protein (Supplementary Information Tables S1 and S2) when compared to the ΔG values for other SDS molecules including the ones corresponding to the putative “second” binding site.

The binding free energy values for the “first” binding sites were: -14.6 ± 3.3 and -33.4 ± 6.0 whereas for the “second” binding site they equalled: -8.7 ± 4.6 and -19.8 ± 10.1 (MM-GBSA and LIE respectively, values given in kcal/mol). Based on these results, not only the “first” binding site was localized but also it was shown that this site is pH-independent when the total binding free energy is considered (this does not apply to the free energy components though). Van der Waals free energy component plays decisive role for binding in this site. This finding supports the thermodynamic parameters (the negative value of ΔH_{ITC}) for the SDS-BSA interactions obtained from ITC experiments.

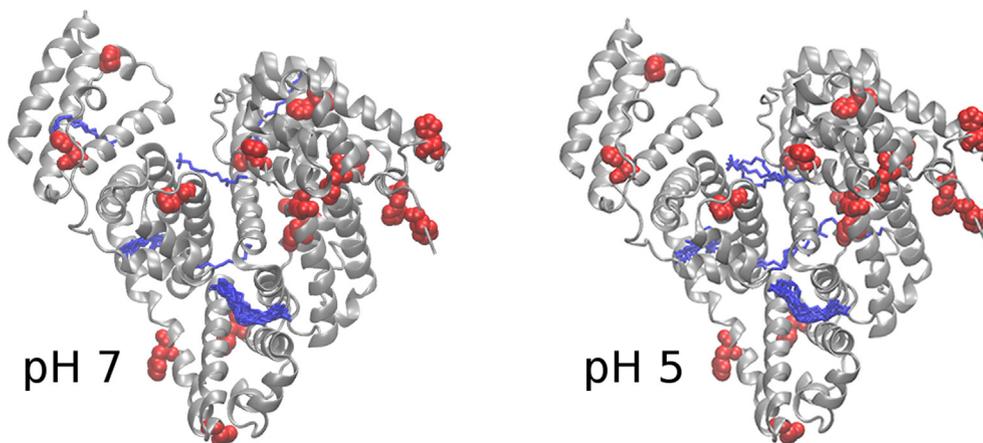


Fig. 6. Top 50 AD3 docking poses (in blue licorice) for docking SDS to BSA (in grey cartoon) under pH 7 (left) and pH 5 (right). HIS residues are represented with red VDW spheres.

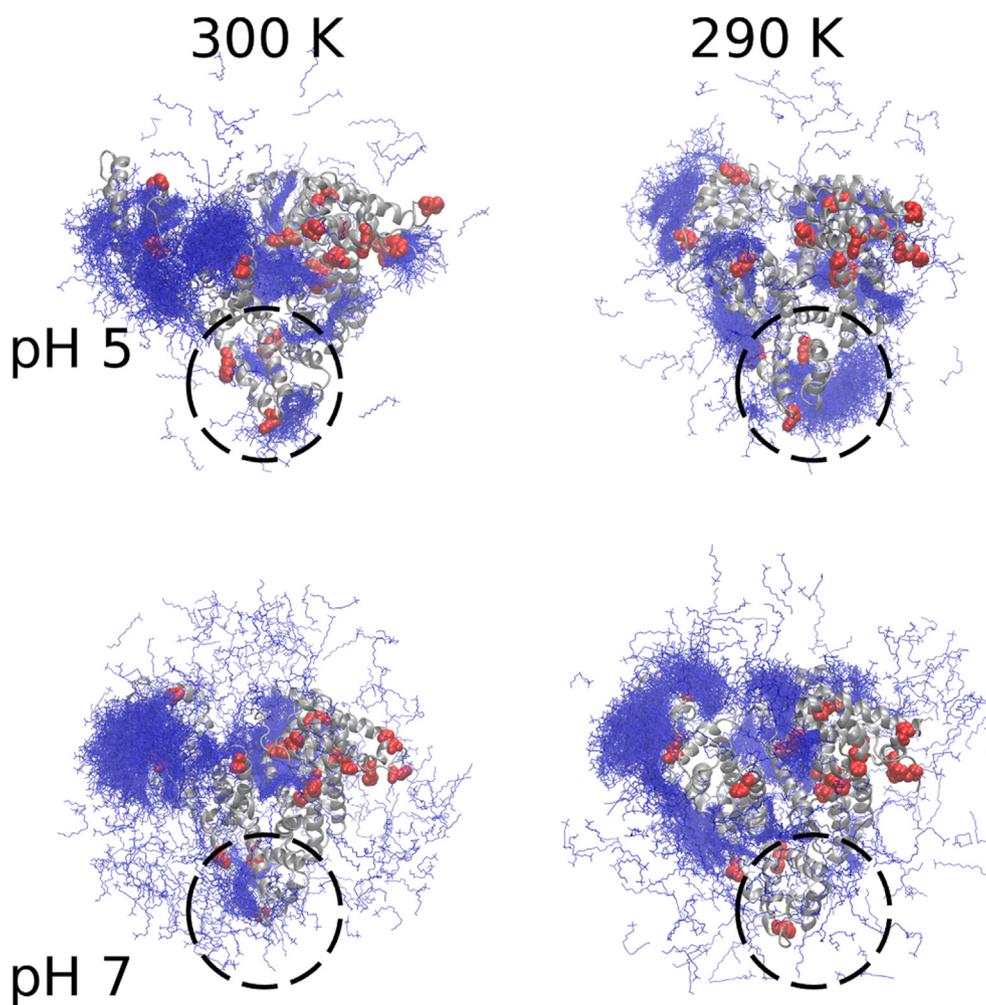


Fig. 7. MD-based analysis of SDS-BSA complexes at different pH and temperatures. BSA protein: grey cartoon; HIS residue: red VDW spheres; SDS molecules after the MD converged (frames are shown each 100 ps): blue licorice; the potential “second” binding site is shown with black circles. The data shown are obtained from two MD simulations.

4. Conclusions

In this study, we showed that a low concentration of SDS (0.009 mg/1 mg SDS/BSA) significantly affects the BSA structure, which may have an impact on its biological functions. It should be stressed that the binding constant and the stoichiometry of the resulting BSA complexes are so-called condition-dependent parameters as their values often depend on experimental conditions (pH and temperature). Furthermore, the determination of the binding properties of BSA is not always straightforward, especially when multiple SDS molecules are potentially bound. In such cases the effect of the SDS competition with other small ligands of BSA should be taken into consideration. At the same time, this phenomenon can be exploited to modulate the binding properties of BSA. Finally, the MD-based computational analysis corroborated the experimental findings and provided atomistic details explaining the differences on the binding of SDS molecules in the high-affinity pH-independent and the low-affinity pH-dependent binding sites.

CRedit authorship contribution statement

Aleksandra Tesmar: Conceptualization, Data curation. **Małgorzata M. Kogut:** Methodology, Investigation. **Krzysztof Żamojć:**

Methodology, Investigation. **Ola Grabowska:** Investigation, Resources. **Katarzyna Chmur:** Investigation, Resources. **Sergey A. Samsonov:** Methodology, Supervision. **Joanna Makowska:** Methodology, Investigation. **Dariusz Wyrzykowski:** Conceptualization, Project administration, Writing - review & editing. **Lech Chmurzyński:** Supervision, 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.

Acknowledgements

This work was supported by the Polish National Science Centre under Grant No. 2016/23/D/ST4/01576. Theoretical research was funded by National Science Centre of Poland, grant number UMO-2018/30/E/ST4/00037. Computational resources were provided by the Polish Grid Infrastructure (PL-GRID, grants plgions-gpu, plggagstr2), ZIH at TU Dresden (grant p_gag) and the local cluster PIASEK at the Faculty of Chemistry, University of Gdansk.

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.molliq.2021.117185>.

References

- [1] A. Taha, E. Ahmed, A. Ismaiel, M. Ashokkumar, X. Xu, S. Pan, H. Hu, Ultrasonic emulsification: an overview on the preparation of different emulsifiers-stabilized emulsions, *Trends Food Sci. Technol.* 105 (2020) 363–377.
- [2] T. Song, F. Gao, S. Guo, Y. Zhang, S. Li, H. You, Y. Du, A review of the role and mechanism of surfactants in the morphology control of metal nanoparticles, *Nanoscale* 13 (7) (2021) 3895–3910.
- [3] J. Song, C. Sun, K. Gul, A. Mata, Y. Fang, Prolamin-based complexes: Structure design and food-related applications, *Compr. Rev. Food Sci. Food Saf.* 20 (2) (2021) 1120–1149.
- [4] Z. Yue, M. Yao, G. Bai, J. Wang, K. Zhuo, J. Wang, Y. Wang, Controllable enzymatic superactivity α -chymotrypsin activated by the electrostatic interaction with cationic gemini surfactants, *RSC Adv.* 11 (2021) 7294–7304.
- [5] J. Aslam, I. Hussain Lone, F. Ansari, A. Aslam, R. Aslam, M. Akram, Molecular binding interaction of pyridinium based gemini surfactants with bovine serum albumin: Insights from physicochemical, multispectroscopic, and computational analysis, *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* 250 (2021) 119350, <https://doi.org/10.1016/j.saa.2020.119350>.
- [6] V. Sharma, O. Yañez, C. Zúñiga, A. Kumar, G. Singh, P. Cantero-López, Protein-surfactant interactions: a multitechnique approach on the effect of Co-solvents over bovine serum albumin (BSA)-cetyl pyridinium chloride (CPC) system, *Chem. Phys. Lett.* 747 (2020) 137349.
- [7] K.K. Andersen, C.L. Oliveira, K.L. Larsen, F.M. Poulsen, T.H. Callisen, P. Westh, J. S. Pedersen, D. Otzen, The role of decorated SDS micelles in sub-cmc protein denaturation and association, *J. Mol. Biol.* 391 (1) (2009) 207–226.
- [8] D.E. Otzen, P. Sehgal, P. Westh, α -Lactalbumin is unfolded by all classes of detergents but with different mechanisms, *J. Coll. Int. Sci.* 329 (2009) 273–283.
- [9] D.E. Otzen, L. Nesgaard, K.K. Andersen, J.H. Hansen, G. Christiansen, H. Doe, P. Sehgal, Aggregation of S6 in a quasi-native state by monomeric SDS, *Biochim. Biophys. Acta* 1784 (2008) 400–414.
- [10] K.K. Andersen, P. Westh, D.E. Otzen, A global study of myoglobin-surfactant interactions, *Langmuir* 24 (2) (2008) 399–407.
- [11] M.M. Nielsen, K.K. Andersen, P. Westh, D.E. Otzen, Unfolding of β -sheet proteins in SDS, *Biophys. J.* 92 (10) (2007) 3674–3685.
- [12] D. Otzen, Protein-surfactant interactions: a tale of many states, *Biochim. Biophys. Acta* 1814 (5) (2011) 562–591.
- [13] M. Aguirre-Ramírez, H. Silva-Jiménez, I.M. Banat, M.A. Díaz De Rienzo, Surfactants: physicochemical interactions with biological macromolecules, *Biotechnol. Lett.* 43 (2021) 523–535.
- [14] S. De, A. Girigoswami, S. Das, Fluorescence probing of albumin-surfactant interaction, *J. Colloid Interface Sci.* 285 (2) (2005) 562–573.
- [15] D. Kelley, D.J. McClements, Interactions of bovine serum albumin with ionic surfactants in aqueous solutions, *Food Hydrocoll.* 17 (1) (2003) 73–85.
- [16] E.L. Gelamo, M. Tabak, Spectroscopic studies on the interaction of bovine (BSA) and human (HSA) serum albumins with ionic surfactants, *Spectrochim. Acta A* 56 (11) (2000) 2255–2271.
- [17] A. Chakraborty, D. Seth, P. Setia, N. Sarkar, Photoinduced electron transfer in a protein-surfactant complex: probing the interaction of SDS with BSA, *J. Phys. Chem. B* 110 (2006) 16607–16617.
- [18] T. Chakraborty, I. Chakraborty, S.P. Moulik, S. Ghosh, Physicochemical and conformational studies on BSA-surfactant interaction in aqueous medium, *Langmuir* 25 (5) (2009) 3062–3074.
- [19] N.J. Turro, X.-G. Lei, K.P. Ananthapadmanabhan, M. Aronson, Spectroscopic probe analysis of protein-surfactant interactions: the BSA/SDS system, *Langmuir* 11 (7) (1995) 2525–2533.
- [20] A. Panuszko, P. Bruździak, J. Zieliński, D. Wyrzykowski, J. Stangret, Effects of urea and trimethylamine-N-oxide on the properties of water and the secondary structure of hen egg white lysozyme, *J. Phys. Chem. B* 113 (44) (2009) 14797–14809.
- [21] J.T. Pelton, L.R. McLean, Spectroscopic methods for analysis of protein secondary structure, *Anal. Biochem.* 277 (2) (2000) 167–176.
- [22] V. Plotnikov, A. Rochalski, M. Brandts, J.F. Brandts, S. Williston, V. Frasca, L.-N. Lin, An autosampling differential scanning calorimeter instrument for studying molecular interactions, *Assay Drug Dev Technol.* 1 (1) (2002) 83–90.
- [23] S.W. Provencher, J. Glöckner, Estimation of protein secondary structure from circular dichroism, *Biochemistry* 20 (1981) 33–37.
- [24] D. A. Case, D. S. Cerutti, T. E. III Cheatham, T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Goetz, D. Greene, N. Homeyer, S. Izadi, A. Kovalenko, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. Mermelstein, K. M. Merz, G. Monard, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D. R. Roe, A. Roitberg, C. Sagui, C. L. Simmerling, W. M. Botello-Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, R. X. Wu, L. Xiao, D. M. York, P. A. Kollman, AMBER 2017; University of California: San Francisco.
- [25] G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson, Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, *J. Comput. Chem.* 19 (14) (1998) 1639–1662.
- [26] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J.L. Klepeis, R.O. Dror, D.E. Shaw, Improved side-chain torsion potentials for the Amber ff99SB protein force field, *Proteins* 78 (2010) 1950–1958.
- [27] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing of a general Amber force field, *J. Comput. Chem.* 25 (9) (2004) 1157–1174.
- [28] C.I. Bayly, P. Cieplak, W. Cornell, P.A. Kollman, A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model, *J. Phys. Chem.* 97 (40) (1993) 10269–10280.
- [29] V. Tsui, D.A. Case, Theory and applications of the generalized Born solvation model in macromolecular simulations, *Biopolymers* 56 (4) (2000) 275–291.
- [30] A. Onufriev, D. Bashford, D.A. Case, Modification of the generalized Born model suitable for macromolecules, *J. Phys. Chem. B* 104 (15) (2000) 3712–3720.
- [31] N. Scholz, T. Behnke, U. Resch-Genger, Determination of the critical micelle concentration of neutral and ionic surfactants with fluorometry, conductometry, and surface tension—a method comparison, *J. Fluoresc.* 28 (1) (2018) 465–476.
- [32] S. Wu, F. Liang, D. Hu, H. Li, W. Yang, Q. Zhu, Development of determining the critical micelle concentration of surfactants by simple and fast titration method, *Anal. Chem.* 92 (2020) 4259–4265.
- [33] M.D. Lad, V.M. Ledger, B. Briggs, R.J. Green, R.A. Frazier, Analysis of the SDS-lysozyme binding isotherm, *Langmuir* 19 (12) (2003) 5098–5103.
- [34] J. Mikšovská, J. Yom, B. Diamond, R.W. Larsen, Spectroscopic and photothermal study of myoglobin conformational changes in the presence of sodium dodecyl sulfate, *Biomacromolecules* 7 (2) (2006) 476–482.
- [35] S. Ghosh, A. Banerjee, A Multitechnique approach in protein/surfactant interaction study: physicochemical aspects of sodium dodecyl sulfate in the presence of trypsin in aqueous medium, *Biomacromolecules* 3 (2002) 9–16.
- [36] E. Fuguet, C. Ràfols, M. Rosés, E. Bosch, Critical micelle concentration of surfactants in aqueous buffered and unbuffered systems, *Anal. Chim. Acta* 548 (1–2) (2005) 95–100.
- [37] J.R. Liley, R.K. Thomas, J. Penfold, I.M. Tucker, J.T. Petkov, P. Stevenson, J.R.P. Webster, Impact of electrolyte on adsorption at the air–water interface for ternary surfactant mixtures above the critical micelle concentration, *Langmuir* 33 (17) (2017) 4301–4312.
- [38] A. Zdziennicka, K. Szymczyk, J. Krawczyk, B. Jańczuk, Critical micelle concentration of some surfactants and thermodynamic parameters of their micellization, *Fluid Ph. Equilibria* 322–323 (2012) 126–134.
- [39] R. Patel, G. Buckton, S. Gaisford, The use of isothermal titration calorimetry to assess the solubility enhancement of simvastatin by a range of surfactants, *Thermochim. Acta* 456 (2) (2007) 106–113.
- [40] P.C. Griffiths, N. Hirst, A. Paul, S.M. King, R.K. Heenan, R. Farley, Effect of ethanol on the interaction between poly(vinylpyrrolidone) and sodium dodecyl, *Langmuir* 20 (2004) 6904–6913.
- [41] M.L. Corrin, W.D. Harkins, The effect of salts on the critical concentration for the formation of micelles in colloidal electrolytes, *J. Am. Chem. Soc.* 69 (1947) 683.
- [42] S.K. Panigrahi, G.R. Desiraju, Strong and weak hydrogen bonds in the protein–ligand interface, *Proteins* 67 (1) (2007) 128–141.
- [43] P. Chakrabarti, R. Bhattacharyya, Geometry of nonbonded interactions involving planar groups in proteins, *Prog. Biophys. Mol. Biol.* 95 (1–3) (2007) 83–137.
- [44] D.L. Cramer, B. Cheng, J. Tian, J.H. Clements, R.M. Wypych, S.F. Martin, Some thermodynamic effects of varying nonpolar surfaces in protein ligand interactions, *European J. Med. Chem.* 208 (2020) 112771.
- [45] X. Du, Yi. Li, Y.-L. Xia, S.-M. Ai, J. Liang, P. Sang, X.-L. Ji, S.-Q. Liu, Insights into Protein–Ligand Interactions: Mechanisms, Models, and Methods, *Int. J. Mol. Sci.* 17 (2016) 144.
- [46] George M. Whitesides, Vijay M. Krishnamurthy, Designing ligands to bind proteins, *Q. Rev. Biophys.* 38 (4) (2005) 385–395.
- [47] Zhiyong Tian, Luyao Tian, Man Shi, Sihan Zhao, Shudi Guo, Wen Luo, Chaojie Wang, Zhihui Tian, Investigation of the interaction of a polyamine-modified flavonoid with bovine serum albumin (BSA) by spectroscopic methods and molecular simulation, *J. Photochem. Photobiol. B: Biology* 209 (2020) 111917, <https://doi.org/10.1016/j.jphotobiol.2020.111917>.
- [48] B.L. Wang, S.B. Kou, Z.Y. Lin, J.H. Shi, Investigation on the binding behavior between BSA and lenvatinib with the help of various spectroscopic and in silico methods, *J. Mol. Struct.* 1204 (2020) 127521.
- [49] X. Wei, Y. Yang, J. Ge, X. Lin, D. Liu, S. Wang, J. Zhang, G. Zhou, S. Li, Synthesis, characterization, DNA/BSA interactions and in vitro cytotoxicity study of palladium(II) complexes of hispolon derivatives, *J. Inorg. Biochem.* 202 (2020) 110857.
- [50] D. Yinhu, M.M. Foroughi, Z. Aramesh-Boroujeni, S. Jahani, M. Peydayesh, F. Borhani, M. Khatami, M. Rohani, M. Dusekj, V. Eigner, The synthesis, characterization, DNA/BSA interactions, molecular modeling, antibacterial properties, and in vitro cytotoxic activities of novel parent and niosome nano-encapsulated Ho(III) complexes, *RSC Adv.* 10 (39) (2020) 22891–22908.
- [51] K. Butowska, K. Żamojć, M. Kogut, W. Kozak, D. Wyrzykowski, W. Wiczak, J. Czub, J. Piosik, J. Rak, The Product of matrix metalloproteinase cleavage of doxorubicin conjugate for anticancer drug delivery: calorimetric, spectroscopic, and molecular dynamics studies on peptide–doxorubicin binding to DNA, *Int. J. Mol. Sci.* 21 (18) (2020) 6923.
- [52] Krzysztof Żamojć, Wiesław Wiczak, Lech Chmurzyński, The influence of the type of substituents and the solvent on the interactions between different

- coumarins and selected TEMPO analogues—Fluorescence quenching studies, *Chem. Phys.* 513 (2018) 188–194.
- [53] Wojciech Dzwolak, Revanur Ravindra, Roland Winter, Hydration and structure—the two sides of the insulin aggregation process, *Phys. Chem. Chem. Phys.* 6 (8) (2004) 1938–1943.
- [54] A. Salis, M. Boström, L. Medda, F. Cugia, B. Barse, D.F. Parsons, B.W. Ninham, M. Monduzzi, Conformational behavior of chemically reactive alanine-rich repetitive protein polymers, *Langmuir* 27 (2011) 11597–11604.
- [55] John A. Schellman, Patrick Oriel, Origin of the Cotton effect of helical polypeptides, *J. Chem. Phys.* 37 (9) (1962) 2114–2124.
- [56] R.W. Woody, I. Jr. Tinoco, Optical Rotation of Oriented Helices. III. Calculation of the Rotatory Dispersion and Circular Dichroism of the Alpha- and 310-Helix, *J. Chem. Phys.* 46 (1967) 4927–4945.
- [57] N. Sreerama, R. W. Woody, Circular Dichroism of Peptides and Proteins. In *Circular Dichroism: Principles and Applications* (Berova, N., Nakanishi, K., and Woody, R. W., Eds.) 2nd ed., pp. 601– 620. Wiley, New York (2000).

PUBLICATION D5

**EFFECT OF TETRAPHENYLBORATE ON
PHYSICOCHEMICAL PROPERTIES OF BOVINE SERUM
ALBUMIN**

Article

Effect of Tetraphenylborate on Physicochemical Properties of Bovine Serum Albumin

 Ola Grabowska [†], Małgorzata M. Kogut [†], Krzysztof Żamojć , Sergey A. Samsonov , Joanna Makowska , Aleksandra Tesmar , Katarzyna Chmur , Dariusz Wyrzykowski ^{*}  and Lech Chmurzyński

Faculty of Chemistry, University of Gdańsk, Wita Stwosza 63, 80-308 Gdańsk, Poland; ola.grabowska@phdstud.ug.edu.pl (O.G.); malgorzata.kogut@phdstud.ug.edu.pl (M.M.K.); krzysztof.zamojc@ug.edu.pl (K.Ż.); sergey.samsonov@ug.edu.pl (S.A.S.); joanna.makowska@ug.edu.pl (J.M.); aleksandra.tesmar@ug.edu.pl (A.T.); k.chmur.820@studms.ug.edu.pl (K.C.); lech.chmurzynski@ug.edu.pl (L.C.)
^{*} Correspondence: dariusz.wyrzykowski@ug.edu.pl
[†] Authors contributed equally.



Citation: Grabowska, O.; Kogut, M.M.; Żamojć, K.; Samsonov, S.A.; Makowska, J.; Tesmar, A.; Chmur, K.; Wyrzykowski, D.; Chmurzyński, L. Effect of Tetraphenylborate on Physicochemical Properties of Bovine Serum Albumin. *Molecules* **2021**, *26*, 6565. <https://doi.org/10.3390/molecules26216565>

Academic Editors: Danuta Witkowska and Simona Collina

Received: 24 September 2021
 Accepted: 28 October 2021
 Published: 29 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: The binding interactions of bovine serum albumin (BSA) with tetraphenylborate ions ($[\text{B}(\text{Ph})_4]^-$) have been investigated by a set of experimental methods (isothermal titration calorimetry, steady-state fluorescence spectroscopy, differential scanning calorimetry and circular dichroism spectroscopy) and molecular dynamics-based computational approaches. Two sets of structurally distinctive binding sites in BSA were found under the experimental conditions (10 mM cacodylate buffer, pH 7, 298.15 K). The obtained results, supported by the competitive interactions experiments of SDS with $[\text{B}(\text{Ph})_4]^-$ for BSA, enabled us to find the potential binding sites in BSA. The first site is located in the subdomain I A of the protein and binds two $[\text{B}(\text{Ph})_4]^-$ ions ($\log K_{(\text{ITC})1} = 7.09 \pm 0.10$; $\Delta G_{(\text{ITC})1} = -9.67 \pm 0.14 \text{ kcal mol}^{-1}$; $\Delta H_{(\text{ITC})1} = -3.14 \pm 0.12 \text{ kcal mol}^{-1}$; $T\Delta S_{(\text{ITC})1} = -6.53 \text{ kcal mol}^{-1}$), whereas the second site is localized in the subdomain III A and binds five ions ($\log K_{(\text{ITC})2} = 5.39 \pm 0.06$; $\Delta G_{(\text{ITC})2} = -7.35 \pm 0.09 \text{ kcal mol}^{-1}$; $\Delta H_{(\text{ITC})2} = 4.00 \pm 0.14 \text{ kcal mol}^{-1}$; $T\Delta S_{(\text{ITC})2} = 11.3 \text{ kcal mol}^{-1}$). The formation of the $\{[\text{B}(\text{Ph})_4]^- \}$ -BSA complex results in an increase in the thermal stability of the alpha-helical content, correlating with the saturation of the particular BSA binding sites, thus hindering its thermal unfolding.

Keywords: bovine serum albumin; sodium tetraphenylborate; sodium dodecyl sulfate; binding site; thermodynamic parameters

1. Introduction

In aqueous media, proteins undergo many physicochemical changes such as conformational alterations, denaturation, folding/unfolding processes, and ligand exchange [1–6]. These phenomena can be invoked by variations in the temperature and/or the pH of a solution. Consequently, the environmental conditions may affect the binding properties of a biomolecule, such as the affinity of low-molecular weight compounds to a protein, the number of potential binding sites, and the stoichiometry of the resulting protein–ligand complexes [7]. As a result, the biological and pharmacological activity of the protein is often highly influenced by the presence of the different ligands in the system. Generally, the nature and the concentration of some potential ligands present in the system, which are capable to compete with other compounds (for example drugs) to a biomolecule, may affect the binding properties and functions of a protein. In the field of physicochemical studies, this phenomenon is applied to the studying of the potential binding sites of BSA by competitive displacement assays [8,9]. There are many techniques used for studying the albumin–ligand interactions. However, due to the presence of two tryptophan residues in BSA, namely Trp-134 (subdomain IA) and Trp-213 (subdomain IIB) [10], fluorescence spectroscopy is the most exploited. An excellent review on the application of direct fluorescence titrations for studying serum albumin–ligand interactions provided by Macii

and Biver reveals the aspects underlying the experimental challenges and recommends new experimental design to gain some additional information on the binding type and sites of a protein [11]. It has been reported that phenylbutazone and warfarin [12] can be successfully employed as probes for determining site I (located in the hydrophobic pocket of subdomain IIA) [13] whereas diazepam and flufenamic acid [14] for site II, located in the hydrophobic cavity of subdomain IIIA.

It is also worth highlighting that albumins are one of the main objects of study in the field of bioinorganic- and metallo-drug design [15–17]. However, in the frame of biological studies, cell culture media used in cytotoxicity assays of new compounds usually contain relatively high amount of BSA (ca. 0.04 mM). This may result in the binding interaction of the tested compound with the cell incubation media component leading to the lowering the concentration of the free (active) form of the investigated substance.

We have focused our attention on bovine serum albumin (BSA), as it represents one of the most common model molecular systems for binding studies [18]. Serum albumins are involved, among others, in transport and biodistribution of endogenous ligands such as fatty acids [19] as well as endogenous substances, for example ibuprofen, warfarin, penicillin or diazepam [11]. They also participate in the metabolism of low-molecular weight compounds. Recently, we have proven that the isothermal titration calorimetry technique supported by experimental and *in silico* methods can be used as an alternative to fluorescence spectroscopy for studying a stoichiometry of the resulting albumin–ligand complexes, the type of the binding interactions, and the number of the binding sites [20]. The studies performed on BSA and sodium dodecyl sulfate (SDS) have revealed that the investigated protein possesses two main binding sites: the first site is pH and temperature independent, and binds two moles of SDS and is located close to Trp-134 residue. In contrast, the total number of SDS bound to the second site of the protein depends on the pH of the solution (pH 5 and pH 7) and the temperature (290 K and 300 K) (Figure 1) [20].

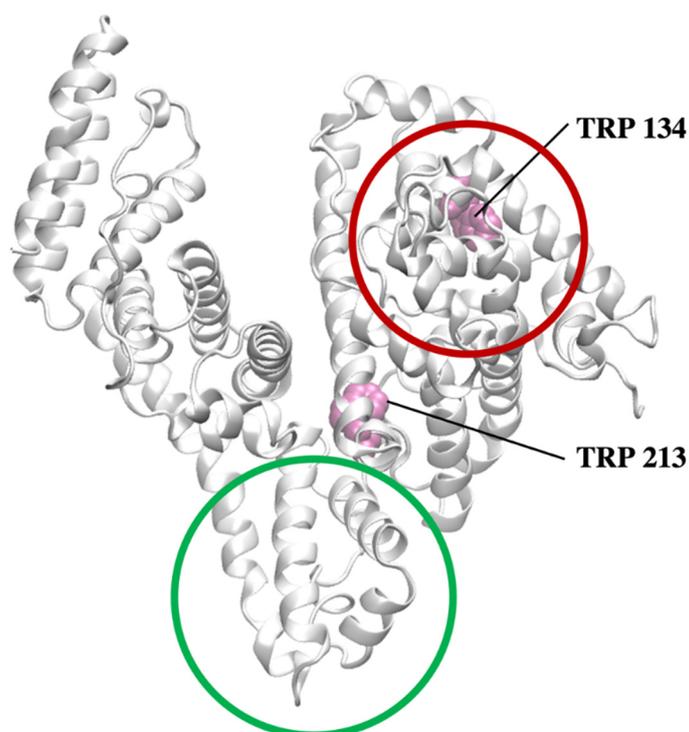


Figure 1. The secondary structure of BSA (grey cartoon) with marked Trp residues (magenta VDW representation). The first and second potential binding sites of SDS are shown with red and green rings, respectively.

In this work, we have exploited the previously gathered information on the SDS–BSA interactions for studying the binding interactions of tetraphenylborate ions ($[B(Ph)_4]^-$) with BSA. We have focused our attention on the $[B(Ph)_4]^-$ ions, as they represent the low-molecular weight compounds with four, bulky hydrophobic phenyl moieties and a negative charge capable of binding through hydrophobic interactions and/or a combination of hydrophobic and electrostatic forces [21] (Figure 2). A set of some complementary methods, namely Isothermal Titration Calorimetry (ITC), steady-state fluorescence spectroscopy, Differential Scanning Calorimetry (DSC) and Circular Dichroism (CD) was employed for assessing the physicochemical nature of the interactions. The effect of the $[B(Ph)_4]^-$ ions binding on the protein structure was also discussed. Finally, new competition ITC experiments using SDS as a competitive ligand have been performed to find BSA binding sites of $[B(Ph)_4]^-$ ions. Then, the experimental results were subsequently verified by a Molecular Dynamics (MD) study. The obtained results of the $[B(Ph)_4]^-$ –BSA interactions were discussed in relation to a previous published study on the interaction of sodium dodecyl sulfate (SDS) with BSA.

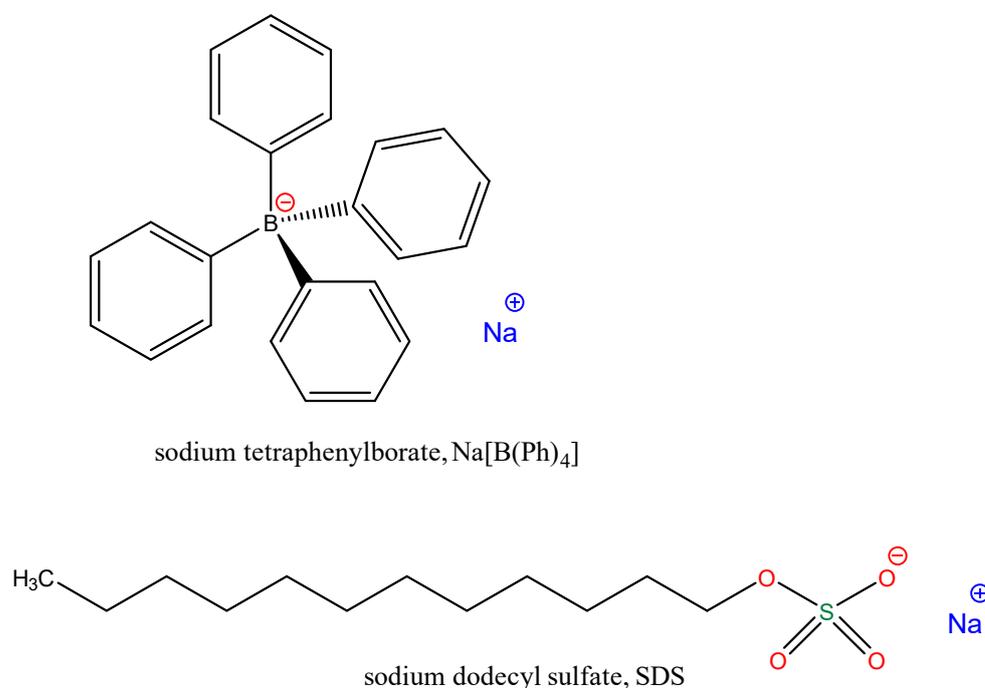


Figure 2. Chemical structure of sodium tetraphenylborate ($Na[B(Ph)_4]$) and sodium dodecyl sulfate (SDS).

2. Results and Discussion

2.1. Isothermal Titration Calorimetry (ITC)

Thermodynamic parameters of the BSA interaction with $[B(Ph)_4]^-$ ions determined by the ITC technique in the cacodylate buffer solution of a pH 7, at 298.15 K, were subsequently compared with the results obtained under the same experimental conditions for the BSA–SDS system [20]. In both investigated cases, namely $\{[B(Ph)_4]^- \}$ –BSA and SDS–BSA, nonlinear least-squares procedures were applied for fitting isotherms to a model that assumes two sets of binding sites. The stoichiometry of the ligand–albumin complexes ($N = [ligand]/[BSA]$), binding constants (K_{ITC}) and the enthalpy change (ΔH_{ITC}) were obtained directly from ITC measurements. The free energy of binding (ΔG_{ITC}) and entropy change (ΔS_{ITC}) were calculated using the standard thermodynamic relationships: $\Delta G_{ITC} = -RT \ln K_{ITC} = \Delta H_{ITC} - T \Delta S_{ITC}$. The obtained parameters are so-called condition-dependent parameters and can be compared with those obtained under the same pH, temperature and the type of a buffer solution [22]. Representative binding isotherms for

the investigated systems are shown in Figure 3, whereas conditional parameters of the interactions are summarized in Table 1.

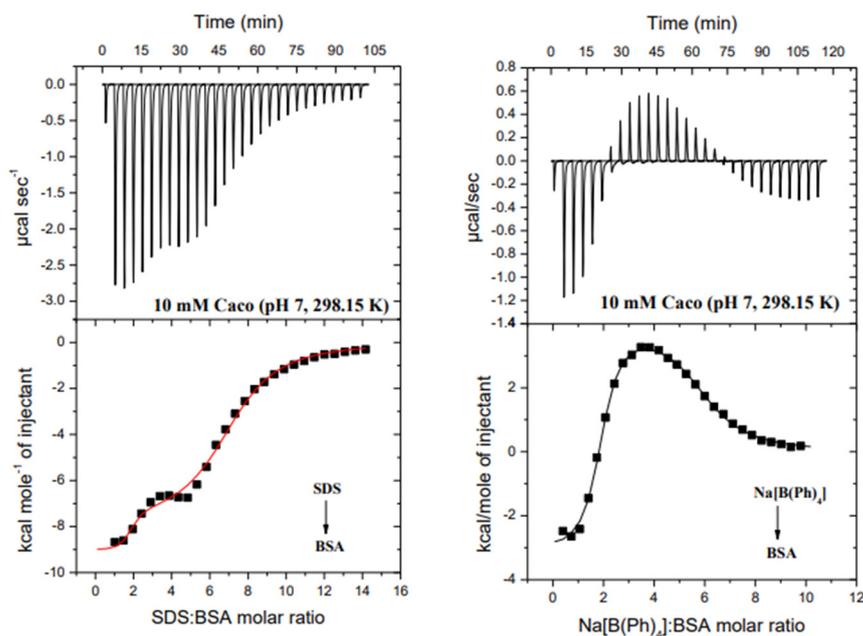


Figure 3. Calorimetric titration isotherms of the binding interactions between SDS and BSA (LEFT) and Na[B(Ph)₄] and BSA (RIGHT) in the 10 mM Caco buffer of pH 7, at 298.15 K.

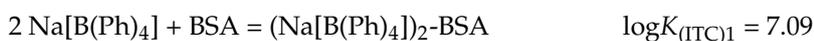
Table 1. The conditional thermodynamic parameters of SDS and Na[B(Ph)₄] binding to BSA (standard deviation values in parentheses) in the 10 mM Caco buffer of pH 7, at 298.15 K.

Parameter	SDS/BSA ⁽¹⁾	Na[B(Ph) ₄]/BSA
N_1	1.75 (±0.14)	1.78 (±0.02)
$\log K_{(ITC)1}$	7.61 (±0.38)	7.09 (±0.10)
$\Delta G_{(ITC)1}$ [kcal mol ⁻¹]	-10.38 (±0.67)	-9.67 (±0.14)
$\Delta H_{(ITC)1}$ [kcal mol ⁻¹]	-9.11 (±0.26)	-3.14 (±0.12)
$T\Delta S_{(ITC)1}$ [kcal mol ⁻¹]	1.27	6.53
N_2	5.33 (±0.29)	4.09 (±0.09)
$\log K_{(ITC)2}$	5.29 (±0.06)	5.39 (±0.06)
$\Delta G_{(ITC)2}$ [kcal mol ⁻¹]	-7.22 (±0.09)	-7.35 (±0.09)
$\Delta H_{(ITC)2}$ [kcal mol ⁻¹]	-7.98 (±0.27)	4.00 (±0.14)
$T\Delta S_{(ITC)2}$ [kcal mol ⁻¹]	-0.76	11.3

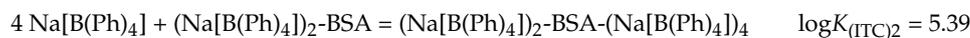
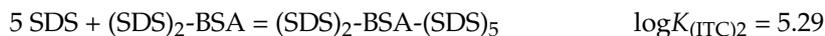
⁽¹⁾ Literature data [17].

ITC data revealed that BSA has two binding sites capable of binding tetraphenylborate ions, as has been previously observed for SDS monomers (Figure 3, Table 1). The simplified chemical equations of the ligand (SDS or [B(Ph)₄]⁻)-BSA complex formation which take into account the stoichiometry of the resulting species are given below:

The first binding site



The second binding site



The first site of albumin binds two $[\text{B(Ph)}_4]^-$ ions followed by the saturation of the second site by four tetraphenylborate ions. The binding constants ($K_{(\text{ITC})}$), and consequently the free energy of binding ($\Delta G_{(\text{ITC})}$), for both SDS–BSA and $[\text{B(Ph)}_4]^-$ –BSA complexes are comparable to one another (in the range of experimental error) indicating their similar thermodynamic stability (Table 1). In contrast to the SDS–BSA interactions in which the charge–charge type or/and van der Waals interactions are involved in the ligand–protein complex formation ($|\Delta H_{(\text{ITC})}| > |T\Delta S_{(\text{ITC})}|$), the binding of $[\text{B(Ph)}_4]^-$ ions in the first site of the albumin is both enthalpy- and entropy-driven. However, the hydrophobic interactions seem to prevail ($|\Delta H_{(\text{ITC})}| < |T\Delta S_{(\text{ITC})}|$) [23].

A different scenario is observed for binding $[\text{B(Ph)}_4]^-$ ions in the second site. The positive value of $\Delta H_{(\text{ITC})2}$ shows that the thermodynamic stability of the resulting complexes strongly depends on the entropy change, while the charge–charge type or/and van der Waals interactions are involved in the SDS binding to the second site (Table 1). Therefore, it is supposed that $[\text{B(Ph)}_4]^-$ -BSA (2nd site) binding occurs mainly via hydrophobic interactions which contribute to an increase in entropy on account of the dehydration of the reactants or as a result of ions exchange occurring in the $[\text{B(Ph)}_4]^-$ binding process. Subsequently, this supposition has been verified by MD simulations.

The obtained findings were employed for studying the potential $[\text{B(Ph)}_4]^-$ binding sites on the surface of the protein. To do this, the first and then the second binding site were saturated with $[\text{B(Ph)}_4]^-$ ions, and checked for whether the albumin is still able to bind SDS monomers (Figure 4). It turned out that BSA with the first site saturated by $[\text{B(Ph)}_4]^-$ ions is still able to bind SDS ($\log K_{(\text{ITC})} = 4.92 \pm 0.04$; $\Delta G_{(\text{ITC})} = -6.73 \pm 0.05 \text{ kcal mol}^{-1}$; $\Delta H_{(\text{ITC})} = -12.85 \pm 0.35 \text{ kcal mol}^{-1}$; $T\Delta S_{(\text{ITC})} = -6.12 \text{ kcal mol}^{-1}$), but in such a system only around four to five molecules of SDS are bound to one site, as opposed to around seven ($N_1 \sim 2$, $N_2 \sim 5$), as was observed in the absence of $[\text{B(Ph)}_4]^-$ ions in the solution (Table 1).

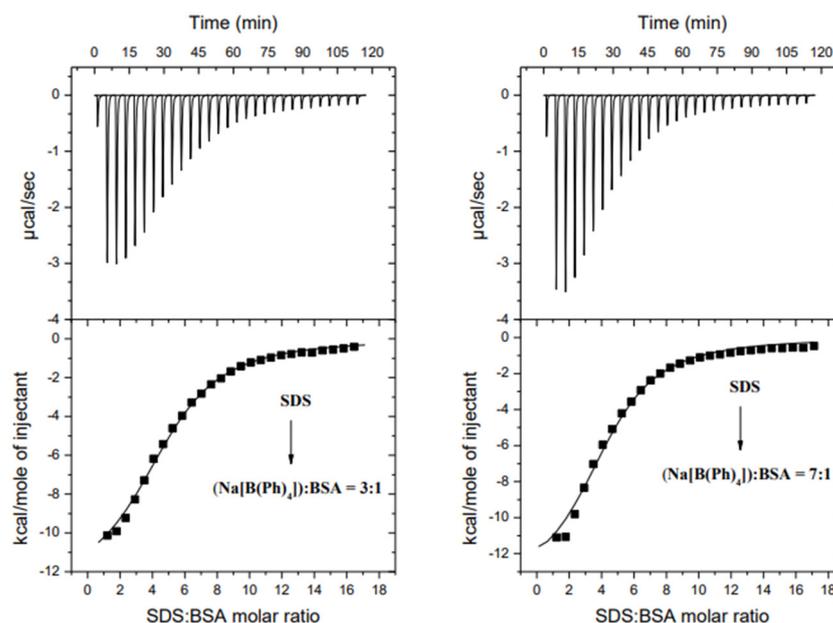


Figure 4. Calorimetric titration isotherms of the binding interactions between SDS and BSA in the presence of $\text{Na[B(Ph)}_4]$ in the $\text{Na[B(Ph)}_4]$:BSA molar ratio 3:1 (left) and 7:1 (right) in the 10 mM Caco buffer of pH 7, at 298.15 K.

Based on these observations, it can be assumed that both ligands ($[\text{B}(\text{Ph})_4]^-$ and SDS) reveal an affinity to the same “first” binding site in BSA. After a saturation of the first site by $[\text{B}(\text{Ph})_4]^-$ there is not enough space for binding of SDS monomers to the same surface of BSA. For this reason, negatively charged SDS anions are forced to bind to the second site of the protein. The above assumption confirms the following observations:

- There are no significant differences in the thermodynamic stability of the ligand–albumin complexes formed in the first binding site. Thus, SDS is not a strong enough ligand to replace $[\text{B}(\text{Ph})_4]^-$ in the first binding site.
- Binding constants of the SDS–BSA complexes formed in the absence and in the presence of $[\text{B}(\text{Ph})_4]^-$ have similar values:

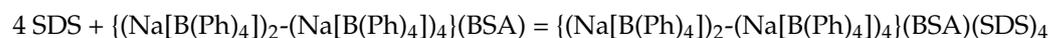
$$5 \text{ SDS} + (\text{SDS})_2\text{-BSA} = (\text{SDS})_2\text{-BSA}-(\text{SDS})_5 \quad \log K_{(\text{ITC})2} = 5.29 (\pm 0.06)$$

$$5 \text{ SDS} + (\text{Na}[\text{B}(\text{Ph})_4])_2\text{-BSA} = (\text{Na}[\text{B}(\text{Ph})_4])_2\text{-BSA}-(\text{SDS})_5 \quad \log K_{(\text{ITC})} = 4.92 (\pm 0.04)$$

In the absence of $[\text{B}(\text{Ph})_4]^-$ in the system, the binding constant of SDS to the first binding site of BSA is expected to be $\log K_{(\text{ITC})}$ approximately 7 (Table 1). The ITC data showed that in cases where the $[\text{B}(\text{Ph})_4]^-$ ions occupy the first site of the protein, the calculated binding constant ($\log K_{(\text{ITC})} \sim 5$) corresponds to the binding of SDS to the second binding site.

- The repulsion interactions between the negatively charged ligands ($[\text{B}(\text{Ph})_4]^-$ and SDS) disfavor occupation by them the same binding site in BSA.

A similar phenomenon has been observed after the saturation of the albumin by $[\text{B}(\text{Ph})_4]^-$ in both binding sites (Figure 4). Under such experimental conditions, the albumin still possesses one binding site capable of binding four moles of SDS monomers ($\log K_{(\text{ITC})} = 4.97 \pm 0.07$; $\Delta G_{(\text{ITC})} = -6.78 \pm 0.09 \text{ kcal mol}^{-1}$; $\Delta H_{(\text{ITC})} = -13.95 \pm 0.71 \text{ kcal mol}^{-1}$; $T\Delta S_{(\text{ITC})} = -7.17 \text{ kcal mol}^{-1}$), according to the following general equation:



Moreover, it is worth noticing that the parameters of the interactions between SDS and BSA within the first as well as both sites being saturated by $[\text{B}(\text{Ph})_4]^-$ are comparable (within the range of the experimental error). This confirms the fact that the investigated ligands show affinity only for the first site in BSA. Upon saturation of the same “first” site in BSA, $[\text{B}(\text{Ph})_4]^-$ and SDS are bound at different sites.

2.2. Steady-State Fluorescence Spectroscopy

Figure 5 presents the fluorescence emission spectra of free BSA and the protein mixed with $\text{Na}[\text{B}(\text{Ph})_4]$ in molar ratios equal to 1:3, 1:7, and 1:15 under the action of increasing concentrations of SDS. First of all, an increase in the amount of $[\text{B}(\text{Ph})_4]^-$ in the system has no impact on the position of an initial strong emission band of tryptophan residues in BSA at approximately 348 nm. Secondly, it can be observed that the intrinsic fluorescence intensity of the albumin (free and saturated with $[\text{B}(\text{Ph})_4]^-$) decreases regularly with the addition of SDS, accompanied by a significant blue shift from 348 nm to 332 nm, which may be a result of conformational changes in the BSA structure under the action of the surfactant (exposure of tryptophan residues to a more hydrophobic environment) [24]. Furthermore, the higher the amount of $[\text{B}(\text{Ph})_4]^-$ ions in the mixture, the lower the changes in the fluorescence intensity of BSA under the action of SDS. These findings are in line with the ITC results, and confirm the fact that upon saturation of BSA with $[\text{B}(\text{Ph})_4]^-$ the changes in the Trp fluorescence intensity are not as pronounced as for the $[\text{B}(\text{Ph})_4]^-$ -free solution. This is related to the fact that under such experimental conditions, the total number of SDS monomers bound to the protein is lower, and the SDS monomers are bound to their second binding site, which is far from sensitive to SDS binding of the Trp-134 and Trp-213 residues [20].

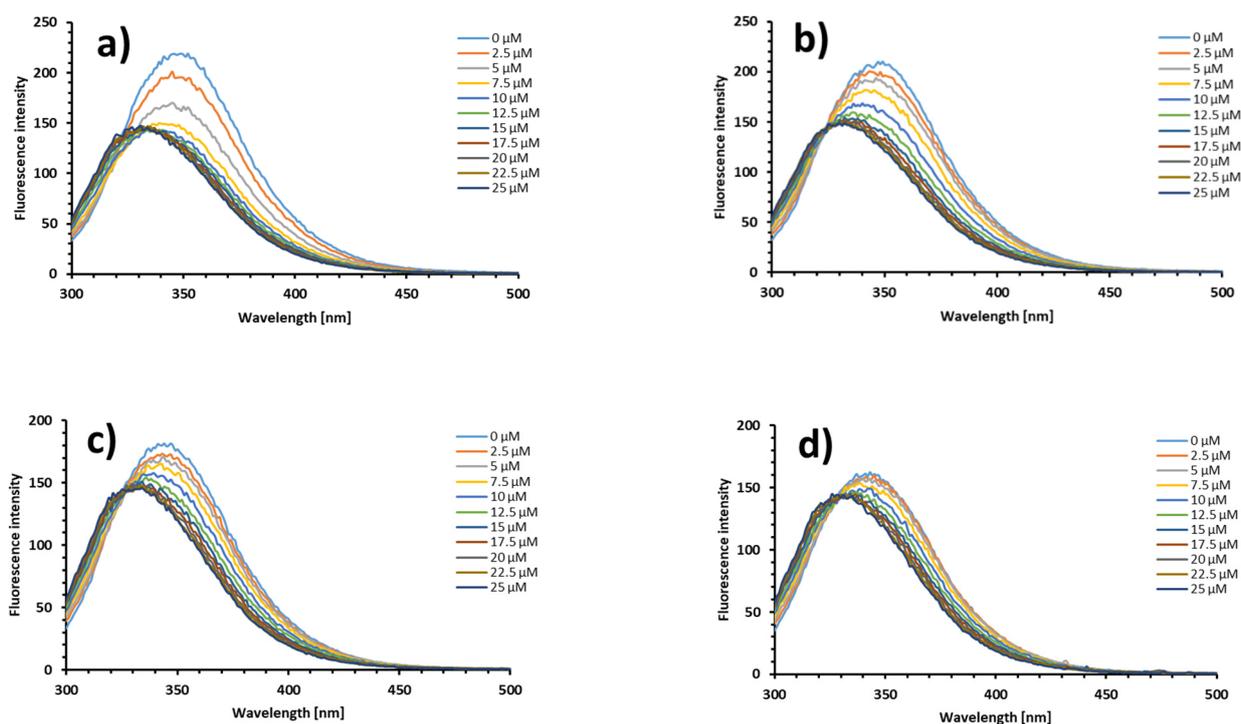


Figure 5. The fluorescence emission spectra of free BSA (a) and the solutions of BSA with Na[B(Ph)₄][−] mixed in molar ratios 1:3 (b); 1:7 (c); and 1:15 (d) in the presence of increasing concentrations of SDS (0–25 μM) in the 10 mM Caco buffer of pH 7.0 at 298.15 K.

To obtain a better insight into the mechanisms of the interactions between BSA (and its mixtures with [B(Ph)₄][−]) and SDS, the obtained results were analyzed according to a well-known Stern–Volmer equation: $\frac{F_0}{F} = 1 + K_{SV}[Q] = 1 + k_q\tau_0[Q]$, where F_0 and F denote the fluorescence intensities in the absence and presence of a quencher (SDS), $[Q]$ is the quencher concentration, K_{SV} is the Stern–Volmer quenching constant, k_q is the bimolecular quenching rate constant, while τ_0 is the lifetime of the fluorophore (BSA) in the absence of quencher [25,26]. The graphs of $\frac{F_0}{F}$ versus $[Q]$ plotted according to the Stern–Volmer equation are shown in Figure 6. Consequently, Table 2 presents the newly determined values of Stern–Volmer quenching constants along with linear correlation coefficients (R^2) and bimolecular quenching rate constants (k_q) recovered for the studied systems. The latter ones were calculated based on the value of the average fluorescence lifetime τ_0 of free BSA (in Tris-HCl, pH 7.0), which equals approximately 6.3 ns [27]. It can be observed that the quenching (binding) constants decrease with the increase in the [B(Ph)₄][−] concentration of the system, which can be assigned to the competition of SDS with [B(Ph)₄][−] in BSA [28,29]. Furthermore, the estimated values of k_q for all systems are of the order $10^{12} \text{ M}^{-1}\cdot\text{s}^{-1}$, which is approximately a few hundred times higher than the maximum value possible for the diffusion-controlled quenching-rate constant ($2.0 \times 10^{10} \text{ M}^{-1}\cdot\text{s}^{-1}$) [30]. Since the values of bimolecular quenching rate constants are considered to be definitive in differentiating between dynamic and static quenching mechanisms, the predominant role of ground-state complexation (static quenching) in the investigated systems was confirmed.

2.3. Differential Scanning Calorimetry (DSC)

The influence of [B(Ph)₄][−] ions on a thermal unfolding transition of BSA was assessed based on DSC experiments. The representative, original DSC curves are shown in Figure 7. Two distinct peaks are observed in the heat-capacity curve for the sample of free BSA. The transition midpoints $T_{m1} = 328.9 \text{ K}$ and $T_{m2} = 346.7 \text{ K}$ were estimated by fitting the experimental C_p vs. T curves to a non-two-state model using Marquardt

nonlinear least-squares method (T_m denotes the temperature value at which a maximum endothermic effect appears). Previously, it has been reported that the appearance of these peaks corresponds to the low-temperature (T_{m1}) and the high-temperature (T_{m2}) transition resulting from the melting of structurally independent parts of the protein [31–34]. The less thermally stable domain comprises the subdomains IIB, IIIA, and IIIB, whereas the subdomains IA, IB, and IIA represent the region (the energetic domain) more resistant to thermal unfolding [35].

The formation of fairly stable Na[B(Ph)₄]-BSA complexes makes the BSA structure more compact and thermostable. This is reflected in the shift of the heat-induced transitions towards a higher range of temperatures with the increase in the Na[B(Ph)₄]:BSA molar ratios in the mixture, namely $T_{m1} = 341.8$ K and $T_{m2} = 351.9$ K for the saturation of the first binding sites (Na[B(Ph)₄]:BSA = 2.5:1) and $T_{m1} = 349.2$ K and $T_{m2} = 356.2$ K for BSA with both binding sites saturated (Na[B(Ph)₄]:BSA = 7:1).

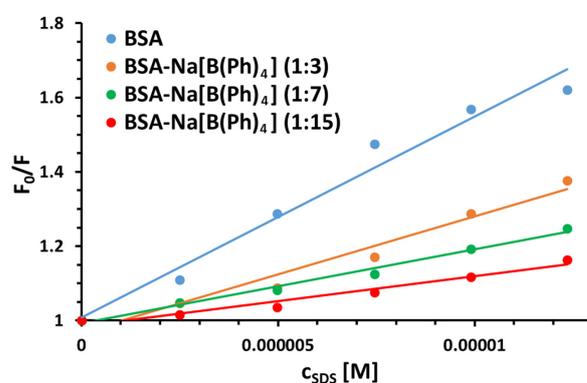


Figure 6. Stern–Volmer plots for the steady-state fluorescence quenching of BSA and its mixtures with Na[B(Ph)₄] (1:3; 1:7; and 1:15) by SDS in the 10 mM Caco buffer of pH 7.0 at 298 K.

Table 2. Stern–Volmer quenching constants, linear correlation coefficients (R^2) and bimolecular quenching rate constants (k_q) recovered for the steady-state fluorescence quenching of pure BSA and its mixtures with Na[B(Ph)₄] (1:3; 1:7; and 1:15) by SDS in the 10 mM Caco buffer of pH 7.0 at 298 K.

System	K_{SV} [10^4 M ⁻¹]	R_2	k_q [M ⁻¹ ·s ⁻¹]
BSA	5.41	0.972	8.58×10^{12}
BSA-Na[B(Ph) ₄] (1:3)	3.12	0.965	4.94×10^{12}
BSA-Na[B(Ph) ₄] (1:7)	1.98	0.987	3.15×10^{12}
BSA-Na[B(Ph) ₄] (1:15)	1.35	0.962	2.14×10^{12}

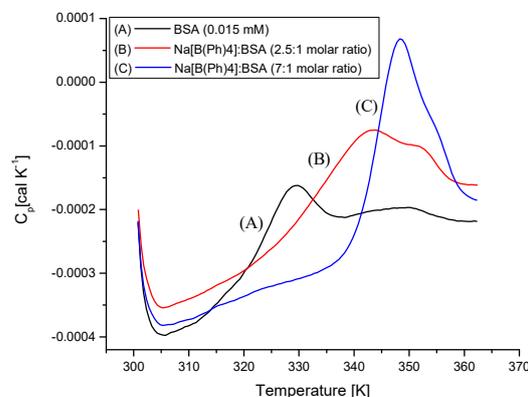


Figure 7. The raw heat capacity data for: (A) free BSA (0.015 mM) and the mixtures (B) with the Na[B(Ph)₄]:BSA molar ratios 2.5:1 (0.0375 mM Na[B(Ph)₄], 0.015 mM BSA) and (C) 7:1 (0.105 mM Na[B(Ph)₄], 0.015 mM BSA) in the 10 mM Caco buffer of pH 7.

2.4. Circular Dichroism (CD) Spectra Analysis

The effect of $\text{Na}[\text{B}(\text{Ph})_4]$ ions' binding and temperature on the secondary structure of BSA was assessed by temperature-dependent CD measurements in the range of temperatures from 298.15 K to 368.15 K (Figures 8–10). The performed experiments showed that the saturation of BSA with $[\text{B}(\text{Ph})_4]^-$ ions at 298.15 K does not contribute to noticeable changes in the secondary structure of the protein. On the other hand, the thermal stability of the α -helix structure increases on account of the formation of the $\text{Na}[\text{B}(\text{Ph})_4]$ –BSA complexes (Table 3). These findings are in agreement with the DSC results.

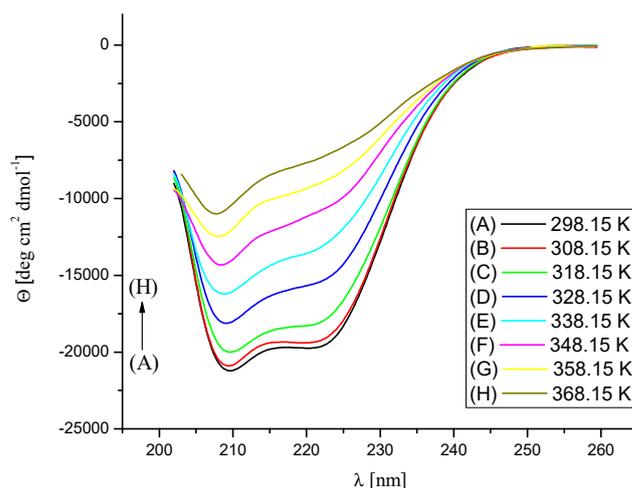


Figure 8. CD spectra of BSA in 10 mM Caco buffer at pH 7, in the temperature range 298.15–368.15 K. The concentration of BSA was maintained at 0.0015 mM.

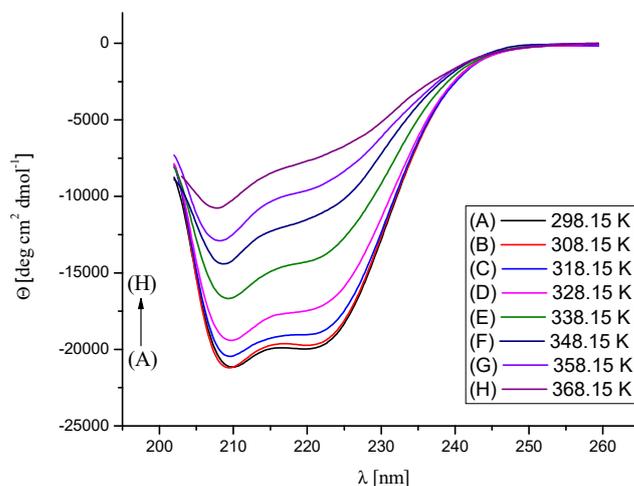


Figure 9. CD spectra of the $\text{Na}[\text{B}(\text{Ph})_4]$ /BSA mixture ($\text{Na}[\text{B}(\text{Ph})_4]$:BSA = 2.5:1 molar ratio) in 10 mM Caco buffer at pH 7, in the temperature range 298.15–368.15 K. The concentration of BSA was maintained at 0.0015 mM.

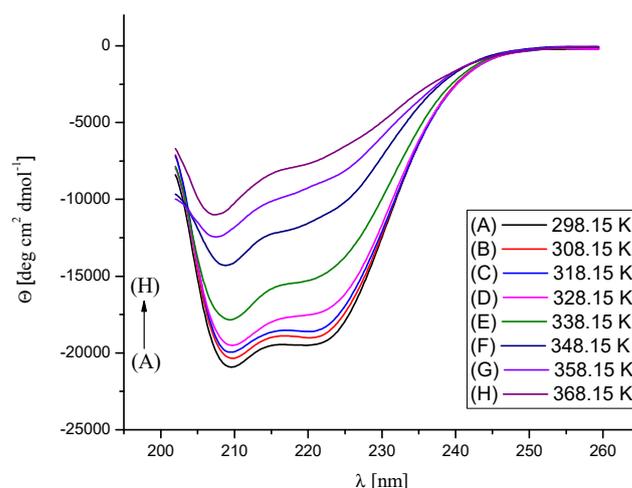


Figure 10. CD spectra of the Na[B(Ph)₄]/BSA mixture (Na[B(Ph)₄]:BSA = 7:1 molar ratio) in 10 mM Caco buffer at pH 7, in the temperature range 298.15–368.15 K. The concentration of BSA was maintained at 0.0015 mM.

Table 3. The secondary structure content (%) of BSA at different Na[B(Ph)₄]:BSA molar ratios in 10 mM Caco buffer at pH 7 in the temperature range 298.15–368.15 K, revealed from CD measurements.

T [K]	BSA	Na[B(Ph) ₄]:BSA	
		2.5:1 (Molar Ratio)	7:1 (Molar Ratio)
The Percentage Content of α-Helix (H); β-Structure (S); β-Turn (Trn); Random Coil (Unrd) [%]			
298.15	59.8 (H); 6.7 (S) 11.1 (Trn); 22.3 (Unrd)	62.5 (H); 5.2 (S) 10.5 (Trn); 21.1 (Unrd)	59.2 (H); 6.1 (S) 11.5 (Trn); 23.1 (Unrd)
308.15	60.5 (H); 8.0 (S) 11.7 (Trn); 20.3 (Unrd)	63.9 (H); 4.5 (S) 10.3 (Trn); 21.3 (Unrd)	59.5 (H); 6.8 (S) 11.5 (Trn); 22.6 (Unrd)
318.15	56.0 (H); 7.3 (S) 13.4 (Trn); 23.6 (Unrd)	61.3 (H); 5.8 (S) 10.9 (Trn); 22.6 (Unrd)	58.2 (H); 7.8 (S) 12.1 (Trn); 22.4 (Unrd)
328.15	50.8 (H); 10.5 (S) 14.1 (Trn); 24.9 (Unrd)	58.0 (H); 7.0 (S) 12.5 (Trn); 22.6 (Unrd)	56.0 (H); 8.4 (S) 12.1 (Trn); 23.8 (Unrd)
338.15	44.4 (H); 11.5 (S) 16.7 (Trn); 27.5 (Unrd)	47.7 (H); 11.3 (S) 15.6 (Trn); 25.9 (Unrd)	50.2 (H); 8.4 (S) 15.8 (Trn); 25.9 (Unrd)
348.15	36.4 (H); 15.0 (S) 19.8 (Trn); 29.4 (Unrd)	39.1 (H); 14.4 (S) 18.6 (Trn); 28.3 (Unrd)	50.5 (H); 9.1 (S) 15.7 (Trn); 25.5 (Unrd)
358.15	25.4 (H); 21.2 (S) 24.1 (Trn); 29.8 (Unrd)	35.4 (H); 15.6 (S) 20.7 (Trn); 27.7 (Unrd)	23.7 (H); 22.4 (S) 22.9 (Trn); 31.0 (Unrd)
368.15	25.3 (H); 22.2 (S) 22.6 (Trn); 30.4 (Unrd)	22.0 (H); 24.9 (S) 22.3 (Trn); 31.0 (Unrd)	22.6 (H); 22.5 (S) 22.5 (Trn); 32.1 (Unrd)

2.5. Molecular Modelling

To gain a deeper understanding of the experimental results at the atomistic level, MD simulation was conducted for the [B(Ph)₄][−]–BSA system. In particular, the MD approach was applied to predict potential binding sites of [B(Ph)₄][−] on the BSA surface and to further characterize and compare them in terms of the binding affinities for these ions. The protein was surrounded by 15 randomly distributed [B(Ph)₄][−] ions and the simulation was performed for 100 ns at 300 K and a pH of 7. After the careful examination of the frames for each [B(Ph)₄][−] ligand, the last 40% of the simulation was used for the further LIE analysis (Table 4). Based on our results, two bindings sites were localized for three

molecules (Figure 11). Interestingly, one binding site is in the proximity to Trp-134, as was observed in the case of our previous SDS–BSA study [20]. Two ions bound to site I and one ion to site II after 40 and 60 ns, respectively, and remained there until the end of the simulation. Their RMSD as well as the RMSD of the protein are shown in Supplementary Figure S1 to demonstrate the convergence. LIE energy values for the ions bound in sites I and II are shown in Supplementary Figure S2, supporting the convergence achieved for the binding in these sites. In order to strictly define the binding event in sites I and II, a 5 Å cutoff criteria was applied for the distance between the residues of the sites and $[B(Ph)_4]^-$ ion atoms. According to LIE calculations, site I has slightly higher affinity towards $[B(Ph)_4]^-$ than site II due to the more favorable van der Waals component of the free binding energy (the electrostatic and van der Waals components calculated for each $[B(Ph)_4]^-$ ion are provided in Table 4). Site I comprises Lys116, Pro119, Glu140, Ile141, Arg144, His145, Leu178, Glu182, Arg185, and Val188, while site II is formed by Leu386, Asn390, Phe402, Leu406, Arg409, Thr410, Lys413, Thr448, Ser479, Arg484, and Phe487. Clearly, site I has a more charged nature and lower propensity for establishing pi–pi and cation–pi interactions than site II. Additionally, when looking at the right panel of Figure 11, it seems that theoretical calculations confirm that hydrophobic rings of phenyl groups contribute to the stabilization of interactions between BSA and $[B(Ph)_4]^-$.

Table 4. LIE free-energy decomposition values for $[B(Ph)_4]^-$ ligands in the $[B(Ph)_4]^-$ –BSA complex. Three $[B(Ph)_4]^-$ ions (no 2, 6 and 11) did not form stable complexes with BSA throughout the simulation. ΔG_{ele} , ΔG_{vdW} and ΔG_{tot} correspond to the electrostatic component, van der Waals component and total energy, respectively.

No of $[B(Ph)_4]^-$	ΔG_{ele} [kcal mol ⁻¹]	ΔG_{vdW} [kcal mol ⁻¹]	ΔG_{tot} [kcal mol ⁻¹]
1	−0.3 ± 0.1	−21.0 ± 2.5	−21.3 ± 2.6
2	N/A	N/A	N/A
3	−0.2 ± 0.1	−23.0 ± 3.2	−23.3 ± 3.2
4	−0.6 ± 0.2	−22.5 ± 2.9	−23.1 ± 2.9
5	−0.2 ± 0.1	−23.4 ± 3.4	−23.6 ± 3.4
6	N/A	N/A	N/A
7	−0.4 ± 0.2	−18.6 ± 5.7	−18.7 ± 5.9
8	−0.6 ± 0.1	−29.9 ± 2.5	−30.5 ± 2.5
9	−0.3 ± 0.1	−15.8 ± 2.5	−16.1 ± 2.5
10	0.7 ± 0.1	−32.2 ± 2.6	−32.8 ± 2.6
11	N/A	N/A	N/A
12	−0.3 ± 0.2	−21.2 ± 2.9	−21.5 ± 2.9
13	−0.2 ± 0.1	−15.2 ± 1.9	−15.4 ± 1.9
14	−0.6 ± 0.2	−28.5 ± 3.2	−29.0 ± 3.3
15	−0.3 ± 0.2	−20.8 ± 3.8	−21.1 ± 3.9

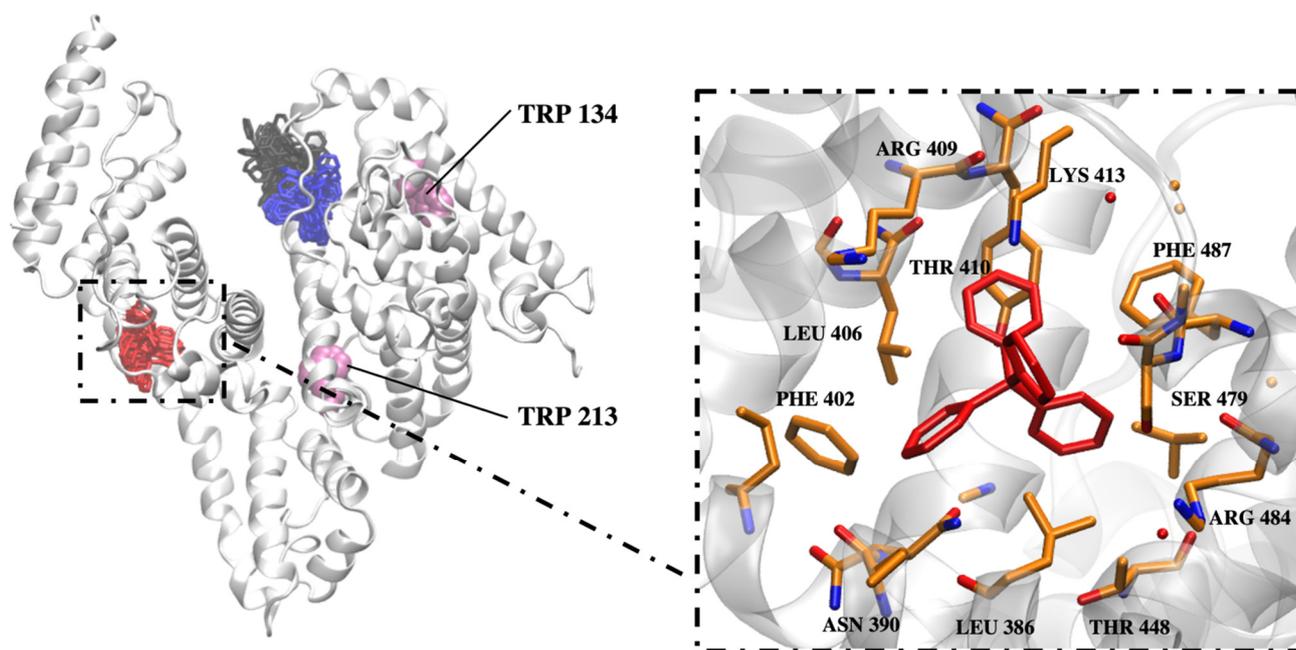


Figure 11. On the left, BSA protein (grey cartoon), Trp residues (magenta VDW representation) and three clusters (licorice) representing $[B(Ph)_4]^-$ ions that bound to the protein with the highest affinity. Colors of the clusters correspond to the color coding used in Table 4. On the right panel, a representative frame for one $[B(Ph)_4]^-$ ion and its protein surrounding.

3. Materials and Methods

3.1. Reagents

Bovine serum albumin (BSA, lyophilized powder, $\geq 96\%$), sodium dodecyl sulfate (SDS, for molecular biology, $\geq 99\%$), sodium tetrphenylborate ($Na[B(Ph)_4]$, $\geq 99.5\%$), sodium dodecyl sulfate (SDS, for molecular biology, $\geq 99\%$) and sodium cacodylate trihydrate (Caco, $\geq 98\%$) were obtained from Merck (Poland) and employed as received without further purification. Double-distilled water with conductivity not exceeding $0.18 \mu S cm^{-1}$ was used for preparations of buffer solutions.

3.2. Isothermal Titration Calorimetry (ITC)

All ITC experiments were performed at 298.15 K using an AutoITC isothermal titration calorimeter (MicroCal Inc. GE Healthcare, Northampton, MA, USA). The details of the measuring device and experimental setup have been described previously [20]. The reagents were dissolved directly in 10 mM Caco buffer of pH 7. The experiments consisted of injecting $10.02 \mu L$ (29 injections, $2 \mu L$ for the first injection only) of (a) 1 mM buffered solution of $Na[B(Ph)_4]$ into the reaction cell which initially contained the 0.02 mM buffered solution of BSA, and (b) 1 mM buffered solution of SDS into the reaction cell which initially contained the $Na[B(Ph)_4]$:BSA mixture of the 3:1 (0.0375 mM $Na[B(Ph)_4]$, 0.0125 mM BSA) or 7:1 (0.0875 mM $Na[B(Ph)_4]$, 0.0125 mM BSA) molar ratio. A background titration, consisting of an identical titrant solution but with the buffer solution in the reaction cell only, was subtracted from each experimental titration on account of the heat of dilution. The titrant was injected at 5 min intervals. Each injection lasted 20 s.

3.3. Steady-State Fluorescence Spectroscopy and UV Spectrophotometry

The stock solutions of BSA, SDS and $Na[B(Ph)_4]$ were prepared in 10 mM Caco buffer of pH 7 (all subsequent dilutions were made with this buffer). The concentration of bovine serum albumin was measured using an extinction coefficient equal to $\epsilon_{280}^{BSA} = 41,180 M^{-1} cm^{-1}$, calculated based on the content of tryptophan ($\epsilon_{280}^W = 5690 M^{-1} cm^{-1}$), tyrosine ($\epsilon_{280}^Y = 1280 M^{-1} cm^{-1}$), and cysteine ($\epsilon_{280}^C = 120 M^{-1} cm^{-1}$) [36]. A maximum ab-

sorbance value of approximately 0.08 at 280 nm (corresponding to a protein concentration of 2 μM) was used to avoid the inner-filter effect.

Steady-state fluorescence experiments were carried out with a Cary Eclipse Varian (Agilent, Santa Clara, CA, USA) spectrofluorometer, equipped with a temperature controller and a 1.0 cm multicell holder. The absorption spectra were recorded on Perkin Elmer Lambda 650 (Waltham, MA, USA) UV/Vis spectrophotometer. In the performed fluorescence titration experiments: (i) 2 mL of pure BSA (2 μM BSA); (ii) 2 mL of $\text{Na}[\text{B}(\text{Ph})_4]$ and BSA mixed in ratio 3:1 (6 μM $\text{Na}[\text{B}(\text{Ph})_4]$, 2 μM BSA); (iii) 2 mL of BSA and $\text{Na}[\text{B}(\text{Ph})_4]$ mixed in ratio 7:1 (14 μM $\text{Na}[\text{B}(\text{Ph})_4]$, 2 μM BSA); and (iv) 2 mL of BSA and $\text{Na}[\text{B}(\text{Ph})_4]$ mixed in ratio 15:1 (30 μM $\text{Na}[\text{B}(\text{Ph})_4]$, 2 μM BSA) were simultaneously titrated with ten 5 μL aliquots of sodium dodecyl sulphate (1 mM). The fluorescence intensity of the band at 348 nm—corresponding to the initial maximum emission of BSA—was used to calculate the quenching constants and then other parameters. Excitation wavelength was always set at 280 nm. All experiments were performed at 298.15 K.

3.4. Differential Scanning Calorimetry (DSC)

Calorimetric (DSC) measurements were made with a VP-DSC microcalorimeter (MicroCal Inc. GE Healthcare, Northampton, USA) at a scanning rate 90 K h^{-1} . All scans were run in 10 mM Caco buffer of pH 7 in the temperature range of 298.15–363.15 K at a scan rate of 1.5 K min^{-1} . The scans were obtained for pure BSA (0.015 mM) and for the solution containing $\text{Na}[\text{B}(\text{Ph})_4]$ and BSA mixed in ratio 2.5:1 (0.0375 mM $\text{Na}[\text{B}(\text{Ph})_4]$, 0.015 mM BSA) and 7:1 (0.105 mM $\text{Na}[\text{B}(\text{Ph})_4]$, 0.015 mM BSA). All the samples were degassed before measurements. These measurements were recorded two times. The details of the measuring device, experimental setup and data processing have been described previously [20].

3.5. Circular Dichroism Spectroscopy (CD)

Circular dichroism (CD) spectra were recorded in water on a Jasco-715 automatic recording spectropolarimeter (Jasco Inc., Easton, MD, USA) for the following systems: pure BSA (0.0015 mM) and two $\text{Na}[\text{B}(\text{Ph})_4]$ /BSA mixtures in stoichiometric molar ratios 2.5:1 (0.00375 mM $\text{Na}[\text{B}(\text{Ph})_4]$, 0.0015 mM BSA) and 7:1 (0.0105 mM $\text{Na}[\text{B}(\text{Ph})_4]$, 0.0015 mM BSA) in the 10 mM Caco buffer of pH 7, in a temperature range from 298.15 K to 368.15 K, every 10 degrees. The details of the measuring device, experimental setup and data processing have been described previously [20]. The secondary structure of BSA under experimental conditions in the absence and presence of $\text{Na}[\text{B}(\text{Ph})_4]$ was determined using the CONTIN/LL program, a variant of the CONTIN method developed by Provencher and Glockner [37] provided within the CDPro software package.

3.6. Theoretical Studies

3.6.1. Structures

The structure of Bovine Serum Albumin (BSA) was obtained from the Protein Data Bank (PDB ID: 4F5S, 2.47 Å). The structure of $[\text{B}(\text{Ph})_4]^-$ (referred to as BPH for simplicity) was parameterized as by Kurt and Temel [38,39] and later built in xleap of AMBER16 package [40].

3.6.2. Molecular Dynamics Simulations

The MD approach was applied to localize potential binding sites and to characterize them in terms of the binding affinities. There are several reasons why such an approach is superior to the classical molecular docking approach, and therefore was used instead of molecular docking in our study: (1) Docking small ligands as SDS in our previous work [20] and $[\text{B}(\text{Ph})_4]^-$ to BSA in this work did not allow for finding more than a single binding site. (2) MD allows for the complete flexibility of the protein residues, which could be crucial for the binding of a small ligand, while rigid receptor molecular docking does not allow for it. Taking into account a structural ensemble of receptors in molecular docking is a possible

alternative which is computationally more expensive than the MD approach. (3) In contrast to molecular docking, solvent is treated explicitly in the MD simulations. (4) Molecular docking scores are far less reliable than MD-derived binding affinities due to the fact that molecular docking scores single, static poses, while multiple frames corresponding to the statistical ensemble of the structures are considered in the MD-based calculations. (5) In most of the state-of-the-art modelling studies, the poses obtained by a molecular docking approach are further analyzed with the MD approach. Therefore, the MD approach, when applied for predicting binding sites, is already self-consistent in comparison to the two-step procedure that includes molecular docking followed by MD. Hence, the MD approach is preferable. (6) Furthermore, we previously demonstrated both convergence and high predictive power of the MD approach when applied to the protein–ion systems [41]. All all-atom molecular dynamics (MD) simulations were performed with the use of AMBER16 software package [40]. The initial structures of the complexes were created by random placement of 15 $[\text{B}(\text{Ph})_4]^-$ ions around the BSA protein. A truncated, octahedron, TIP3P periodic box of an 8 Å water layer from the box's border to the solute was used to solvate complexes. The ff14SBonlysc force field for the protein [42] and gaff force field [43] with RESP charges [44] obtained in the antechamber module of AMBER16 [40] for $[\text{B}(\text{Ph})_4]^-$ were used. The net negative charge of the system was neutralized with Na^+ counterions. Energy minimization was carried out in two steps: beginning with 500 steepest descent cycles and 10^3 conjugate gradient cycles with 100 kcal/mol/Å² harmonic force restraints on solute atoms, continued with 3×10^3 steepest descent cycles and 3×10^3 conjugate gradient cycles without any restraints. Following minimization steps, the system was heated up from 0 to 300 K for 10 ps with harmonic force restraints of 100 kcal/mol/Å² on solute atoms. Then, the system was equilibrated at 300 K and 10^5 Pa in isothermal isobaric ensemble for 100 ps. Afterwards, an independent prediction MD run was carried out in the same isothermal isobaric ensemble for 100 ns. The particle mesh Ewald method for treating electrostatics and the SHAKE algorithm for all the covalent bonds containing hydrogen atoms were implemented in the MD simulations. Such parameters within the MD simulation were shown to be sufficient for a system of ions with a protein, which was substantially bigger than BSA, allowing for proper prediction of the ion binding sites [41].

3.6.3. Binding Free-Energy Calculations

Energetic postprocessing of the trajectories and per-residue energy decomposition was performed for the system with the use of Linear Interaction Energy (LIE) with dielectric constant of 80. Frames from the MD simulation were carefully examined with VMD [45]. The last 40% of the simulation was chosen for the analysis, since in the corresponding frames no more events of ion dissociations or associations were observed. LIE takes into account only the electrostatic component (scaled by the dielectric constant) and the van der Waals component of the binding free energy. Although the binding free energies obtained by this method are not expected to represent absolute binding free energies, these calculations are useful to compare the binding strength in different binding sites.

4. Conclusions

A few complementary experimental methods supported by in silico analysis have successfully been applied to describe the physicochemical nature of tetraphenylborate ions' interactions with BSA. The obtained results were subsequently discussed in relation to a previously studied system, namely the interaction of sodium dodecyl sulphate (SDS) with BSA. The described approach enables us to find the BSA potential binding sites of $[\text{B}(\text{Ph})_4]^-$.

In conclusion, two main binding sites of $[\text{B}(\text{Ph})_4]^-$ were unveiled. They are located in the subdomain IA and IIIA, respectively, of the investigated protein. Both tetraphenylborate ions and SDS molecules reveal a similar affinity to the first specific binding site in BSA which is located close to the Trp-134 residue. In turn, the second binding site on the BSA surface

for investigated ligands are different. The binding of $[B(Ph)_4]^-$ ions in the subdomain IIIA is an entropy-driven process, whereas SDS in the subdomain IIB is enthalpy-driven.

It has been proven that $[B(Ph)_4]^-$ ions participate in the stabilization of the tertiary structure of BSA at higher temperatures on account of the formation of fairly stable $[B(Ph)_4]^-$ -BSA complexes of a different stoichiometry ($[B(Ph)_4]^-$:BSA = 2:1 and 5:1). This phenomenon underlying its biological activity could be responsible for the protection of the protein from thermal unfolding.

The results clearly show that the presence of competitive ligands may have some implications on the concentration of the compounds tested (as has been found for $[B(Ph)_4]^-$ ions or SDS) in the cell-culture media containing albumin. On the one hand, BSA, an important component of the cell culture media, binding to the compounds of a hydrophobic nature, as $[B(Ph)_4]^-$ or SDS, can lead to a decrease in the concentration of free, active species. In consequence, this can affect the cytotoxic action of the tested compound. On the other hand, the saturation of the potential binding sites of BSA by a nontoxic, low-molecular-weight ligand may prevent the interactions of albumin with other biologically active compounds, thus affecting their dose-dependent action.

The findings presented in this contribution are worth taking into consideration during the analysis of the cytotoxicity assays of the compounds being characterized.

Supplementary Materials: The following are available online. Figure S1: RMSD of protein (upper panel) and $[B(Ph)_4]^-$ ions in site I (middle panels) and in site II (bottom panel) in the course of the MD simulation, Figure S2: LIE free energy of binding for $[B(Ph)_4]^-$ ions in site I (upper and middle panels) and in site II (bottom panel) in the course of the MD simulation.

Author Contributions: Conceptualization, O.G., M.M.K., S.A.S. and D.W.; methodology, S.A.S., D.W.; formal analysis, M.M.K., K.Ž. and D.W.; investigation, O.G., M.M.K., K.Ž., J.M., A.T., K.C. and D.W.; data curation, O.G., M.M.K., A.T., K.C.; writing—original draft preparation, O.G., M.M.K., K.Ž., S.A.S., D.W. and L.C.; writing—review and editing, M.M.K., K.Ž., S.A.S., D.W. and L.C.; visualization, O.G., M.M.K., K.Ž. and D.W.; supervision, S.A.S., D.W. and L.C.; project administration, D.W.; funding acquisition, K.Ž. and S.A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Polish National Science Centre under Grants No. UMO-2016/23/D/ST4/01576 (K.Ž.), UMO-2018/30/E/ST4/00037 (S.A.S.).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

Sample Availability: All compounds used in the study are commercially available (for provider, see Section 3).

References

1. Pessoa, J.C.; Correia, I. Misinterpretations in Evaluating Interactions of Vanadium Complexes with Proteins and Other Biological Targets. *Inorganics* **2021**, *9*, 17. [[CrossRef](#)]
2. Chi, E.Y.; Krishnan, S.; Randolph, T.W.; Carpenter, J.F. Physical Stability of Proteins in Aqueous Solution: Mechanism and Driving Forces in Nonnative Protein Aggregation. *Pharm. Res.* **2003**, *20*, 1325–1336. [[CrossRef](#)] [[PubMed](#)]
3. March, D.; Bianco, V.; Franzese, G. Protein Unfolding and Aggregation near a Hydrophobic Interface. *Polymers* **2021**, *13*, 156. [[CrossRef](#)] [[PubMed](#)]
4. Quevedo, M.; Karbstein, H.P.; Emin, M.A. Influence of thermomechanical treatment and pH on the denaturation kinetics of highly concentrated whey protein isolate. *J. Food Eng.* **2021**, *292*, 110294. [[CrossRef](#)]
5. Rose, G.D.; Fleming, P.J.; Banavar, J.R.; Maritan, A. A backbone-based theory of protein folding. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 16623–16633. [[CrossRef](#)] [[PubMed](#)]
6. Williams, R.J.P.; Fraústo da Silva, J.J.R. *The Natural Selection of the Chemical Elements—The Environment and Life's Chemistry*; Oxford University Press: Oxford, UK, 1996.

7. Jahmide-Azizi, N.; Oliva, R.; Gault, S.; Cockell, C.; Winter, R. The Effects of Temperature and Pressure on Protein-Ligand Binding in the Presence of Mars-Relevant Salts. *Biology* **2021**, *10*, 687. [[CrossRef](#)]
8. Seedher, N.; Agarwal, P. Competitive binding of fluoroquinolone antibiotics and some other drugs to human serum albumin: A luminescence spectroscopic study. *Luminescence* **2013**, *28*, 562–568. [[CrossRef](#)] [[PubMed](#)]
9. Maciążek-Jurczyk, M.; Sułkowska, A.; Bojko, B.; Równicka, J.; Sułkowski, W. Fluorescence analysis of competition of phenylbutazone and methotrexate in binding to serum albumin in combination treatment in rheumatology. *J. Mol. Struct.* **2009**, *924–926*, 378–384. [[CrossRef](#)]
10. Zhang, G.; Wang, A.; Jiang, T.; Guo, J. Interaction of the irisfloreutin with bovine serum albumin: A fluorescence quenching study. *J. Mol. Struct.* **2008**, *891*, 93–97. [[CrossRef](#)]
11. Macii, F.; Biver, T. Spectrofluorimetric analysis of the binding of a target molecule to serum albumin: Tricky aspects and tips. *J. Inorg. Biochem.* **2021**, *216*, 111305. [[CrossRef](#)] [[PubMed](#)]
12. Ni, Y.; Su, S.; Kokot, S. Spectrofluorimetric studies on the binding of salicylic acid to bovine serum albumin using warfarin and ibuprofen as site markers with the aid of parallel factor analysis. *Anal. Chim. Acta* **2006**, *580*, 206–215. [[CrossRef](#)]
13. Sudlow, G.; Birkett, D.J.; Wade, D.N. The characterization of two specific drug binding sites on human serum albumin. *Mol. Pharmacol.* **1975**, *11*, 824–832.
14. Ding, F.; Li, N.; Han, B.; Liu, F.; Zhang, L.; Sun, Y. The binding of C.I. Acid Red 2 to human serum albumin: Determination of binding mechanism and binding site using fluorescence spectroscopy. *Dyes Pigments* **2009**, *83*, 249–257. [[CrossRef](#)]
15. Handing, K.B.; Shabalin, I.G.; Kassaar, O.; Khazaipoul, S.; Blindauer, C.A.; Stewart, A.J.; Chruszcz, M.; Minor, W. Circulatory zinc transport is controlled by distinct interdomain sites on mammalian albumins. *Chem. Sci.* **2016**, *7*, 6635–6648. [[CrossRef](#)] [[PubMed](#)]
16. Sciortino, G.; Sanna, D.; Lubinu, G.; Maréchal, J.; Garribba, E. Unveiling V^{IV}O²⁺ Binding Modes to Human Serum Albumins by an Integrated Spectroscopic–Computational Approach. *Chem. A Eur. J.* **2020**, *26*, 11316–11326. [[CrossRef](#)]
17. Sciortino, G.; Maréchal, J.D.; Garribba, E. Integrated approaches to characterize the systems formed by vanadium with proteins and enzymes. *Inorg. Chem. Front.* **2021**, *8*, 1951–1974. [[CrossRef](#)]
18. Sun, X.-Y.; Bi, S.-Y.; Wu, J.; Zhao, R. Study on the interaction of amprolium HCl and dinitolmide in animal-derived food products with BSA by multiple spectroscopies and molecular modeling techniques. *J. Biomol. Struct. Dyn.* **2019**, *37*, 4283–4291. [[CrossRef](#)] [[PubMed](#)]
19. Abedin, J.; Mahbub, S.; Rahman, M.M.; Hoque, A.; Kumar, D.; Khan, J.M.; El-Sherbeeney, A.M. Interaction of tetradecyltrimethylammonium bromide with bovine serum albumin in different compositions: Effect of temperatures and electrolytes/urea. *Chin. J. Chem. Eng.* **2021**, *29*, 279–287. [[CrossRef](#)]
20. Tesmar, A.; Kogut, M.M.; Żamojć, K.; Grabowska, O.; Chmur, K.; Samsonov, S.A.; Makowska, J.; Wyrzykowski, D.; Chmurzyński, L. Physicochemical nature of sodium dodecyl sulfate interactions with bovine serum albumin revealed by interdisciplinary approaches. *J. Mol. Liq.* **2021**, *340*, 117185. [[CrossRef](#)]
21. Pawar, S.K.; Jaldappagari, S. Interaction of repaglinide with bovine serum albumin: Spectroscopic and molecular docking approaches. *J. Pharm. Anal.* **2019**, *9*, 274–283. [[CrossRef](#)] [[PubMed](#)]
22. Grosseohme, N.E.; Spuches, A.M.; Wilcox, D.E. Application of isothermal titration calorimetry in bioinorganic chemistry. *J. Biol. Inorg. Chem.* **2010**, *15*, 1193–1194. [[CrossRef](#)]
23. Cramer, D.L.; Cheng, B.; Tian, J.; Clements, J.H.; Wypych, R.M.; Martin, S.F. Some thermodynamic effects of varying nonpolar surfaces in protein-ligand interactions. *Eur. J. Med. Chem.* **2020**, *208*, 112771. [[CrossRef](#)] [[PubMed](#)]
24. Sharma, V.; Yañez, O.; Alegria-Arcos, M.; Kumar, A.; Thakur, R.C.; Cantero-López, P. A physicochemical and conformational study of co-solvent effect on the molecular interactions between similarly charged protein surfactant (BSA-SDBS) system. *J. Chem. Thermodyn.* **2020**, *142*, 106022. [[CrossRef](#)]
25. Lakowicz, J.R. *Principles of Fluorescence Spectroscopy*, 3rd ed.; Springer: Singapore, 2006. [[CrossRef](#)]
26. Makowska, J.; Żamojć, K.; Wyrzykowski, D.; Wicz, W.; Chmurzyński, L. Copper(II) complexation by fragment of central part of FBP28 protein from *Mus musculus*. *Biophys. Chem.* **2018**, *241*, 55–60. [[CrossRef](#)]
27. Gelamo, E.; Tabak, M. Spectroscopic studies on the interaction of bovine (BSA) and human (HSA) serum albumins with ionic surfactants. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2000**, *56*, 2255–2271. [[CrossRef](#)]
28. Ghosh, S.; Dey, J. Interaction of bovine serum albumin with N-acyl amino acid based anionic surfactants: Effect of head-group hydrophobicity. *J. Colloid Interface Sci.* **2015**, *458*, 284–292. [[CrossRef](#)]
29. Ni, Y.; Zhang, X.; Kokot, S. Spectrometric and voltammetric studies of the interaction between quercetin and bovine serum albumin using warfarin as site marker with the aid of chemometrics. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2009**, *71*, 1865–1872. [[CrossRef](#)]
30. Makowska, J.; Żamojć, K.; Wyrzykowski, D.; Żmudzińska, W.; Uber, D.; Wierzbicka, M.; Wicz, W.; Chmurzyński, L. Probing the binding of Cu²⁺ ions to a fragment of the Aβ_(1–42) polypeptide using fluorescence spectroscopy, isothermal titration calorimetry and molecular dynamics simulations. *Biophys. Chem.* **2016**, *216*, 44–50. [[CrossRef](#)] [[PubMed](#)]
31. Michnik, A. Thermal stability of bovine serum albumin DSC study. *J. Therm. Anal. Calorim.* **2003**, *71*, 509–519. [[CrossRef](#)]
32. Precupas, A.; Sandu, R.; Leonties, A.R.; Anghel, D.-F.; Popa, V.T. Complex interaction of caffeic acid with bovine serum albumin: Calorimetric, spectroscopic and molecular docking evidence. *New J. Chem.* **2017**, *41*, 15003–15015. [[CrossRef](#)]
33. Yamasaki, M.; Yano, H.; Aoki, K. Differential scanning calorimetric studies on bovine serum albumin: I. Effects of pH and ionic strength. *Int. J. Biol. Macromol.* **1990**, *12*, 263–268. [[CrossRef](#)]

34. Murayama, K.; Tomida, M. Heat-Induced Secondary Structure and Conformation Change of Bovine Serum Albumin Investigated by Fourier Transform Infrared Spectroscopy. *Biochemistry* **2004**, *43*, 11526–11532. [[CrossRef](#)] [[PubMed](#)]
35. Bohlooli, M.; Moosavi-Movahedi, A.A.; Ghaffari-Moghaddam, M.; Saboury, A.A.; Khajeh, M.; Najafi, S.; Poormolaie, N.; Taghavi, F.; Poursasan, N.; Sanchooli, M.; et al. Comparative study of thermal domains analyzing of glycated and non-glycated human serum albumin. *Thermochim. Acta* **2014**, *594*, 24–30. [[CrossRef](#)]
36. Pelton, J.T.; McLean, L.R. Spectroscopic Methods for Analysis of Protein Secondary Structure. *Anal. Biochem.* **2000**, *277*, 167–176. [[CrossRef](#)]
37. Provencher, S.W.; Glöckner, J. Estimation of protein secondary structure from circular dichroism. *Biochemistry* **1981**, *20*, 33–37. [[CrossRef](#)] [[PubMed](#)]
38. Kurt, B.; Temel, H. Parametrization of boronates using VFFDT and paramfit for molecular dynamics simulation. *Molecules* **2020**, *25*, 2196. [[CrossRef](#)]
39. Kurt, B.; Temel, H. Development of AMBER parameters for molecular dynamics simulations of boron compounds containing aromatic structure. *Chem. Phys. Lett.* **2021**, *775*, 138656. [[CrossRef](#)]
40. Case, D.A.; Cerutti, D.S.; III Cheatham, T.E.; Darden, T.A.; Duke, R.E.; Giese, T.J.; Gohlke, H.; Goetz, A.W.; Greene, D.; Homeyer, N.; et al. *AMBER 2017*; University of California: San Francisco, CA, USA, 2017.
41. Potthoff, J.; Bojarski, K.K.; Kohut, G.; Lipska, A.G.; Liwo, A.; Kessler, E.; Ricard-Blum, S.; Samsonov, S.A. Analysis of Procollagen C-Proteinase Enhancer-1/Glycosaminoglycan Binding Sites and of the Potential Role of Calcium Ions in the Interaction. *Int. J. Mol. Sci.* **2019**, *20*, 5021. [[CrossRef](#)] [[PubMed](#)]
42. Maier, J.A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K.E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713. [[CrossRef](#)]
43. Wang, J.; Wolf, R.M.; Caldwell, J.W.; Kollman, P.A.; Case, D.A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174. [[CrossRef](#)]
44. Bayly, C.I.; Cieplak, P.; Cornell, W.; Kollman, P.A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model. *J. Phys. Chem.* **1993**, *97*, 10269–10280. [[CrossRef](#)]
45. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38. [[CrossRef](#)]

PUBLICATION D6

**INDUCED CIRCULAR DICHROISM AS A TOOL TO
MONITOR THE DISPLACEMENT OF LIGANDS BETWEEN
ALBUMINS**



Contents lists available at ScienceDirect

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy

journal homepage: www.elsevier.com/locate/saa

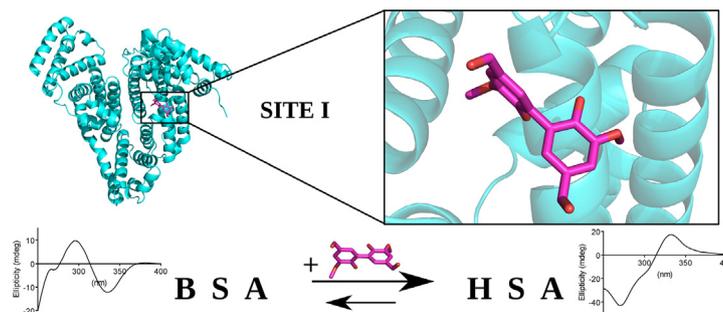
Induced circular dichroism as a tool to monitor the displacement of ligands between albumins

Luiza de Carvalho Bertozo^{a,1}, Małgorzata Kogut^{b,1}, Martyna Maszota-Zieleniak^{b,1}, Sergey A. Samsonov^{b,*}, Valdecir F. Ximenes^{a,*}^a Department of Chemistry, Faculty of Sciences, UNESP – São Paulo State University, 17033-360 Bauru, São Paulo, Brazil^b Faculty of Chemistry, University of Gdańsk, Wita Stwosza 63, 80-308 Gdańsk, Poland

HIGHLIGHTS

- The induction of chirality (ICD) is a tool for studying protein–ligand interactions.
- Dimers of methyl vanillate and vanillin are susceptible to ICD in albumin.
- The complexation in bovine and human albumins generates inverted ICD signs.
- The movement of the dimers between albumins can be monitored by inversion of ICD.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 17 January 2022

Received in revised form 27 April 2022

Accepted 9 May 2022

Available online 14 May 2022

Keywords:

Induced circular dichroism
Albumin
Axial chirality
Biaryl compounds
Molecular dynamics
Quantum mechanics

ABSTRACT

The induction of chirality in a ligand can be a powerful analytical tool for studying protein–ligand interactions. Here, we advanced by applying the technique to monitor the inversion of the induced circular dichroism (ICD) spectrum when ligands move between human and bovine serum albumin proteins (HSA and BSA). ICD experiments were performed using dimers of methyl vanillate (DVT) and vanillin (DVN). The sign and spectra shape were dependent on the albumin type. DVN presented a positive maximum in 312 nm when complexed with HSA and a negative one in BSA. It was possible to induce and follow the time-dependent displacement of the ligand from BSA ($2.2 \times 10^6 \text{ M}^{-1}$) to HSA ($6.6 \times 10^5 \text{ M}^{-1}$) via ICD inversion. The Molecular Mechanics Generalized Born Surface Area approach was used to calculate the binding free energy of the conformers, and a dissociation pathway for each system was proposed using Umbrella Sampling calculations. Four energy minima dihedral angle conformers were identified, and the corresponding CD spectra were calculated using the quantum chemistry approach. Then, weighted spectra for the conformationally accessible conformers were obtained based on each conformer's Boltzmann probability distribution. In conclusion, the methodology described in the manuscript might be helpful in monitoring the movement of ligands between proteins that they bind.

© 2022 Elsevier B.V. All rights reserved.

1. Introduction

The induction of optical activity in non-chiral compounds is intrinsically associated with intermolecular forces and the microenvironment (chiral solvents, proteins, DNA, membranes, cyclodextrin, etc.). This photophysical phenomenon has extensive

* Corresponding authors.

E-mail addresses: sergey.samsonov@ug.edu.pl (S.A. Samsonov), valdecir.ximenes@unesp.br (V.F. Ximenes).¹ These authors contributed equally.

applications in studying host–guest complexation and phenomena related to supramolecular chemistry [1–8]. Specifically, in protein–ligand analyses, the chirality induction directly shows the complexation [7,9,10]. The emerging circular dichroism signal is the induced circular dichroism (ICD) spectrum from this interaction. As generally defined, ICD is the effect of perturbations of the symmetry of non-chiral molecules resulting in the generation of non-zero rotational strengths, the theoretical counterpart of the circular dichroism phenomenon [1,11]. In short, the overall system symmetry perturbation can be correlated with the protein–ligand complex formation. Typically, the ligand is not a CD-detectable species in the bulk solution since it can be interconverted in enantiomeric conformers by single bond rotation. However, when the protein preferentially binds a specific conformer, the overall symmetry is broken, and ICD emerges.

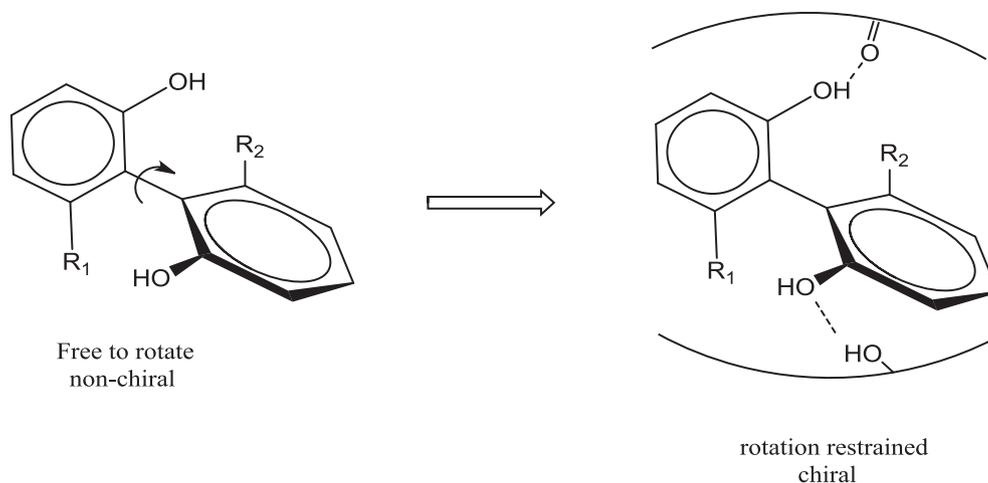
ICD has become a powerful analytical tool in supramolecular chemistry. Some recent applications can be listed as follows: chirally-arranged molecular assemblies of non-chiral cationic gemini surfactants detected by ICD [12]; interaction of TEMPO (2,2,6,6-tetramethylpiperidine 1-oxyl) radicals with cyclodextrins [13]; the study of phosphine gold(I) aryl acetylide urea complexes through [14]; the correlation between the ICD of Congo red and the formation of amyloid fibrils in albumin [15]. A family of molecules susceptible to ICD in host–guest associations is the biaryl compounds. This class of molecules, even in the absence of a chiral center, can be optically active due to the energetic restraint for the free rotation of the single bond that connects the aryl rings, i.e., to present axial chirality. However, most are not optically active once the energetic barrier is low enough to allow free rotation at room temperature [16]. In other words, chiral conformers, generated by the energetically free rotation around the single bond that connects the aromatic rings, exist in pairs in the bulk solution. This condition might be altered in the protein–ligand association when a preferential axial enantiomer is bound [17–20]. Scheme 1 exemplifies the ICD generation in the biaryl system due to its interaction in a chiral host. From these statements, it can be concluded that ICD has excellent potential as an analytical tool. However, it is worth noting that some practical aspects must be fulfilled to exploit this methodology. The ICD bands are usually of low intensity compared to the intrinsic far-UV CD signal of proteins. As such, ICD bands below 250 nm are not helpful due to the spectral superposition. Moreover, if the studied ligand is a chiral molecule, then a potentially useful ICD signal is that beyond the wavelength range of the intrinsic CD spectrum.

Human serum albumin (HSA) makes up about 60% of the blood plasma proteins. Besides several physiological functions, such as supporting the oncotic pressure, this protein acts as a drug carrier for numerous endogenous and exogenous compounds; it directly influences the drug's pharmacokinetics and toxicokinetics [21–24]. To this task, hydrophobic cavities, spread in the structural domains of HSA, are easily accessed by endogenous and exogenous compounds. These hydrophobic cavities present some degree of specificity. Hence, a pocket located in the subdomain IIA is named drug-binding site-I, Sudlow's site I, or yet warfarin binding site, due to its specificity to the anticoagulant. The other widely studied and well-established, located in the subdomain IIIA, is named drug-binding site II, Sudlow's site II or benzodiazepine binding site, due to its higher specificity to diazepam and correlates [24,25]. Serum albumin is ubiquitous in mammals. The serum albumins of different species, even though with a high degree of homology, have particularities. For instance, differences in the drug binding affinities have been reported [26–29]. This subject is relevant since it is quite common to use albumin from different species, for instance, bovine (BSA), as a model for the human protein. Here, we contribute to the subject using ICD spectroscopy to study and compare the interactions of biaryl ligands using HSA and BSA. We demonstrated an inversion of the ICD sign and its dependence on the protein organism source for the first time. This spectroscopic property might be beneficial to follow the movement of a ligand between the homologous proteins and open a novel investigative methodology.

2. Experimental and theoretical methods

2.1. Chemicals and solutions

Bovine serum albumin (BSA) fatty acid-free and essentially globulin free (A7030), human serum albumin (HSA) fatty acid-free and essentially globulin free (A3782), warfarin, ibuprofen, digitoxin, diazepam, and phenylbutazone were purchased from Sigma-Aldrich Chemical Co. (St. Louis, MO, USA). 6,6'-dihydroxy-5,5'-dimethoxy-[1,1'-biphenyl]-3,3'-dicarboxylate (dimer of methyl vanillate, DVT) and 6,6'-dihydroxy-5,5'-dimethoxy-[1,1'-biphenyl]-3,3'-dicarbaldehyde (dimer of vanillin, DVN) were available in our laboratory and synthesized as described [30,31]. Stock solutions of DVT, DVN, warfarin, ibuprofen, digitoxin and phenylbutazone (10 mmol L^{-1}) were prepared in dimethyl sulfox-



Scheme 1. The concept of chirality induction in a biaryl compound due to the complexation in a chiral host.

ide. Albumins were dissolved in 50 mmol L⁻¹ phosphate buffer at pH 7.0 to give a 1.0 mmol L⁻¹ stock solution.

2.2. Complexation and binding affinity studies

The ICD signal was used to evaluate complexations of DVT, DVN with HSA, and BSA. The reaction mixtures were constituted by HSA or BSA (30 μM) and the ligands DVT or DVN (30 μM) in a medium formed by 50 mmol L⁻¹ phosphate buffer, pH 7.0 at 25 °C. The study of binding sites preferences was performed by titration of reaction mixtures with albumin binding-site markers. The albumin binding-site markers were warfarin, ibuprofen, diazepam, digitoxin, and phenylbutazone. The measurements were performed on a Jasco J-815 spectropolarimeter equipped with a thermostatically controlled cell holder (Jasco, Japan). The spectra were obtained with 1 nm step resolution, 1 s of response time, three accumulations, and a 50 nm/min scanning speed using a 3 mL quartz cuvette with a 10 mm path length and a magnetic stirrer. The baseline (50 mmol L⁻¹ phosphate buffer) was subtracted from all measurements. The mixtures were incubated for 5 min at 25 °C before the spectra acquisition. For the thermal denaturation studies, the ICD signal was monitored at fixed wavelengths at 314 nm and 334 nm to DVT and DVN, respectively. The temperature was raised at 2.5 °C/min, and remained for 5 min in the targeted before measurement.

2.3. Molecular docking simulations

The structures of BSA and HSA were obtained from the Protein Data Bank (PDB ID: 4F5S, 2.47 Å, and 5 × 52, 3.00 Å, respectively) [32,33]. These two structures correspond to complete sequences of the soluble mature proteins with high enough resolution to be used as a starting structure in the docking and following MD simulations. The structures of the ligands DVT and DVN were prepared in Avogadro [34,35]. The ligand charges were obtained by RESP procedure [36] in an antechamber module of AMBER16 and further used for docking. Autodock3 software [37] was used to perform molecular docking. Grid box size of 60 Å × 48 Å × 66 Å was used that fully covered either site-I or site-II of BSA or HSA protein with the default grid step of 0.375 Å. The Lamarckian genetic algorithm was deployed for 1000 independent runs. The size of 300 for the initial population of 10⁵ generations for terminations conditions was chosen 9995 × 10⁵ energy evaluations were performed. Clustering was performed using the DBSCAN algorithm [38] on the top of 50 docking poses. One cluster of each system was formed, and five structures were chosen for further analysis. In total, eight independent systems have been analyzed.

2.4. Molecular dynamics simulations

All-atom molecular dynamic (MD) simulations of complexes obtained from molecular docking were performed using the AMBER16 software package [39]. A truncated octahedron TIP3P periodic box of 10 Å water layer from the box's border to the solute was used to solvate the complexes. The charge was neutralized with Na⁺ counterions. Energy minimization was carried out in two steps: 500 steepest descent cycles and 10³ conjugate gradient cycles with 100 kcal mol⁻¹ Å⁻² harmonic force restraints; 3 × 10³ steepest descent cycles and 3 × 10³ conjugate gradient cycles without any restraints. Following the minimization steps, the system was equilibrated at 300 K and 10³ Pa in an isothermal, isobaric ensemble for 500 ps. Afterward, the prediction MD was carried out in the same isothermal, isobaric ensemble using ff99SB [40] and gaff [41] force fields for 25 ns for the protein and ligand parts. The particle mesh Ewald method for treating electrostatics and

the SHAKE algorithm for the covalent bonds containing hydrogen atoms were implemented in the MD simulations.

2.5. Binding free energy calculations

Energetic post-processing of the trajectories and per-residue energy decomposition was performed for all the systems using Molecular Mechanics Generalized Born Surface Area (MM-GBSA) [42] and a model with the surface area and Born radii as default parameters as implemented in igb = 2 model in AMBER16. All frames from MD simulations were analyzed. Additionally, the MM-GBSA analysis evaluated binding free energies corresponding to the minima found in the Umbrella Sampling calculations (see next section).

2.6. Umbrella sampling calculations

Dissociation pathway: Umbrella sampling (US) MD simulations were performed to compute the free energy along the dissociation pathway (reaction coordinate) in AMBER [43]. In US, a series of windows were evenly located along with the reaction coordinates, and conformational sampling in these windows was achieved by enforcing an external biasing potential. The coordinate reaction sampling in each window overlapped with those in the adjacent windows so that the unbiased potential mean force (PMF) was reproduced by removing the biasing potential. Once all the US MD simulations were finished, the data collected from the separate simulation windows were combined along with the reaction coordinates. Subsequently, these data were used to calculate the PMF profile for the whole structural transition process by the weighted histogram analysis method (WHAM) using the code developed by Alan Grossfield [44]. In total, 8 independent US simulations were performed for BSA-DVT, BSA-DVN, HSA-DVT, HSA-DVN for site-I and site-II. In each case, the initial distance between the centers of masses of the protein and the ligand was about 10 Å, and the final value was about 50 Å with the step of 1 Å, producing 40 windows in total (in each window, the MD simulation was carried out for 10 ns). In the WHAM, the tolerance of iteration was set to 0.001 and the temperature to 300 K.

2.7. Ligand rotation around the covalent bond between two aromatic moieties of DVN/DVT

The US protocol was applied to study the energetic profile of DVT/DVN orientation change in terms of the dihedral angle between the planes defined by the aromatic moieties of DVT/DVN on the BSA/HSA surfaces or in the absence of protein. Based on the MM-GBSA results, the complex with the lowest binding energy was chosen as a starting conformation (Table 1). Finally, the 25 ns MD simulation was performed for each window with the step size equal to 3°. The process produced 120 structures in total per simulation, and they were used for further analysis. The obtained dihedral angle values were then processed with WHAM. In this approach, the periodic coordinates system generated the PMF from 0° to 360° using 120 bins with a tolerance for reconstructing the PMF of 0.01 kcal/mol.

Table 1
Dissociation analysis based on US calculations for each complex.

Complex, protein–ligand (site)	Barrier, kcal/mol	Complex, protein–ligand (site)	Barrier, kcal/mol
HSA-DVT (I)	56.3	HSA-DVT (II)	57.6
HSA-DVN (I)	60.3	HSA-DVN (II)	46.5
BSA-DVT (I)	57.5	BSA-DVT (II)	64.8
BSA-DVN (I)	51.8	BSA-DVN (II)	59.9

2.8. Calculation of circular dichroism spectra

For each conformation corresponding to the four DVT/DVN free energy minima, GAUSSIAN09 [45] with the basis B3LYP/6-31G* was used for energy optimization and for calculating the circular dichroism spectra.

2.9. Data analysis and visualization

Data analysis and its graphical representation were done with the R-package [46]. Each trajectory was visualized with VMD [47].

3. Results and discussion

3.1. Protein-dependent ICD signal

Methyl divanillate (DVT) and divanillin (DVN) are dimers of methyl vanillate and vanillin, respectively (Fig. 1). These biaryl compounds have been explored in the context of their health-beneficial biological effects [30,48–50]. Besides, they are ligands of albumin and susceptible to ICD. It means that even though achiral molecules, CD spectra are observed when they are bonded in the protein cavities. This effect is explained by their biaryl structures, which enable the existence of axial chirality [1,11,18,30,49]. Thus, once inside the protein, a preferential axial conformation is obtained due to specific intermolecular interactions with the amino acids that surround the compound. The consequence is the appearance of an ICD signal [1,11,51].

Interestingly, for DVT and DVN, the ICD spectra' shape depends on the albumin type. As shown in Fig. 1A, in BSA, DVT acquired an ICD spectrum with a maximum of 306 nm on the positive side and 280 nm on the negative one. For HSA, the signal was inverted, with a peak at 324 nm on the negative side and 294 nm on the positive one. For DVN, the spectra were inverted and almost symmetric above 312 nm (Fig. 1B). These results suggest that different chiral conformations were stabilized in the protein cavities. It can be

explained by microenvironmental changes in the binding pockets of HSA and BSA, which, indeed, have been reported [21 and references therein]. Finally, below 290 nm, the analysis of ICD is not helpful since proteins have a strong CD signal below this value (see Supporting Information, Fig. S1). For this reason, our study was focused on the spectra above 290 nm.

3.2. Binding sites specificities and interactions

Even though some preferences for site-I have been reported [30,31], DVT and DVN have affinities to the albumins' major binding sites. This propensity can be noted by analyzing the alteration in the ICD spectra when ligands (drugs) were added to displace DVT and DVN from the proteins (Fig. 2). The site-marker drugs were: warfarin and phenylbutazone (site-I), ibuprofen and diazepam (site-II), and digitoxin (site-III) [52–55]. As shown, the addition of warfarin (WAR) induced the displacement of DVT in HSA. However, even applying a four-fold molar excess, the biaryl compound was not wholly removed from the protein cavity (Fig. 2A). Phenylbutazone (PBZ), also a site-I drug [24,56], did not displace DVT; instead, it induced an increase in the ICD band, and the band related to PBZ was also progressively increased (Fig. S2A). This result indicates that both DVT and PBZ were occupying site-I concomitantly. It is worth noting that PBZ is also susceptible to ICD. The PBZ interaction with albumin generates a positive ICD spectrum with a peak at 290 nm [57].

DVT was also sensitive to digitoxin (DIG), suggesting some affinity to the site-III of HSA [55] (Fig. 2C). On the other hand, ibuprofen (IBP, site-II) caused a peculiar effect, as instead of decreasing the ICD signal, an increase was obtained (Fig. 2B). It is indicative that DVT was not displaced and shows that IBP was simultaneously bound in the protein. It can be suggested that some alteration in the protein tridimensional structure caused by ibuprofen reflected the interaction with DVT provoking an increased ICD signal. Even though less intense, a similar effect was observed in diazepam (DIA), which shares with IBP specificity to site-II [24,58] (Fig. S2B).

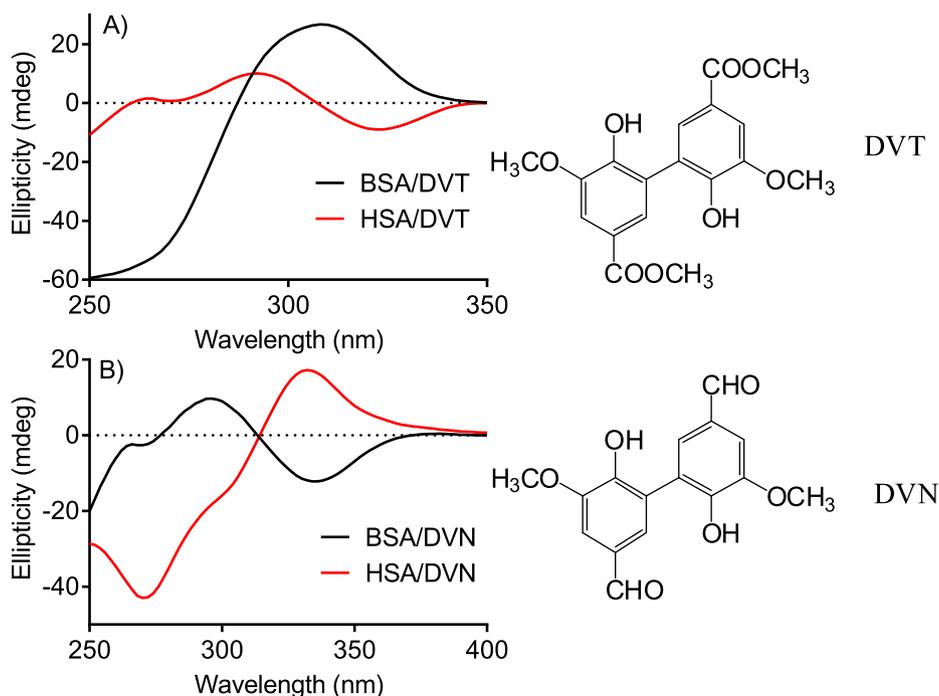


Fig. 1. The molecular structures of methyl divanillate (DVT) and divanillin (DVN) and their induced circular dichroism spectra (A, DVT) and (B, DVN) were acquired by complexation in HSA and BSA. Experimental condition: DVT, DVN, HSA and BSA (30 μ M), phosphate buffer, pH 7.0, 50 mM, 25 $^{\circ}$ C.

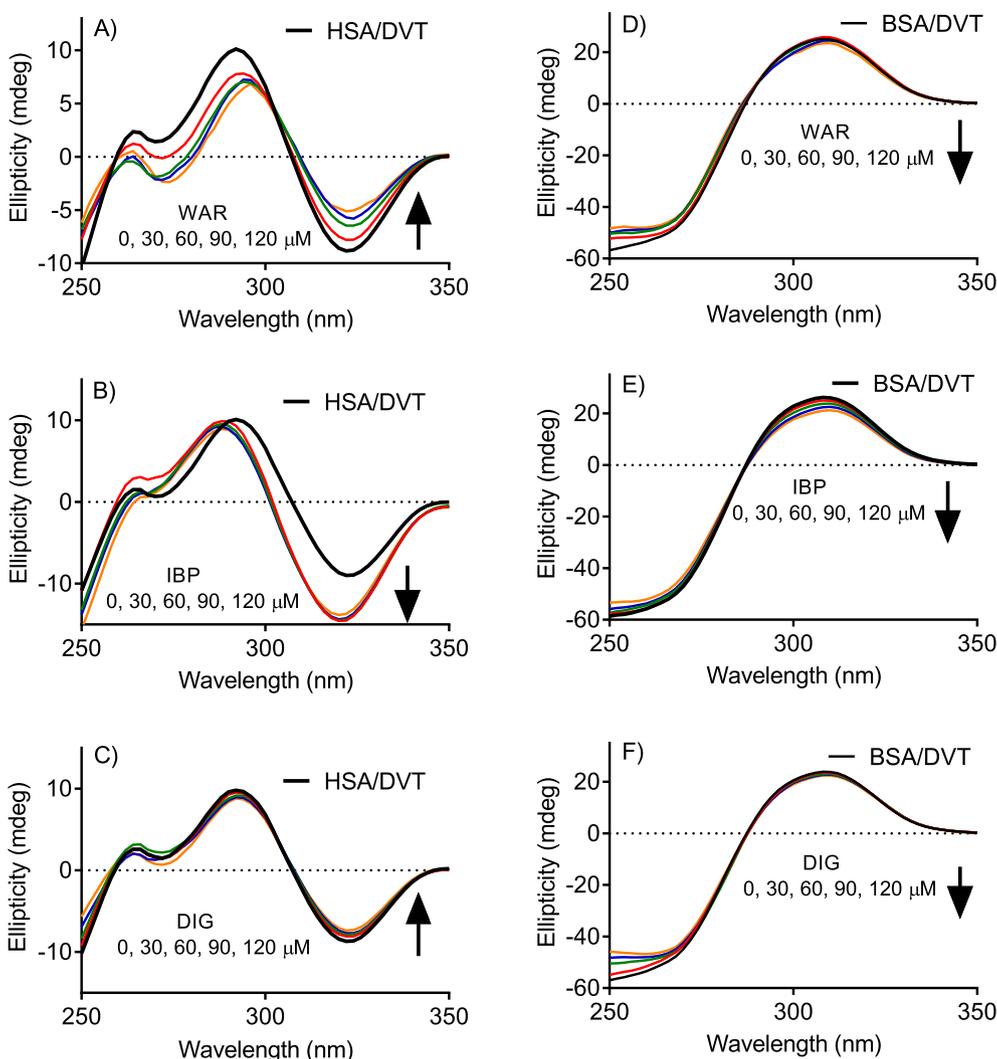


Fig. 2. Effect of pharmaceutical drugs on the ICD spectrum of DVT in HSA and BSA. (A, D) addition of warfarin, (B, E) addition of ibuprofen, (C, F) addition of digitoxin. Experimental conditions: DVT, HSA, and BSA (30 μ M) in phosphate buffer, pH 7.0, 50 mM, 25 $^{\circ}$ C.

Regarding the effects on BSA, the pharmaceutical drugs provoked similar, even though less intense, effects compared to HSA (Fig. 2D–2F and Fig. S2C–S2D). A more significant difference was obtained for IBP since a decrease in the signal was observed instead of increasing, as noted in HSA (Fig. 2E). For DVN, similar results were obtained. Again, WAR induced a more effective displacement, and IBP caused the intensification of the ICD spectra. The significant difference compared to DVT was that BSA, and not HSA, was the more susceptible protein to these pharmaceuticals (Fig. S3A–S3F).

Overall, these findings strengthen the concept that DVT and DVN can bind in all the major drug-binding sites of albumin. Besides, due to the absence of spectral superposition and the selectivity of ICD spectra, the findings unequivocally showed that more than one ligand could simultaneously bind to albumin. These results also showed that the interaction of a pharmaceutical drug could change the protein's tridimensional structure. As shown, IBP induced the intensification of the DVT and DVN signals. In other words, the complexation of IBP at site-II altered the tridimensional structure of the albumin and, consequently, the intermolecular interactions of the DVT and DVN. A possible interpretation would be an increase in the binding affinity of these ligands induced by IBP. However, it does not seem to be the case

since the presence of IBP, instead of a boost, yielded a slight decrease in the association constant of DVN with BSA (Fig. S4). Hence, the explanation must rely on altering the spatial arrangement of the amino acids that interact with DVT and DVN and the consequent alteration in their chiral conformation inside the protein. In short, the effect of IBP can be interpreted as an allosteric-like effect. Indeed, cooperativity and allosteric modulation among binding sites have been described to albumins [59–61].

3.3. ICD as a tool for protein denaturation study

Additional evidence of the interaction between the studied ligands and albumins was obtained by measuring their displacement caused by temperature increases. The removal of DVT and DVN from the protein cavities resulted in the loss of the ICD spectrum. The ICD signal decrease followed a typical thermal-induced denaturation curve of proteins (Fig. 3A–3D) [62]. A denaturation temperature (T_m) of around 60 $^{\circ}$ C was obtained from this curve. This value is close to reported ones for albumins, usually obtained by monitoring intrinsic protein CD intensity loss at 222 nm [62]. From these findings, it can be concluded that the removal of the compounds followed the denaturation of the protein. Both DVT and DVN were equally effective.

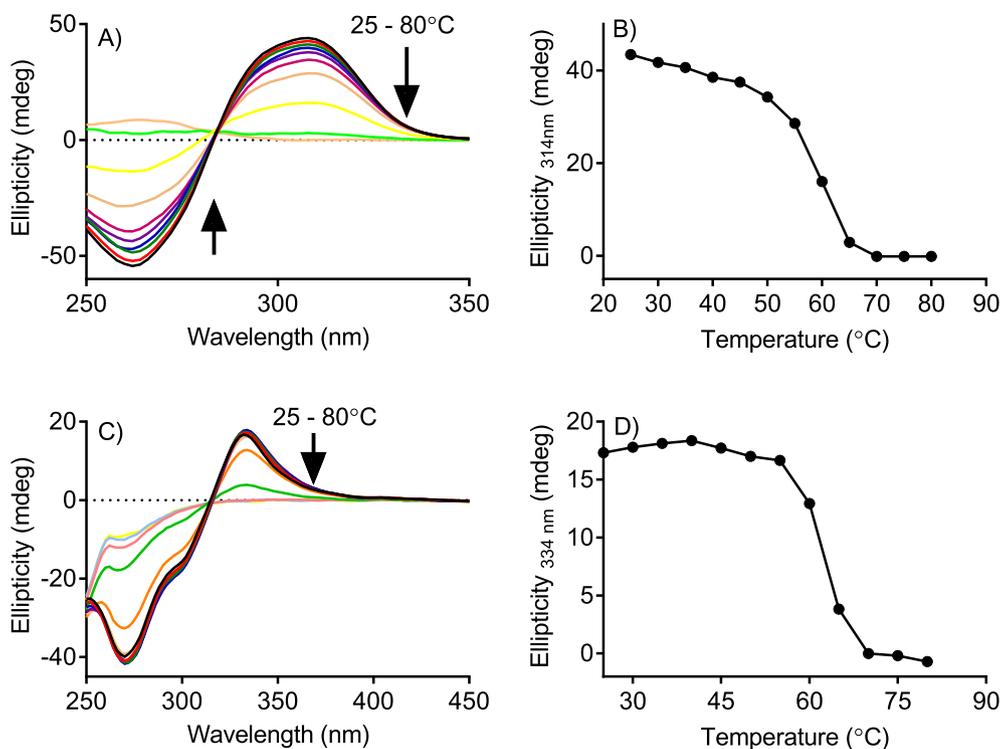


Fig. 3. Thermal induced denaturation of BSA and HSA followed by ICD. (A,B) Denaturation of BSA monitored by DVT's ICD. (C,D) Denaturation of HSA monitored by DVN's ICD. Experimental condition. Albumin, DVT, and DVN 30 μM in 50 mM, pH 7.0 phosphate buffer. Temperature raised at 2.5 $^{\circ}\text{C}/\text{min}$, remained for 5 min in the targeted temperature before spectra record.

3.4. Competition: BSA versus HSA

The study of the interaction of DVT or DVN with albumins can help monitor the movement of these ligands between the proteins. This technique could be beneficial for studying alteration in proteins and their mediums. The following findings justify this statement:

- (i) The ICD spectra are exclusively linked to the host-guest interaction.
- (ii) Induced chirality is not present in the majority of the micro-molecules; hence, the ICD spectra are less susceptible to spectral superposition and inner filter effect, which are common in absorption and fluorescence-based techniques.
- (iii) The ICD spectral shape of DVT and DVN was dependent on the albumin type.
- (iv) HSA has a higher affinity for DVT and DVN than BSA. Regarding the last statement, the binding affinities of DVT with HSA and BSA were reported as $7.6 \times 10^5 \text{ M}^{-1}$ and $3.5 \times 10^5 \text{ M}^{-1}$, respectively [29]. For DVN, the measurements were performed in this work and resulted in $2.2 \times 10^6 \text{ M}^{-1}$ and $6.6 \times 10^5 \text{ M}^{-1}$ to HSA and BSA, respectively (Fig. S5).

Based on these properties, we studied the alteration in the ICD spectra provoked by the simultaneous addition of both proteins in the medium. The complex BSA-DVT was titrated with HSA and vice-versa. As shown, the addition of HSA in the complex BSA-DVT provoked progressive alteration in the ICD spectrum of DVT, revealing its displacement from BSA to HSA (Fig. 4). This result is a clear indication of the higher affinity of DVT to HSA compared to BSA and corroborated with the binding constants as described above. In agreement, the opposite was not observed, i. e., BSA was not able to invert the ICD spectrum of DVT in HSA (Fig. S6).

The displacement of DVN from BSA to HSA was also clearly detected by the change in the ICD spectrum. In this case, a technical advantage of DVN compared to DVT was its almost symmetrical and inverted bands above 314 nm (Fig. 5A). After adding only 10 μM of HSA, the signal of 30 μM of BSA was canceled, and at equimolar concentrations, the ICD spectrum was utterly inverted, showing that DVN complexed with BSA was displaced by HSA (Fig. 5B and 5C). The isosbestic point at 314 nm is consistent with an equilibrium $\text{BSA} \leftrightarrow \text{DVN} \leftrightarrow \text{HSA}$ (Fig. 5B). Fig. 5D shows the time-dependent displacement of DVN from BSA to HSA. Finally, corroborating the previous results, BSA could not remove DVN from HSA (Fig. 5E).

3.5. Molecular modeling

Molecular docking was initially performed to investigate other potential binding sites in BSA/HSA for the analyzed ligands, apart from site-I and site-II. In each investigated system, only one cluster of structures was obtained, which suggests that once the ligand is near either site-I or site-II of BSA/HSA, there are no other sites where a ligand would preferentially bind. The representative five structures for each system were obtained and used for further MD simulation (RMSD of the MD simulations are provided in Supplementary Fig. S7) and free energy analysis. Binding free energy (ΔG) computation can play a crucial role in prioritizing compounds to be evaluated experimentally on their affinity for target proteins. The more negative the ΔG value is, the more stable the complex is. One of the most popular end-point methods, MM-GBSA, was used in this study. The results obtained from the MM-GBSA calculations are summarized in Fig. 6 and Table S1 in Supporting Information.

In addition to the MM-GBSA free energy calculations, a ligand dissociation pathway and its energy analysis for each system were proposed with the US calculations. This technique involves pulling

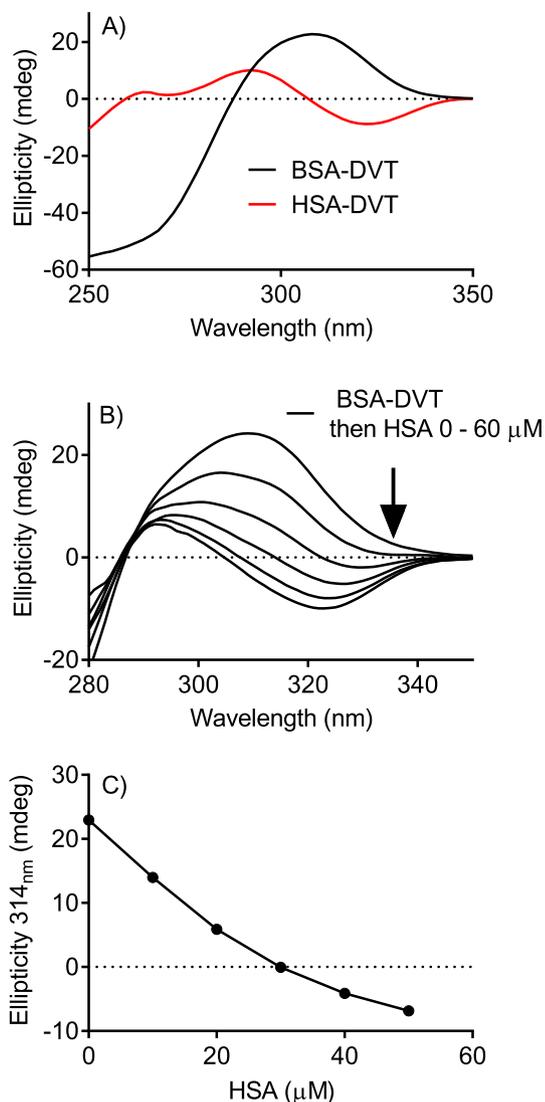


Fig. 4. Monitoring the displacement of DVT from BSA to HSA by inversion of ICD. (A) ICD spectra of DVT (30 μM) in BSA (30 μM) and HSA (30 μM) in 50 mM, pH 7.0 phosphate buffer. (B,C) Titration of the complex BSA (30 μM) - DVT (30 μM) with HSA monitored by DVT's ICD.

away a ligand from its binding partner and providing an energetic barrier that a ligand needs to overcome, expressed in PMF. Therefore, the higher the energetic barrier, the less probable the dissociation in the complex (Table 1). The following conclusions can be drawn from the US calculations:

- (i) HSA-DVN binding is stronger than BSA-DVN according to the dissociation but not the MM-GBSA analysis.
- (ii) Site-I of HSA is preferred by both DVN and DVT based on the MM-GBSA calculations and to DVN according to the dissociation analysis. For DVT, dissociation analysis yields very similar results for both sites.
- (iii) The calculations showed that site-II of the BSA is preferred, which does not agree with the literature and experimental data.

All the above findings should be addressed carefully since the dissociation analysis could not be expected to reflect the mechanical behavior in the analyzed systems appropriately. When a ligand is pulled away from the binding site and the protein's backbone is restrained, the modeled dissociation description is not entirely physically correct since, in the absence of the restraints, the dissociation does not occur in the simulation. Due to the applied restraints, the protein is partially distorted, which could substantially impact the results in terms of the PMF profile.

3.6. Dihedral angle analysis: Minimum energy conformations

For DVN and DVT bound in site-I of BSA and HSA, the change of energy of the system was analyzed depending on the dihedral angle between the planes of the ligands aromatic moieties, which defined chirality and compared with the energy dependence on the same angle for the unbound molecules. The complexes with the lowest ΔG values were chosen based on the results shown in Table S1. The starting point for each US MD simulation was at the value of the dihedral angle near 120° . This dihedral angle was measured for the complex with the lowest ΔG based on the MM-GBSA calculations. For clarity, a range between -180° and 180° for the plots was chosen (Fig. 7). Therefore, the minimum 4 was the initial energetic value for each complex. Considering the energetic barriers, only minima 3 and 4 should be considered since it is improbable for the ligand to overcome the 15 kcal/mol barrier separating these minima from two other ones. It is worth noticing that only for the BSA-DVT complex, the global minimum was the minimum 3 instead of 4. Altogether, these findings support that the studied ligands bind with higher affinity to the HSA than BSA, which agreed with the experimental results.

The free energy minima obtained by the dihedral angle sampling were calculated using the MM-GBSA approach (Table 2). The most contributing residues were defined and summarized in Tables S2-S5 and Figs. 8 and 9. The MM-GBSA energy for minima obtained by the dihedral angle rotation does not agree with the US-derived energy values, highlighting the fundamental differences in the applied methodologies. Nevertheless, from the MM-GBSA calculations, the per residue decomposition was extracted and allowed to find the most important amino acid binders. These residues are very similar independently of the angle value.

3.7. Quantum chemical calculation: Simulation of CD spectra

The CD spectra for DVT and DVN were obtained from quantum chemical calculations. Fig. S8 shows the CD spectra for DVN and DVT calculated for the four conformations corresponding to the free energy minima using GAUSSIAN09. Then, weighted spectra for the minima 3 and 4 were obtained based on each conformer's Boltzmann probability distribution (based on PMF, Fig. 7). Clearly, the weighted spectra corresponding to minima 3 and 4 are different (Fig. 10). From the fact that for HSA and BSA, different conformations in different minima are dominant, corresponding differences in spectra are expected for the complexes with these two proteins. For DVT, the weighted spectra fit well with the experimental data regarding the sign of the difference originating from the protein organism source (compare Fig. 1 and Fig. 8). For DVN, the weighted spectra were also dependent on the albumin organism but did not follow the experimental result.

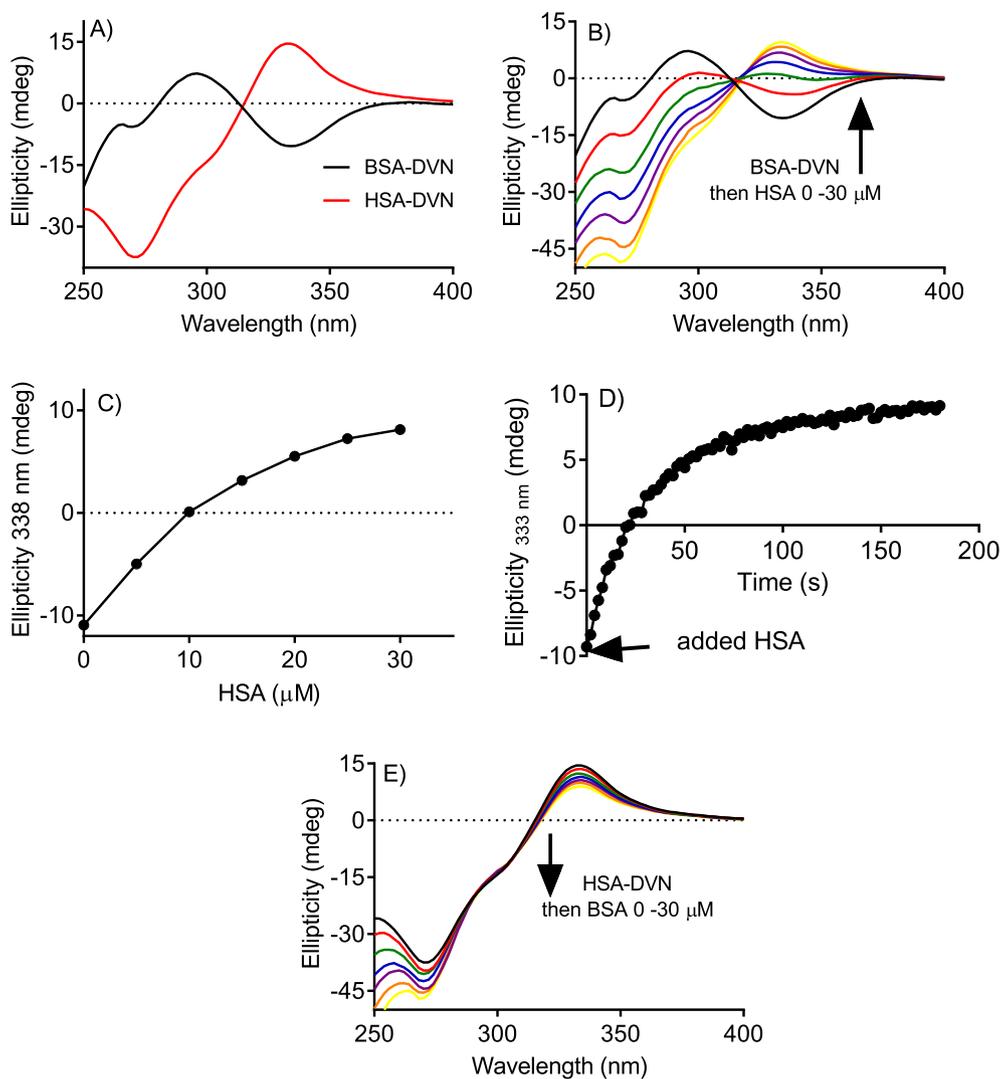


Fig. 5. Monitoring the displacement of DVN from BSA to HSA by inversion of ICD. (A) ICD spectra of DVN (30 μM) in BSA (30 μM) and HSA (30 μM) in 50 mM pH 7.0 phosphate buffer. (B,C) Titration of the complex BSA (30 μM)-DVN (30 μM) with HSA monitored by DVN's ICD. (D) Kinetic of DVN displacement from BSA due to addition of HSA. (E) Titration of the complex HSA (30 μM)-DVN (30 μM) with BSA monitored by DVN's ICD.

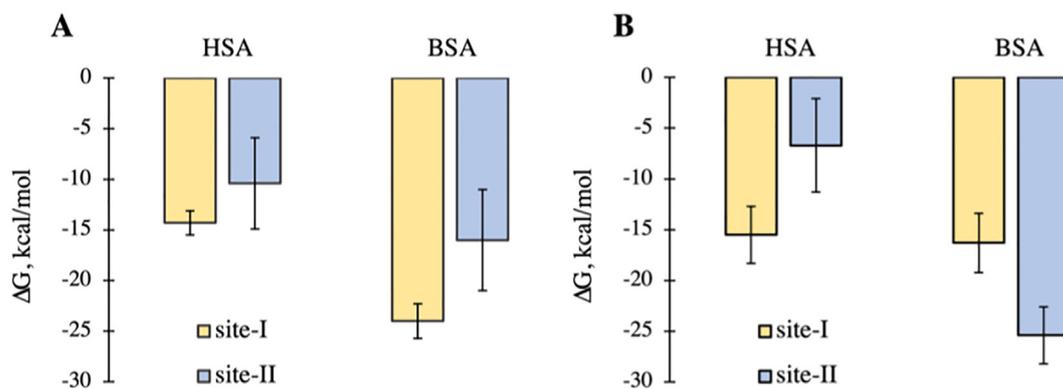


Fig. 6. Barplots represent free binding energy (ΔG) obtained from the MM-GBSA calculations for DVT-HSA/BSA (A) and DVN-HSA/BSA (B) complexes.

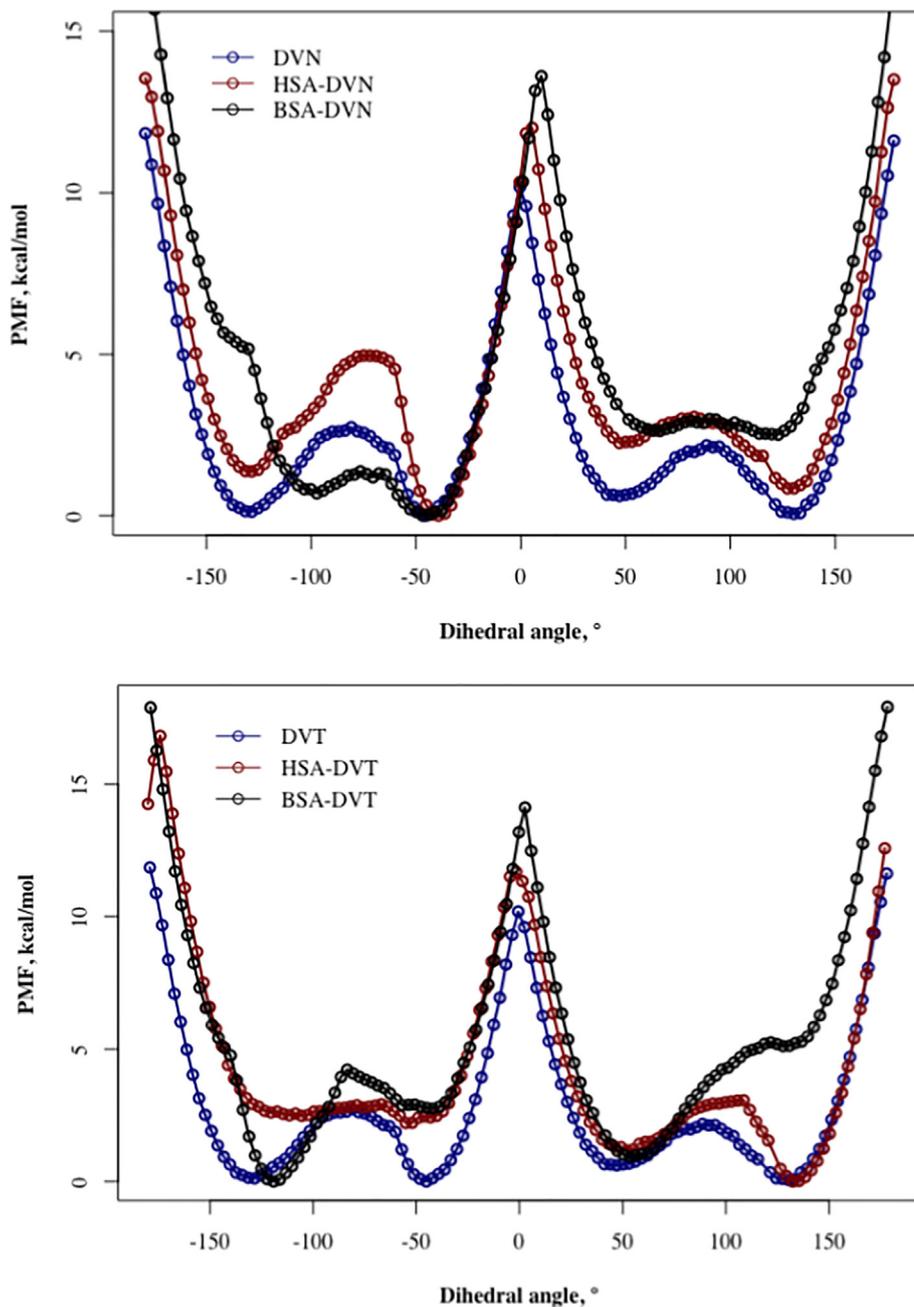


Fig. 7. PMF of DVN (up) and DVT (down) in the unbound state and complexed with BSA and HSA from the US MD simulations.

Table 2

Analysis of the free energy minima (MM-GBSA).

Minimum n°	MM-GBSA, kcal/mol			
	BSA-DVT	BSA-DVN	HSA-DVT	HSA-DVN
1	-18.3	-21.5 ± 2.3	-13.0	-14.6 ± 2.0
2	-19.0	-24.1 ± 2.0	-13.3	-19.8 ± 1.5
3	-26.7	-21.2 ± 3.1	-12.2	-15.6 ± 2.1
4	N/A	-25.2 ± 2.1	-15.3	-16.6 ± 1.9

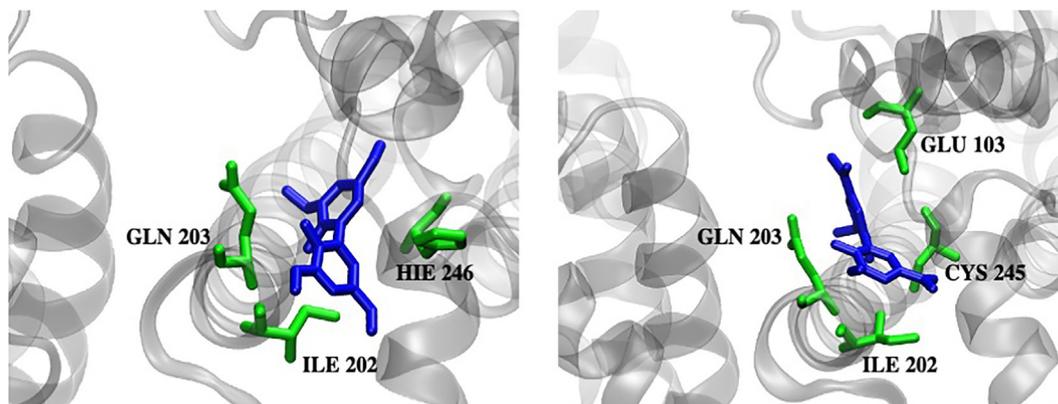


Fig. 8. Most energetically contributing residues (similar for all minima) in the complex DVN-BSA (left) and DVT-BSA (right).

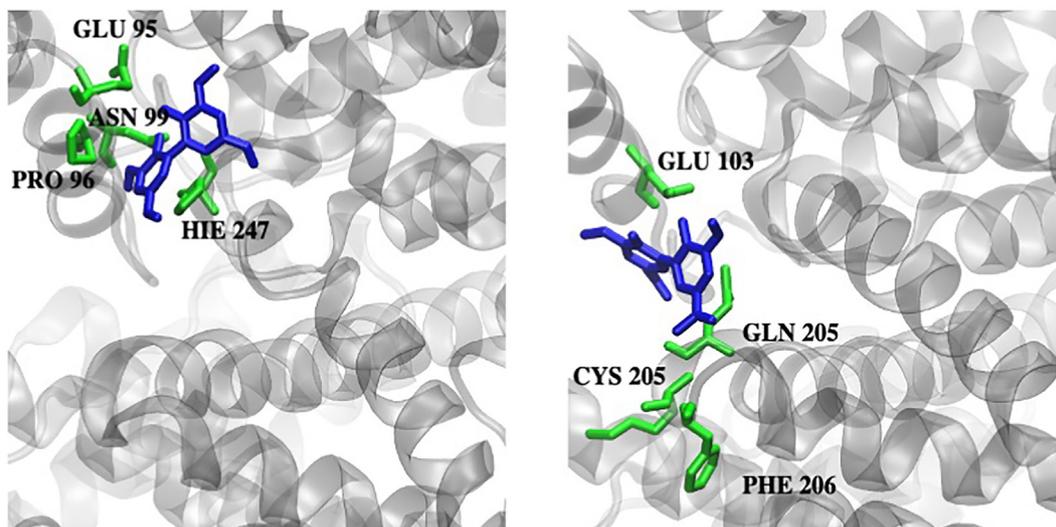


Fig. 9. Most energetically contributing residues (similar for all minima) in the complex DVN-HSA (left) and DVT-HSA (right).

4. Conclusions

The induction of chirality in a ligand due to the restriction of its conformational space upon binding in the binding pocket of a protein can be a powerful analytical tool for studying protein–ligand interaction. First, the phenomenon itself is an unequivocal demonstration of the complexation. Indeed, it is generated by interactions with the amino acid residues of the binding site, leading to stabilizing a specific ligand conformation that, outside of the protein, would not be preferential and, therefore, not chiral. Second, due to its nature, the spectra of circular dichroism are usually less susceptible to spectral superposition and inner filter effects, which are a drawback in molecular spectroscopy-based techniques. Here, a new feature related to ICD was discovered, providing an advance in its application to study albumin and its ligands. As we have demonstrated, it was possible to monitor the movement of DVT and DVN from BSA to HSA using the inversion of their ICD signals. In other words, a competition of two proteins by the ligand was monitored by CD in real-time. As the phenomenon is related to

the protein structural integrity, it could be applied to study other factors that potentially affect the outcome and susceptibility of albumins to the medium, the presence of different ligands, and structural alterations. Using MD to estimate the contribution of minima energy conformers and quantum mechanics to calculate the weighted CD spectra, the inversion of the sign was consistent with the experimental for DVT and, at least qualitatively, explained. We propose the application of ICD inversion as a procedure to study alteration in the structure of albumins and as a real-time method to compare the binding efficacy of different organism sources of this drug-carrier protein.

CRediT authorship contribution statement

Luiza de Carvalho Bertozo: Conceptualization, Methodology, Investigation. **Małgorzata Kogut:** Methodology, Investigation. **Martyna Maszota-Zieleniak:** Methodology, Investigation. **Sergey A. Samsonov:** Supervision, Resources, Data curation, Methodology, Writing – original draft, Writing – review & editing.

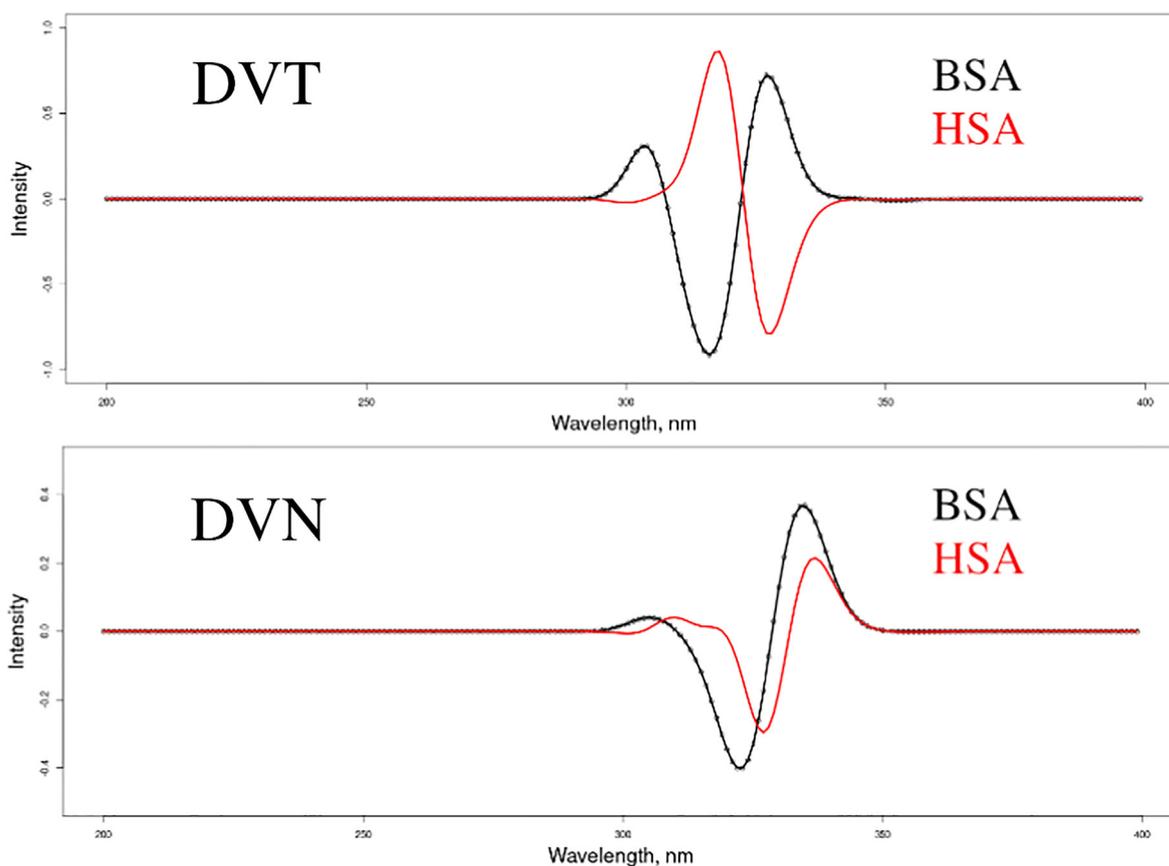


Fig. 10. Weighted CD spectra for DVT (up) and DVN (down) were calculated for the conformations corresponding to the free energy minima with GAUSSIAN09.

Valdecir Farias Ximenes: Conceptualization, Supervision, Resources, Methodology, Investigation, Data curation, 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.

Acknowledgments

The authors gratefully acknowledge financial support from National Council for Scientific and Technological Development (CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico, grant number: 303485/2019-1 and 163201/2020-0), National Institute of Science and Technology - INCT BioNat, grant number: 465637/2014-0, State of Sao Paulo Research Foundation (FAPESP, Fundação de Amparo à Pesquisa do Estado de São Paulo, grant number: 2019/18445-5), and from National Science Centre of Poland (UMO-2018/30/E/ST4/00037). The molecular dynamics simulations were performed on the PROMETHEUS cluster provided by Polish Grid Infrastructure (PL-GRID, grant plggagstr3) as well as on the local “piasek” cluster. We thank Prof. Satoru Tsushima (Helmholtz-Zentrum Dresden-Rossendorf, Germany) for technical support with the quantum chemical ICD calculations.

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.saa.2022.121374>.

References

- [1] D. Tedesco, C. Bertucci, Induced circular dichroism as a tool to investigate the binding of drugs to carrier proteins: Classic approaches and new trends, *J. Pharm. Biomed. Anal.* 113 (2015) 34–42, <https://doi.org/10.1016/j.jpba.2015.02.024>.
- [2] A.R. Holzwarth, E. Langer, H. Lehner, K. Schaffner, Absorption, Luminescence, Solvent-induced Circular Dichroism and H NMR Study of Bilirubin Dimethyl Ester: Observation of Different Forms in Solution, *Photochem. Photobiol.* 32 (1980) 17–26, <https://doi.org/10.1111/j.1751-1097.1980.tb03981.x>.
- [3] S. Müller, J. Paulus, J. Mattay, H. Ihmels, V.I. Dodero, N. Sewald, Photocontrolled DNA minor groove interactions of imidazole/pyrrole polyamides, *Beilstein J. Org. Chem.* 16 (2020) 60–70, <https://doi.org/10.3762/bjoc.16.8>.
- [4] N. Holmgaard List, J. Knoop, J. Rubio-Magnieto, J. Idé, D. Beljonne, P. Norman, M. Surin, M. Linares, Origin of DNA-Induced Circular Dichroism in a Minor-Groove Binder, *J. Am. Chem. Soc.* 139 (2017) 14947–14953, https://doi.org/10.1021/jacs.7b05994/suppl_file/ja7b05994_si_001.pdf.
- [5] O.F. Silva, N.M. Correa, J.J. Silber, R.H. De Rossi, M.A. Fernández, Supramolecular Assemblies Obtained by Mixing Different Cyclodextrins and AOT or BHDC Reverse Micelles, *Langmuir* 30 (2014) 3354–3362, <https://doi.org/10.1021/la404584q>.
- [6] E. Kiss, V.A. Szabó, P. Horváth, Simple circular dichroism method for selection of the optimal cyclodextrin for drug complexation, *J. Incl. Phenom. Macrocycl. Chem.* 95 (2019) 223–233, <https://doi.org/10.1007/s10847-019-00938-2/tables/6>.
- [7] A. Tramarin, D. Tedesco, M. Naldi, M. Baldassarre, C. Bertucci, M. Bartolini, New insights into the altered binding capacity of pharmaceutical-grade human serum albumin: site-specific binding studies by induced circular dichroism spectroscopy, *J. Pharm. Biomed. Anal.* 162 (2019) 171–178, <https://doi.org/10.1016/j.jpba.2018.09.022>.
- [8] W. Zuo, Z. Huang, Y. Zhao, W. Xu, Z. Liu, X.J. Yang, C. Jia, B. Wu, Chirality sensing of choline derivatives by a triple anion helicate cage through induced circular dichroism, *Chem. Commun.* 54 (2018) 7378–7381, <https://doi.org/10.1039/c8cc03883j>.
- [9] M. Losytskyy, N. Chornenka, S. Vakarov, S.M. Meier-Menches, C. Gerner, S. Potocki, V.B. Arion, E. Gumienna-Kontecka, Y. Voloshin, V. Kovalska, Sensing of Proteins by ICD Response of Iron(II) Clathrochelates Functionalized by Carboxyalkyl/sulfide Groups, *Biomol.* 10 (2020) 1602, <https://doi.org/10.3390/biom10121602>.
- [10] L.d.C. Bertozo, E.A. Pilot, A.N. Lima, P.T. de Resende Lara, A.L. Scott, V.F. Ximenes, Interaction between 1-pyrenesulfonic acid and albumin: Moving

- inside the protein, *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* 208 (2019) 243–254.
- [11] S. Allénmark, Induced circular dichroism by chiral molecular interaction, *Chirality* 15 (2003) 409–422, <https://doi.org/10.1002/CHIR.10220>.
- [12] Y. Okazaki, N. Ryu, T. Buffeteau, S. Pathan, S. Nagaoka, E. Pouget, S. Nlate, H. Ihara, R. Oda, Induced circular dichroism of monoatomic anions: silica-assisted the transfer of chiral environment from molecular assembled nanohelices to halide ions, *Chem. Commun.* 54 (2018) 10244–10247, <https://doi.org/10.1039/C8CC05449E>.
- [13] G. Ionita, S. Mocanu, I. Matei, Conformational preferences of TEMPO type radicals in complexes with cyclodextrins revealed by a combination of EPR spectroscopy, induced circular dichroism and molecular modeling, *Phys. Chem.* 22 (2020) 12154–12165, <https://doi.org/10.1039/D0CP01937B>.
- [14] J. Dubarle-Offner, J. Moussa, H. Amouri, B. Jouvet, L. Bouteiller, M. Raynal, Induced Circular Dichroism in Phosphine Gold(I) Aryl Acetylide Urea Complexes through Hydrogen-Bonded Chiral Co-Assemblies, *Chemistry* 22 (2016) 3985–3990, <https://doi.org/10.1002/CHEM.201504441>.
- [15] D.N. De Vasconcelos, V.F. Ximenes, Albumin-induced circular dichroism in Congo red: Applications for studies of amyloid-like fibril aggregates and binding sites, *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.* 150 (2015) 321–330, <https://doi.org/10.1016/j.saa.2015.05.089>.
- [16] G. Haberhauer, C. Tepper, C. Wölper, D. Bläser, Tropos, Nevertheless Conformationally Stable Biphenyl Derivatives, *European J. Org. Chem.* 2013 (2013) 2325–2333, <https://doi.org/10.1002/ejoc.201300087>.
- [17] I. Matei, S. Ionescu, M. Hillebrand, Kaempferol-human serum albumin interaction: Characterization of the induced chirality upon binding by experimental circular dichroism and TDDFT calculations, *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* 96 (2012) 709–715, <https://doi.org/10.1016/j.saa.2012.07.043>.
- [18] D. Venturini, B. Pastrello, M.L. Zeraik, I. Pauli, A.D. Andricopulo, L.C. Silva-Filho, V.S. Bolzani, N.H. Morgon, A.R. de Souza, V.F. Ximenes, Experimental, DFT and docking simulations of the binding of diapocynin to human serum albumin: induced circular dichroism, *RSC Adv.* 5 (2015) 62220–62228, <https://doi.org/10.1039/C5RA10960D>.
- [19] L. Di Bari, S. Ripoli, S. Pradhan, P. Salvadori, Interactions between quercetin and Warfarin for albumin binding: A new eye on food/drug interference, *Chirality* 22 (2010) 593–596, <https://doi.org/10.1002/CHIR.20794>.
- [20] B. Pastrello, G.C. dos Santos, L.C.D. Silva-Filho, A.R. de Souza, N.H. Morgon, V.F. Ximenes, Novel aminoquinoline-based solvatochromic fluorescence probe: Interaction with albumin, lysozyme and characterization of amyloid fibrils, *Dye. Pigment* 173 (2020), <https://doi.org/10.1016/j.dyepig.2019.107874>
- [21] D.A. Belinskaia, P.A. Voronina, V.I. Shmurak, R.O. Jenkins, N.V. Goncharov, Serum Albumin in Health and Disease: Esterase, Antioxidant, Transporting and Signaling Properties, *Int. J. Mol. Sci.* 22 (2021) 10318, <https://doi.org/10.3390/IJMS221910318>.
- [22] H. Rimac, Ž. Debeljak, M. Bojić, L. Miller, Displacement of Drugs from Human Serum Albumin: From Molecular Interactions to Clinical Significance, *Curr. Med. Chem.* 24 (2017) 1930–1947, <https://doi.org/10.2174/0929867324666170202152134>.
- [23] D. Pilati, K.A. Howard, Albumin-based drug designs for pharmacokinetic modulation, *Expert Opin. Drug Metab. Toxicol.* 16 (2020) 783–795, <https://doi.org/10.1080/17425255.2020.1801633>.
- [24] A. Shamsi, A. Ahmed, M.S. Khan, M. Al Shahwan, F.M. Husain, B. Bano, Understanding the binding between Rosmarinic acid and serum albumin: In vitro and in silico insight, *J. Mol. Liq.* 311 (2020), <https://doi.org/10.1016/j.molliq.2020.113348>
- [25] J. Ghuman, P.A. Zunszain, I. Petitpas, A.A. Bhattacharya, M. Otagiri, S. Curry, Structural Basis of the Drug-binding Specificity of Human Serum Albumin, *J. Mol. Biol.* 353 (2005) 38–52, <https://doi.org/10.1016/j.jmb.2005.07.075>.
- [26] I. Petitpas, A.A. Bhattacharya, S. Twine, M. East, S. Curry, Crystal Structure Analysis of Warfarin Binding to Human Serum Albumin: Anatomy of Drug Site I, *J. Biol. Chem.* 276 (2001) 22804–22809, <https://doi.org/10.1074/JBC.M100575200>.
- [27] M. Poór, Y. Li, G. Matisz, L. Kiss, S. Kunsági-Máté, T. Koszegi, Quantitation of species differences in albumin-ligand interactions for bovine, human and rat serum albumins using fluorescence spectroscopy: A test case with some Sudlow's site I ligands, *J. Lumin.* 145 (2014) 767–773, <https://doi.org/10.1016/j.jlumin.2013.08.059>.
- [28] T. Kosa, T. Maruyama, M. Otagiri, Species Differences of Serum Albumins: I. Drug Binding Sites, *Pharm. Res.* 14 (1997) 1607–1612, <https://doi.org/10.1023/A:1012138604016>.
- [29] S. Shityakov, A. Fischer, K.P. Su, A.A. Hussein, T. Dandekar, J. Broscheit, Novel Approach for Characterizing Propofol Binding Affinities to Serum Albumins from Different Species, *ACS Omega* 5 (2020) 25543–25551, https://doi.org/10.1021/ACSEOMEGA.0C01295/SUPPL_FILE/AO0C01295_SI_002.ZIP.
- [30] D.N. de Vasconcelos, A.N. Lima, E.A. Philot, A.L. Scott, I.A. Ferreira Boza, A.R. de Souza, N.H. Morgon, V.F. Ximenes, Methyl divanillate: redox properties and binding affinity with albumin of an antioxidant and potential NADPH oxidase inhibitor, *RSC Adv.* 9 (2019) 19983–19992, <https://doi.org/10.1039/C9RA02465D>.
- [31] D. Venturini, A.R. de Souza, I. Caracelli, N.H. Morgon, L.C. da Silva-Filho, V.F. Ximenes, J. Chamani, Induction of axial chirality in divanillin by interaction with bovine serum albumin, *PLoS One* 12 (6) (2017) e0178597.
- [32] A. Kawai, V.T.G. Chuang, Y. Kouno, K. Yamasaki, S. Miyamoto, M. Anraku, M. Otagiri, Crystallographic analysis of the ternary complex of octanoate and N-acetyl-L-methionine with human serum albumin reveals the mode of their stabilizing interactions, *Biochim. Biophys. Acta - Proteins Proteomics.* 2017 (1865) 979–984, <https://doi.org/10.1016/j.bbapap.2017.04.004>.
- [33] A. Bujacz, Structures of bovine, equine and leporine serum albumin, *Acta Crystallogr. Sect. D Biol. Crystallogr.* 68 (2012) 1278–1289, <https://doi.org/10.1107/S0907444912027047>.
- [34] Avogadro - Free cross-platform molecular editor - Avogadro, (n.d.), <http://avogadro.cc/> (accessed November 1, 2021).
- [35] M.D. Hanwell, D.E. Curtis, D.C. Lonie, T. Vandermeersch, E. Zurek, G.R. Hutchison, Avogadro: an advanced semantic chemical editor, visualization, and analysis platform, *J. Cheminformatics.* 4 (2012) 1–17, <https://doi.org/10.1186/1758-2946-4-17>.
- [36] C.I. Bayly, P. Cieplak, W. Cornell, P.A. Kollman, A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model, *J. Phys. Chem.* 97 (2002) 10269–10280, <https://doi.org/10.1021/j100142A004>.
- [37] G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson, Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, *J. Comput. Chem.* 19 (1998) 1639–1662, [https://doi.org/10.1002/\(SICI\)1096-987X\(19981115\)19:14](https://doi.org/10.1002/(SICI)1096-987X(19981115)19:14).
- [38] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, Comparison of multiple Amber force fields and development of improved protein backbone parameters, *Proteins* 65 (2006) 712–725, <https://doi.org/10.1002/PROT.21123>.
- [39] D.A. Case, D.S. Cerutti, T.E. Cheatham III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, D. Greene, N. Homeyer, S. Izadi, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, 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, D.M. York, P.A. Kollman, AMBER 2017, University of California, San Francisco, 2017, <http://ambermd.org/contributors.html> (accessed November 1, 2021).
- [40] V. Tsui, D.A. Case, Theory and applications of the generalized born solvation model in macromolecular simulations, *Biopolymers* 56 (2001) 75–291, [https://doi.org/10.1002/1097-0282\(2000\)56:4<275::AID-BIP10024>3.0.CO;2-E](https://doi.org/10.1002/1097-0282(2000)56:4<275::AID-BIP10024>3.0.CO;2-E).
- [41] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing of a general amber force field, *J. Comput. Chem.* 25 (2004) 1157–1174, <https://doi.org/10.1002/JCC.20035>.
- [42] D. Hamelberg, J.A. McCammon, Standard free energy of releasing a localized water molecule from the binding pockets of proteins: Double-decoupling method, *J. Am. Chem. Soc.* 126 (2004) 7683–7689, <https://doi.org/10.1021/JA0377908>.
- [43] A. Onufriev, D. Bashford, D.A. Case, Modification of the Generalized Born Model Suitable for Macromolecules, *J. Phys. Chem. B.* 104 (2000) 3712–3720, <https://doi.org/10.1021/JP994072S>.
- [44] S. Kumar, J.M. Rosenberg, D. Bouzida, R.H. Swendsen, P.A. Kollman, The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method, *J. Comput. Chem.* 13 (1992) 1011–1021, <https://doi.org/10.1002/JCC.540130812>.
- [45] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, Gaussian 09, Revision A.02 Gaussian, Inc., Wallingford CT, 2016. <https://gaussian.com/g09citation/> (accessed November 1, 2021).
- [46] R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (2018) Available online at <https://www.R-project.org/>. (accessed November 1, 2021).
- [47] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, *J. Mol. Graph.* 14 (1996) 33–38, [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5).
- [48] P. Jantaree, K. Lirdprapamongkol, W. Kaewsri, C. Thongsornkleeb, K. Choowongkorn, K. Atjanasupatt, S. Ruchirawat, J. Svasti, Homodimers of Vanillin and Apocynin Decrease the Metastatic Potential of Human Cancer Cells by Inhibiting the FAK/PI3K/Akt Signaling Pathway, *J. Agric. Food Chem.* 65 (2017) 2299–2306, <https://doi.org/10.1021/acs.jafc.6b05697>.
- [49] D. Lamoral-Théys, L. Pottier, F. Kerff, F. Dufresne, F. Proutière, N. Wauthoz, P. Neven, L. Ingrassia, P. Van Antwerpen, F. Lefranc, M. Gelbcke, B. Pirotte, J.-L. Kraus, J. Nève, A. Kornienko, R. Kiss, J. Dubois, Simple di- and trivanillates exhibit cytostatic properties toward cancer cells resistant to pro-apoptotic stimuli, *Bioorg. Med. Chem.* 18 (2010) 3823–3833, <https://doi.org/10.1016/j.bmc.2010.04.047>.
- [50] S.R. Potje, J.A. Troiano, M.E. Graton, V.F. Ximenes, A.C.M.S. Nakamura, C. Antoniali, Hypotensive and vasorelaxant effect of Diapocynin in normotensive rats, *Free Radic. Biol. Med.* 106 (2017) 148–157, <https://doi.org/10.1016/j.freeradbiomed.2017.02.026>.
- [51] A.R. de Souza, I.A.F. Boza, V.F. Ximenes, M.I. Yoguín, M.-J. Dávila-Rodríguez, N. H. Morgon, I. Caracelli, Elucidation of the induced chirality of dansylglycine by

- its interaction with human serum, *Quim. Nova.* 42 (2019) 135–142, <https://doi.org/10.21577/0100-4042.20170341>.
- [52] C.E. Petersen, C.E. Ha, S. Curry, N.V. Bhagavan, Probing the structure of the warfarin-binding site on human serum albumin using site-directed mutagenesis accessed November 28, 2018 *Proteins* 47 (2002) 116–125. <http://www.ncbi.nlm.nih.gov/pubmed/11933059>.
- [53] G. Sudlow, D.J. Birkett, D.N. Wade, The characterization of two specific drug binding sites on human serum albumin accessed November 29, 2018 *Mol. Pharmacol.* 11 (1975) 824–832. <http://www.ncbi.nlm.nih.gov/pubmed/1207674>.
- [54] K. Yamasaki, V.T.G. Chuang, T. Maruyama, M. Otagiri, Albumin–drug interaction and its clinical implication, *Biochim. Biophys. Acta - Gen. Subj.* 2013 (1830) 5435–5443, <https://doi.org/10.1016/j.bbagen.2013.05.005>.
- [55] A. Nozaki, M. Hori, T. Kimura, H. Ito, T. Hatano, Interaction of Polyphenols with Proteins: Binding of (–)-Epigallocatechin Gallate to Serum Albumin, Estimated by Induced Circular Dichroism, *Chem. Pharm. Bull.* 57 (2009) 224–228, <https://doi.org/10.1248/CPB.57.224>.
- [56] A. Nerusu, K. Chinthapalli, R. Subramanyam, Role of Herborn (K240E) and Milano Slow (D375H) human serum albumin variants towards binding of phenylbutazone and ibuprofen, *Int. J. Biol. Macromol.* 134 (2019) 645–652, <https://doi.org/10.1016/j.ijbiomac.2019.05.075>.
- [57] L. Bertozo, E. Tavares Neto, L. Oliveira, V. Ximenes, Oxidative Alteration of Trp-214 and Lys-199 in Human Serum Albumin Increases Binding Affinity with Phenylbutazone: A Combined Experimental and Computational Investigation, *Int. J. Mol. Sci.* 19 (2018) 2868, <https://doi.org/10.3390/ijms19102868>.
- [58] K. Yamasaki, K. Sakurama, K. Nishi, H. Watanabe, T. Maruyama, H. Seo, M. Otagiri, K. Taguchi, Characterization of the Interaction of Daptomycin With Site II on Human Serum Albumin, *J. Pharm. Sci.* 109 (2020) 2919–2924, <https://doi.org/10.1016/j.xphs.2020.06.011>.
- [59] P. Ascenzi, A. Bocedi, S. Notari, G. Fanali, R. Fesce, M. Fasano, Allosteric Modulation of Drug Binding to Human Serum Albumin, *Mini-Reviews, Med. Chem.* 6 (2006) 483–489, <https://doi.org/10.2174/138955706776361448>.
- [60] L. Leboffe, A. di Masi, F. Polticelli, V. Trezza, P. Ascenzi, Structural basis of drug recognition by human serum albumin, *Curr. Med. Chem.* 26 (2019) 4907–4931, <https://doi.org/10.2174/0929867326666190320105316>.
- [61] K. Yamasaki, S. Hyodo, K. Taguchi, K. Nishi, N. Yamaotsu, S. Hirono, V.T.G. Chuang, H. Seo, T. Maruyama, M. Otagiri, J. Chamani, Long chain fatty acids alter the interactive binding of ligands to the two principal drug binding sites of human serum albumin, *PLoS One* 12 (6) (2017) e0180404, <https://doi.org/10.1371/JOURNAL.PONE.0180404>.
- [62] N.K. Das, N. Ghosh, A.P. Kale, R. Mondal, U. Anand, S. Ghosh, V.K. Tiwari, M. Kapur, S. Mukherjee, Temperature induced morphological transitions from native to unfolded aggregated States of human serum albumin, *J. Phys. Chem. B.* 118 (2014) 7267–7276, <https://doi.org/10.1021/JJP5030944>.

PUBLICATION D7

**MODELING GLYCOSAMINOGLYCAN-PROTEIN
COMPLEXES**



Modeling glycosaminoglycan–protein complexes

Małgorzata M. Kogut^a, Mateusz Marcisz^a and
Sergey A. Samsonov

Abstract

Glycosaminoglycans are long linear and complex polysaccharides that are fundamental components of the mammalian extracellular matrix. Therefore, it is crucial to appropriately characterize molecular structure, dynamics, and interactions of protein-glycosaminoglycans complexes for improving understanding of molecular mechanisms underlying GAG biological function. Nevertheless, this proved challenging experimentally, and theoretical techniques are beneficial to construct new hypotheses and aid the interpretation of experimental data. The scope of this mini-review is to summarize four specific aspects of the current theoretical approaches for investigating noncovalent protein-glycosaminoglycan complexes such as molecular docking, free binding energy calculations, modeling ion impact, and addressing the phenomena of multipose binding of glycosaminoglycans to proteins.

Addresses

Faculty of Chemistry, University of Gdańsk, ul. Wita Stwosza 63, 80-308, Gdańsk, Poland

Corresponding author: Samsonov, Sergey A (sergey.samsonov@ug.edu.pl)

^a Authors contributed equally.

has been conducted to understand protein-GAG interactions, GAGs still represent considerable challenges for both experimental and computational methods [7]. In the past decade, the computational analysis of protein-GAG systems has demonstrated its potential power to complement experimental procedures. Despite the considerable progress in this field, there are persisting serious computational challenges in the treatment of GAGs owing to the determining role of electrostatics, abundant solvent- and ion-mediated interactions, high flexibility, pseudosymmetry, and periodicity of GAGs. The protocols are still missing the concept of GAG's specificity, and there is a lack of precisely developed approaches and available experimental structures for GAG-containing systems. While there are several excellent reviews focusing on computational advances in protein-GAG research offering a more in-depth overview on this broad topic [7–9], our mini-review concentrates on the four selected aspects of modeling noncovalent protein-GAG complexes: molecular docking; free energy calculations; roles of ions; and multipose binding (Figure 2). We primarily focus on the recent achievements in the field and provide the necessary background.

Current Opinion in Structural Biology 2022, 73:102332

This review comes from a themed issue on **Protein-Carbohydrate complexes and glycosylation**

Edited by Sylvie Ricard-Blum and Joseph Zaia

For complete overview of the section, please refer the article collection - [Protein-Carbohydrate complexes and glycosylation](#)

Available online 10 February 2022

<https://doi.org/10.1016/j.sbi.2022.102332>

0959-440X/© 2022 Elsevier Ltd. All rights reserved.

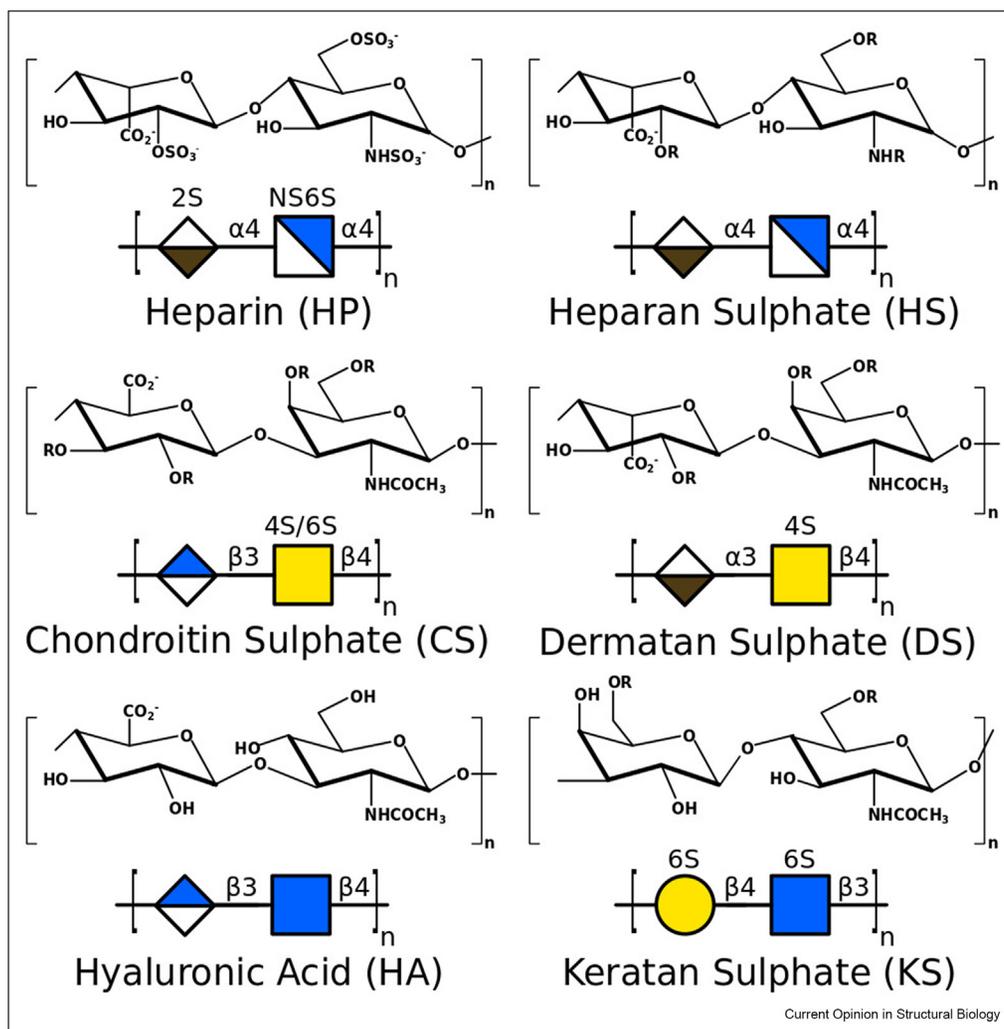
Introduction

Glycosaminoglycans (GAGs) are long anionic linear oligosaccharides made up of disaccharide repeats [1] of diverse composition (Figure 1). They are the key players in numerous biological processes in the extracellular matrix by establishing interactions with protein targets as cytokines, growth factors, and collagen in the tissue-specific context [2]. The disruption of GAG-directed molecular mechanisms results in cancer [3,4], Alzheimer's, Parkinson's diseases [5], and tissue regeneration abnormalities [6]. Although extensive research

Molecular docking of glycosaminoglycans

The computational field of GAGs is still lagging behind the ones of proteins, nucleic acids, and small-drug molecules which represent greater interest for the research community. That is one of the reasons why designing specific docking programs for GAGs is less advanced than for other classes of biologically relevant ligands. GAGs are known to be challenging molecules for docking owing to their length, periodicity, flexibility, linearity, and negative charge. Although some of their interactions with proteins are specific [10], others are electrostatically driven. All those properties make it an arduous task to develop docking approaches containing a sophisticated energy function tailored to GAGs that would be precise and reliable for these highly flexible and charged molecules. There is plenty of conventional docking software originally optimized for other ligands, and many of them have been applied to dock GAGs. However, most of them do not perform at the required quality level [11,12]. Therefore, only some were deployed, and Autodock3 (AD3), which proved to be the most successful [11,13] is the most widely used but

Figure 1



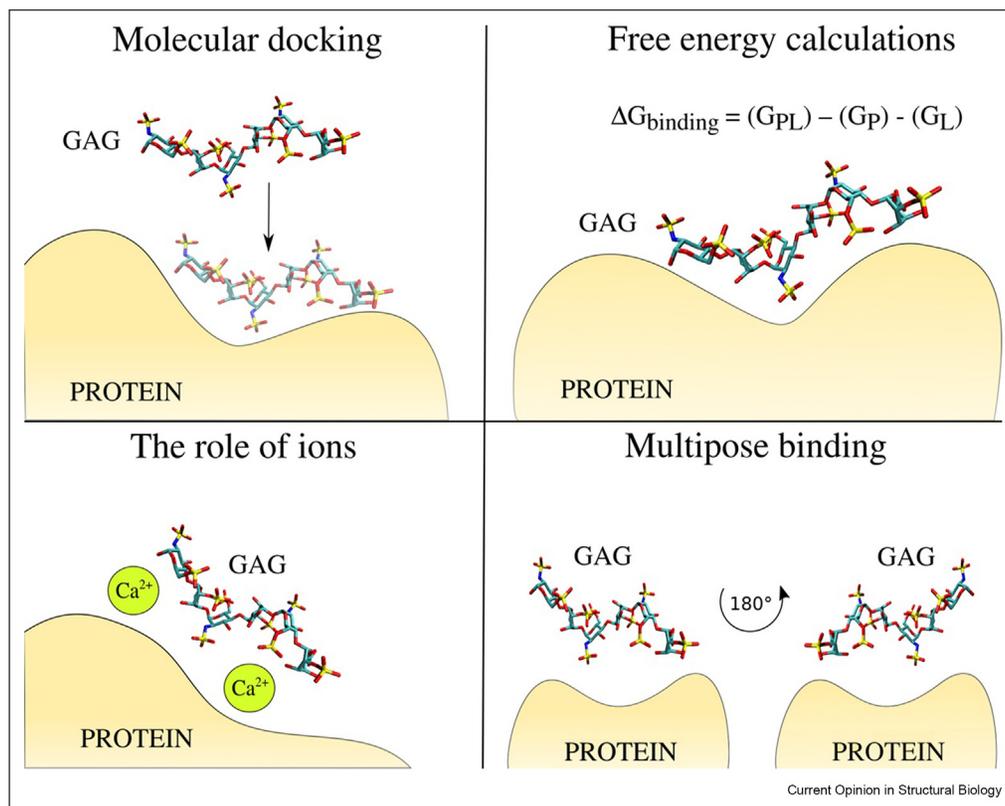
Repeating disaccharide units of glycosaminoglycans with their SNFG representation.

still has flaws. The most fundamental one is the limit of 32 torsional degrees of freedom for the docked molecule that renders longer GAGs (>dp8) unfeasible to be docked flexibly. Therefore, most studies still focus on short GAGs [13].

Among the docking programs dedicated to GAGs, based on the AD program, there are Vina-Carb [14] and its advanced derivative GlycoTorch Vina [15], which both outperformed AD Vina [16] and Glide [17] for a protein-GAG data set [15]. It is also worth mentioning that there are online docking software servers such as ClusPro [18], which in 2014 introduced heparin parameters [19], HADDOCK (High Ambiguity Driven biomolecular DOCKing) [20], or SwissDock [21]. Despite their limitations, some of them proved successful for several systems: ClusPro was used in a recent study on SARS-CoV-2 (severe acute respiratory

syndrome coronavirus 2) and heparin sulfate interactions [22], HADDOCK demonstrated its efficiency in a GAG-related study where CXCL-8 (interleukin 8) interactions with heparin were analyzed [23]. Alternatively, one may manually place GAG near the predicted binding site and follow the molecular dynamics (MD) simulations to find the binding pose [24,25]. To overcome the challenges experienced by conventional docking, principally different approaches were built to dock GAGs. As an example, Dynamic Molecular Docking, which is the combination of molecular docking and MD, was proposed [26]. This steered-MD technique applies the additional potential to move a GAG (ligand) from a distant position toward the binding site on the receptor's surface. The major disadvantage of this method is the required knowledge on a binding site, which is not always available. Moreover, it may use heavy computational resources owing to the required size of

Figure 2



Reviewed aspects of glycosaminoglycans in computational studies. Four aspects of glycosaminoglycans in computational studies discussed in this review: molecular docking (top left); free energy calculations (top right); role of ions (bottom left); and multipose binding (bottom right). Free energy calculations contain the equation describing energy being estimated from the free energies of protein (G_P), ligand (G_L), and complex (G_{PL}).

the periodic boundary box and the use of the explicit solvent model. One more technique to dock GAGs that also works for longer molecules is a fragment-based approach [27]. In this method, the protein's surface is sampled by docking of GAG trimeric fragments, which afterward are assembled in a long chain based on their overlaps. This simple idea allows docking longer GAGs without any limitations associated with their length. However, if the GAG docking site is near the negatively charged amino acid residues, this method could fail to dock trimeric fragments near such residues. As a result, longer docked GAG fragments would not be obtained. A novel approach called repulsive scaling replica exchange molecular dynamics [28] seems to tackle the mentioned problems, and it performs well with GAGs [29]. In repulsive scaling replica exchange molecular dynamics, van der Waals' radii are increased in different replicas (while not affecting other types of interactions in the system). It allows a robust and extensive search for the proper binding sites and poses on the protein surface while leaving the docked molecule and the receptor sidechains flexible (Table 1). To summarize, there is still

room to improve protein-GAG docking tools that should deploy GAG-specific scoring functions.

Free energy calculations of protein-GAG complexes

Successful free energy calculations, which are crucial to understanding protein-GAG systems, similar to molecular docking, face the challenges originated in GAG nature. A binding free energy analysis, molecular mechanics (MM)/Poisson-Boltzmann surface area, applied for the first time to protein-GAG systems by Gandhi et al. [30], together with its approximation MM/generalized Born surface area, is most common for these systems. In both approaches, the evaluation of the free energy is described as a sum of *in vacuo* MM energy terms and a solvation free energy term in implicit solvent calculated for the minimized frames of the MD trajectory in the implicit solvent [31]. Those techniques proved to work in general satisfactorily but not yet precisely enough [32,33]. When applied to protein-GAG systems, these approaches demonstrated to be practical

Table 1

Comparison of four different docking methods and their properties. As an example of conventional docking tool, Autodock—k3 was chosen.

Method	Conventional docking (AD3)	Dynamic Molecular Docking	Fragment-based approach	RS-REMD
Speed	Fast/Average	Average/slow	Average	Average
Usage complexity	Low	Average	High	High
Protein flexibility	No	Full	No	Partial (sidechains)
GAG flexibility	Partial	Full	Full	Full
Solvent model	Implicit	Explicit	Implicit	Implicit
Binding site information	Not required	Required	Not required	Not required
GAG length	< dp8 ^a	Unlimited	Unlimited ^b	Unlimited

AD3, Autodock 3; GAG, glycosaminoglycan; RS-REMD, repulsive scaling replica exchange molecular dynamics.

^a Docking GAGs longer than dp8 makes docking results unreliable.

^b Poses limited to the surface electrostatic potential of the receptor.

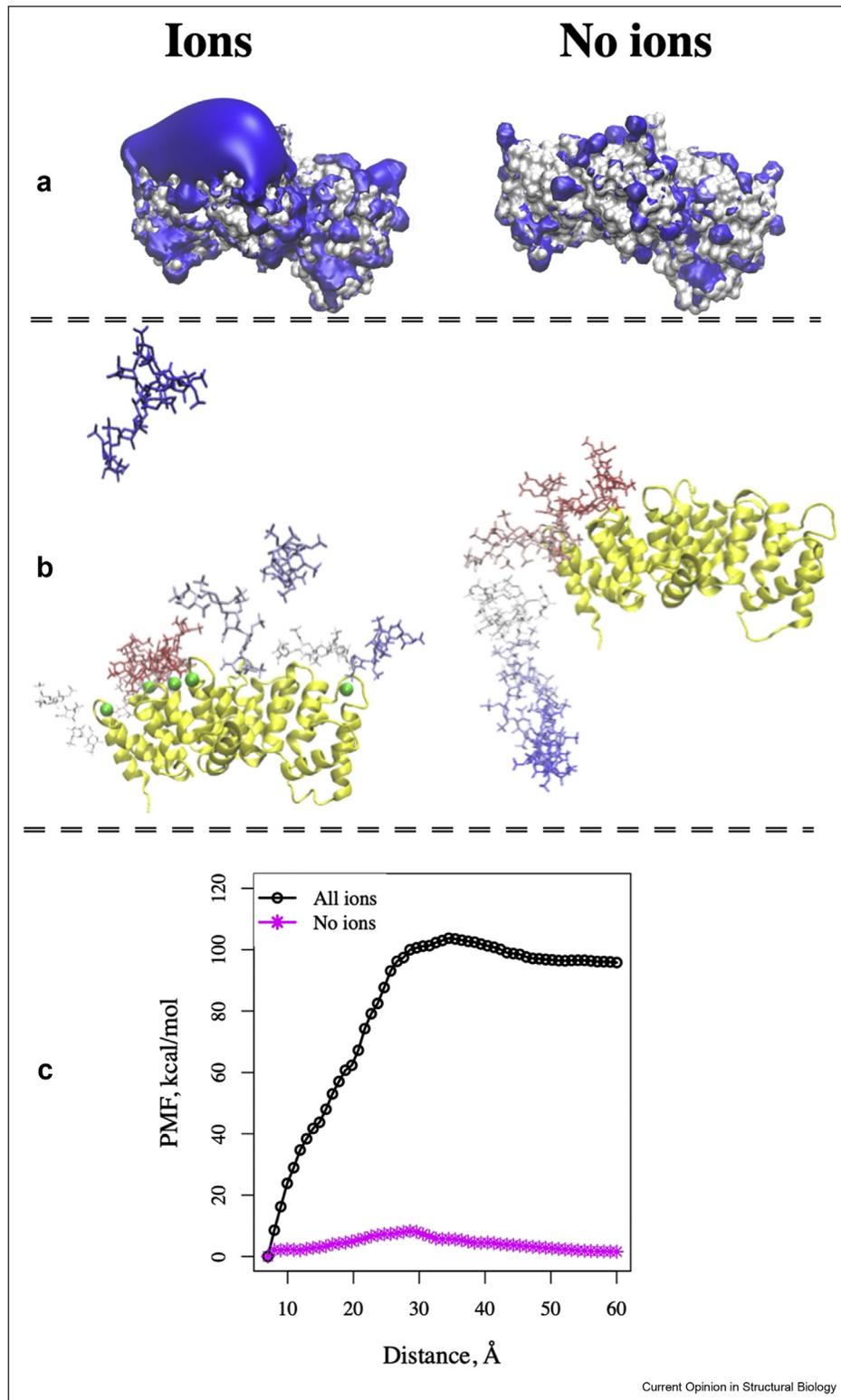
by properly ranking GAG-binding affinities. In addition, they allowed for atomistic interpretation of the experimental data within interdisciplinary studies [34,35]. LIE (linear interaction energy) is another, although less popular, technique, for energy analysis of GAGs, and was applied in several GAG studies [29,36,37]. In LIE, electrostatic energy is calculated *in vacuo* and scaled by a dielectric constant. Although this technique is computationally inexpensive, it is also less accurate. However, if rigorously calibrated with extensive experimental data, LIE can outperform other methods and could be especially promising for scoring protein-GAG interactions. A more accurate but computationally more demanding tool for energy assessment is potential of mean force that can be determined by, for example, umbrella sampling, which can provide binding energy and kinetic characteristics [37]. The potential of mean force can also be obtained by using the Jarzyński equation in steered MD simulations [38]. However, achieving convergence/low error margin is hard to consider this methodology practical for quantitative comparisons. Some techniques suggest replacing the MM force field with QM (quantum mechanics) calculations or combining it into QM/MM methods [32,39]. Nonetheless, there was only limited interest and, hence little research conducted with the use of QM/MM methods in the GAG-related studies.

The role of ions in protein-ion-glycosaminoglycan complexes

Ions play an essential role in physiology and biochemistry, and there are numerous experimental studies describing interactions between GAGs and ions [40,41]. The importance of ions for protein-GAG systems was shown experimentally for APP (amyloid precursor protein) [42], HCII (heparin cofactor II) [43], endostatin [44], FGF1 (fibroblast growth factor 1), and IL-7 (interleukin 7) [45]. Zn²⁺ attenuation of the binding affinity was demonstrated for endostatin-heparin interactions [46]. Recently, several computational studies

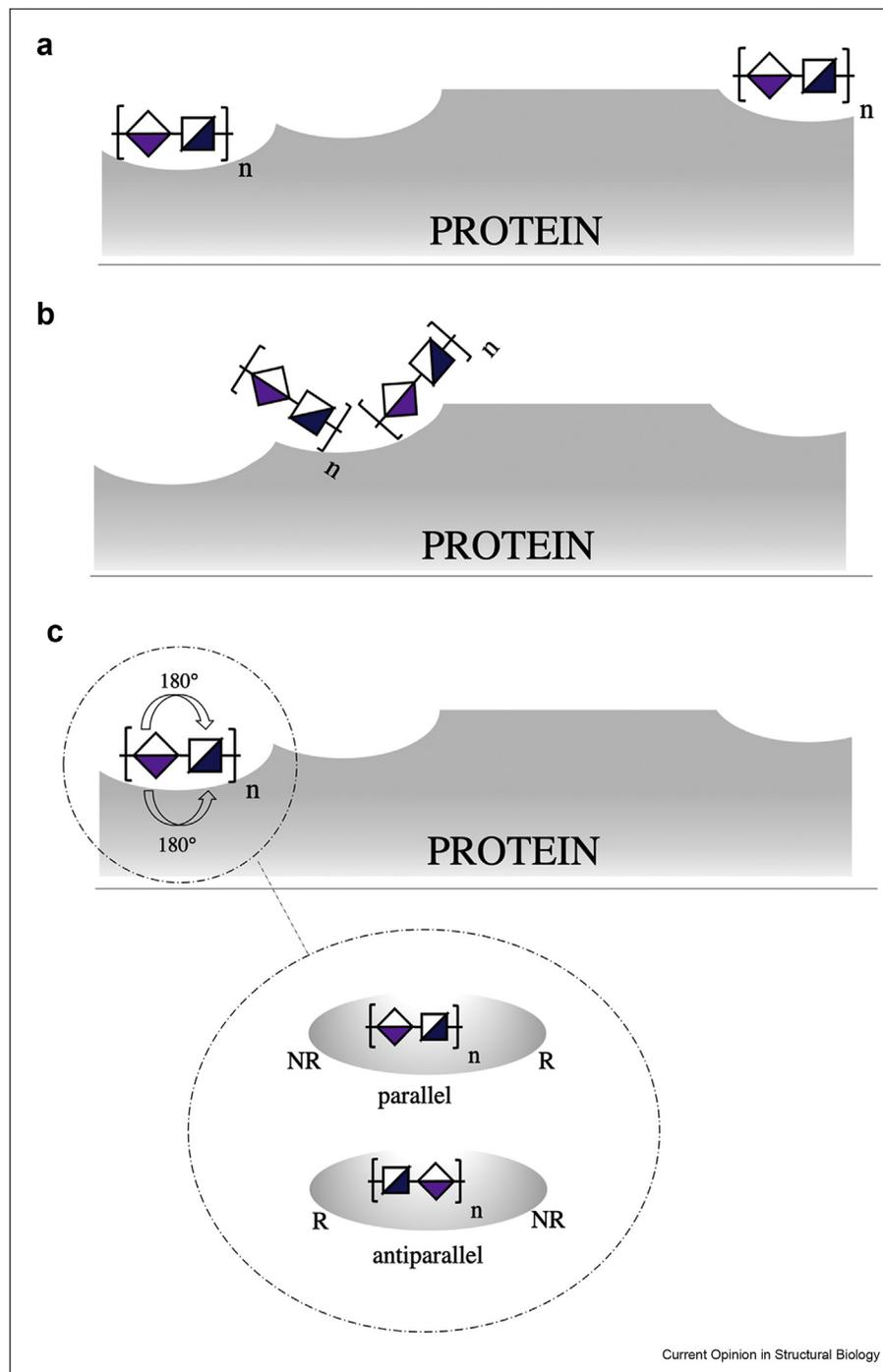
highlighted the importance of ions in the context of molecular modeling of protein-GAG complexes and their free energy analysis to improve the protocols used to treat the ionic environment in GAG-binding studies. The effect of ions on protein-GAG binding using modeling approaches was endeavored by Potthoff et al. [47]. Their findings indicated that calcium ions either bind to GAG before they interact with procollagen C-proteinase enhancer-1 or stabilize the structure and conformation of full-length procollagen C-proteinase enhancer-1. Although several computational approaches were used to predict calcium ion-binding sites on the protein surface, considering calcium ions as part of the protein receptor for docking does not apply to all systems. The latest theoretical findings by Kogut et al. [37] suggested that the presence of calcium ions has a tremendous impact on the annexin-HP binding site. This study aimed to get a deeper insight into protein-ion-GAG interactions using *in silico* techniques to verify the accuracy and sensitivity of the most deployed molecular modeling tools. The results indicated how computational strategy might help to inspect annexin-GAG interactions in the presence and absence of calcium ions at the atomic level. Unlike Potthoff et al. [47] Kogut et al. [37] considered the ions as a part of the receptor using the already reported crystal structures of protein-ion-GAG complexes for the analysis [48,49]. There was clear evidence suggesting that the presence of ions influences the electrostatic potential of the protein surface and renders the dissociation path and its energetic characteristics for a GAG when inspected with the umbrella sampling method (Figure 3). Altogether, these findings contribute to understanding the shortcomings of computational methodologies applicable to protein-GAG systems. The limited research in this field is owing to the lack of available X-ray structures that describe protein-ion-GAG complexes [50]. Another challenge that needs to be faced is the appropriate choice of ion parameters. As it was recently shown by Guvench and Whitmore [51], GAG compactness could

Figure 3



Modeling protein-ion-glycosaminoglycan complexes. **(a)** Electrostatic potential isosurfaces for annexin II (surface representation, in blue +4 kcal/mol/e) in the presence and absence of Ca^{2+} calculated with the Poisson-Boltzmann surface area approach. **(b)** The heparin dissociation pathway for PDB ID 2HYV complex in the presence and absence of Ca^{2+} calculated with the umbrella sampling approach. In dark red: initial position of heparin, and in dark blue: final position of heparin. **(c)** The potential of mean force for the dissociation pathways corresponding to the simulations depicted in panel **(b)** [34].

Figure 4



Multipose binding in protein-glycosaminoglycan complexes. Schematic representation of three possible modes of multiple binding in protein-GAG complexes. The blue squares and purple diamonds represent two different monosaccharide units: **(a)** Multiple binding pockets on the protein surface for one GAG molecule. **(b)** One binding site on the protein surface but different GAG conformations on binding. **(c)** Parallel and antiparallel orientations of GAG binding. The gray ellipsoids represent protein surface with reducing (R) and nonreducing (NR) ends of the oligosaccharide chain in parallel and antiparallel orientations.

be significantly affected by the presence of ions, whereas subtle differences in the applied force field parameters for ions can have dramatic impact on the dynamic and conformational behavior of the GAG polymers. In numerous MM models, the ions are spherical, and their interactions are determined exclusively by Lennard–Jones parameters and the charge. Saxena and Sept [52] introduced a model where the total charge of an ion is distributed into *n*-dummy centers that reproduce the ion's coordination features more appropriately. Similar models have also been proposed for manganese [53], zinc [54], magnesium [55], and nickel [56]. The appropriate design of computational experiments shall not be forgotten to avoid confining the treatment to nonphysiological concentrations of ions [41] as well as the adequate and thoughtful analysis of the obtained results [57].

Multipose binding in protein-glycosaminoglycan complexes

Multipose binding is the property of certain protein-ligand complexes that exhibit different ligand binding modes. The experimentally solved structures for such complexes have been reported [58–60]. Atkovska *et al.* [61] conducted a high-throughput docking study of small molecules and implemented multipose binding in the scoring procedure by considering multiple docking solutions in binding affinity predictions. The take-home message from this work was that careful consideration of multipose binding in docking might give the ability to predict the binding affinity more effectively. Implementation of multipose binding in the scoring scheme yields a better assessment of the binding affinity of the analyzed complex to a different extent depending on the properties of the complex and the selection of the considered poses. The power to distinguish various contributions of each mode to the binding may lead to a more efficient optimization process in rational design, as a proper understanding of how each mode influences the binding. Multipose binding is characteristic for GAGs owing to their pseudosymmetry and periodicity. Figure 4 summarizes three potential modes of GAG multipose binding. Rother *et al.* [62] evaluated interactions of native and chemically sulfated GAG derivatives on the activity of TIMP-3 (tissue inhibitor of metalloproteinase-3). They revealed that differences in their sulfation pattern might be responsible for binding structures that implied GAG's multipose binding. Furthermore, Penk *et al.* [63] investigated the interaction of chemokine (C-X-C motif) ligand 14 (CXCL14) by using *in silico* approaches, NMR spectroscopy, microscale thermophoresis, and analytical heparin affinity chromatography. Results suggested that distinct GAG sulfation patterns confer specificity beyond simple electrostatic interactions that usually represent the driving forces in protein-GAG interactions. They determined *in silico* three binding sites, two of which were energetically more favorable. The most favorable one

agreed best with the data on chemical shift changes obtained by NMR spectroscopy. Although the binding poses for different GAGs are structurally similar when visualizing the trajectories from MD simulations, there was dependence of N-loop and C-terminal alpha-helix residue contribution to GAGs binding on GAG type and charge suggesting specificity of CXCL14-GAG interactions. It is important to note that a GAG can be bound in antiparallel energetically comparable orientations on the protein surface, and these orientations are, thus, difficult to distinguish not only experimentally but also computationally [25]. The very first analysis of the impact of the GAG chain polarity on the interactions with fibroblast growth factors 1 and 2 was carried out by Bojarski and Samsonov [64]. Heparin was predicted to bind to these proteins in the same binding sites but with different orientations, whereas the orientation reported in the experimental structure might be favorable. The probability of the bound GAG orientation change decreases with the increase of heparin chain length. In addition, a GAG can potentially change its orientation by dissociation followed by re-association with the protein rather than rotation in a bound state on the protein surface. This study provides a novel view on the impact of the GAG polarity on the specificity of protein-GAG complex formation — an essential aspect in correctly understanding the intermolecular interactions in these systems.

Conclusions

Although recent advances in computational power and techniques have enabled us to take a step forward in modeling noncovalent protein-GAG complexes, there is still room for improvement. While there is plenty of docking software available, the programs dedicated to GAGs are limited and require careful consideration. There was little interest in further refinement or development of binding free energy calculation approaches specifically for the protein complexes with GAGs as it was for other protein ligands as proteins, nucleic acids, or small-drug molecules. Because improved calculation methods would be of great benefit in the GAG field, the next promising step could be replacing the MM force field with QM calculations or the combination of these two. New theoretical protocols are being developed and tested to study the interactions between proteins, GAGs, and ions. Much progress has also been made on the development of new models for ions. Therefore, it would be of considerable interest to conduct comparative studies for GAG-containing systems where several types of ion parameters are used. Furthermore, the multipose binding should not be disregarded in the GAG rational design because this will help to improve binding sites and pose prediction when experimental data are unavailable. Altogether, protein-GAG system modeling, although still being in the early phase of its development, contributes substantially to overcoming the significant challenges in this research field.

Funding

This research was funded by the National Science Centre of Poland (UMO-2018/30/E/ST4/00037).

Conflict of interest statement

Nothing declared.

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest

** of outstanding interest

- Esko JD, Kimata K, Lindahl U: *Proteoglycans and sulfated glycosaminoglycans. Essentials of glycobiology*. 3rd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2015–2017. Chapter 17.
 - Vallet SD, Clerc O, Ricard-Blum S: **Glycosaminoglycan–protein interactions: the first draft of the glycosaminoglycan interactome**. *J Histochem Cytochem* 2021, **69**:93–104.
 - Wei J, Hu M, Huang K, Lin S, Du H: **Roles of proteoglycans and glycosaminoglycans in cancer development and progression**. *Int J Mol Sci* 2020, **21**:5983.
 - Morla S: **Glycosaminoglycans and glycosaminoglycan mimetics in cancer and inflammation**. *Int J Mol Sci* 2019, **20**:1963.
 - Jin W, Zhang F, Linhardt RJ: **Glycosaminoglycans in neurodegenerative diseases**. *Adv Exp Med Biol* 2021, **1325**:189–204.
 - Paganini C, Costantini R, Superti-Furga A, Rossi A: **Bone and connective tissue disorders caused by defects in glycosaminoglycan biosynthesis: a panoramic view**. *FEBS J* 2019, **286**:3008–3032.
 - Almond A: **Multiscale modeling of glycosaminoglycan structure and dynamics: current methods and challenges**. *Curr Opin Struct Biol* 2018, **50**:58–64.
 - Sankaranarayanan NV, Nagarajan B, Desai UR: **So you think computational approaches to understanding glycosaminoglycan-protein interactions are too dry and too rigid? Think again!** *Curr Opin Struct Biol* 2018, **50**:91–100.
 - Paiardi G, Milanese M, Wade RC, d'ursi P, Rusnati M: **A bittersweet computational journey among glycosaminoglycans**. *Biomolecules* 2021, **11**:739.
 - Petitou M, Casu B, Lindahl U: **197601983, a critical period in the history of heparin: the discovery of the antithrombin binding site**. *Biochimie* 2003, **85**:83–89.
 - Uciechowska-Kaczmarzyk U, de Beauchene I, Samsonov SA: **Docking software performance in protein-glycosaminoglycan systems**. *J Mol Graph Model* 2019, **90**:42–50.
- A molecular docking benchmarking study was performed for a non-redundant dataset of 28 protein-glycosaminoglycan structures. 14 docking programs were compared in terms of their performance for these systems.
- Griffith AR, Rogers CJ, Miller GM, Abrol R, Hsieh-Wilson LC, Goddard WA: **Predicting glycosaminoglycan surface protein interactions and implications for studying axonal growth**. *Proc Natl Acad Sci USA* 2017, **114**:13697–13702.
 - Frank M: **Computational docking as a tool for the rational design of carbohydrate-based drugs**. *Carbohydrat Drugs* 2014, **12**:53–72.
 - Nivedha AK, Thieker DF, Makeneni S, Hu H, Woods RJ: **Vina-carb: improving glycosidic angles during carbohydrate docking**. *J Chem Theor Comput* 2016, **12**:892–901.
 - Boittier ED, Burns JM, Gandhi NS, Ferro V: **GlycoTorch Vina: docking designed and tested for glycosaminoglycans**. *J Chem Inf Model* 2020, **60**:6328–6343.
- GlycoTorch Vina, a novel GAG-specific molecular docking software, which takes into account ring puckering flexibility and GAG glycosidic linkage structural features, was developed based on the carbohydrate-specific molecular docking program VinaCarb. The program outperformed Autodock Vina and Glide for a curated protein-GAG complex benchmark.
- Eberhardt J, Santos-Martins D, Tillack AF, Forli S: **AutoDock Vina 1.2.0: new docking methods, expanded force field, and Python bindings**. *J Chem Inf Model* 2021, **61**:3891–3898.
 - Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelly M, Perry JK, Shaw DE, Francis P, Shenkin PS: **Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy**. *J Med Chem* 2004, **47**:1739–1749.
 - Desta IT, Porter KA, Xia B, Kozakov D, Vajda S: **Performance and its limits in rigid body protein-protein docking**. *Structure* 2020, **28**:1071–1081.
 - Mottarella SE, Beglov D, Beglova N, Nugent MA, Kozakov D, Vajda S: **Docking server for the identification of heparin binding sites on proteins**. *J Chem Inf Model* 2014, **54**:2068–2078.
 - Van Zundert GCP, Rodrigues JPGLM, Trellet M, Schmitz C, Kastiris PL, Karaca E, Melquionhorougvan Dijk M, deVries SJ, Bonvin AMJJ: **The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes**. *J Mol Biol* 2016, **428**:720–725.
 - Grosdidier A, Zoete V, Michielin O: **SwissDock, a protein-small molecule docking web service based on EADock DSS**. *Nucleic Acids Res* 2011, **39**. Web Server issue.
 - Clausen TM, Sandoval DR, Spliid CB, Pihl J, Perrett HR, Painter CD, Narayanan A, Majowicz SA, Kwong EM, McVicar RN, Thacker BE, Glass CA, Yang Z, Torres JL, Golden GJ, Bartles PL, Porell RN, garretson AF, Laubach L, Feldman J, Yin X, Pu Y, Hauser BM, Cardonna TM, Kellman BP, Martino C, Gordts PLSM, Chanda SK, Schmidt AG, Godula K, Leibel SL, Jose J, Corbett KD, Ward AB, Carlin AF, Esko JD: **SARS-CoV-2 infection depends on cellular heparan sulfate and ACE2**. *Cell* 2020, **183**:1043–1057.
- For the first time, an atomistic model of interactions between SARS-CoV-2 spike protein and heparan sulfate was proposed. In this multifaceted interdisciplinary study, the authors showed that heparan sulfate could act as a scaffold for establishing ACE2 and spike protein tertiary complex.
- Joseph PRB, Mosier PD, Desai UR, Rajarathnam K: **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.
 - Sandoval Daniel R, Toledo Alejandro Gomez, Painter Chelsea D, Tota Ember M, Osman Sheikh XM, West Alan MV, Frank Martin M, Wells Lance, Xu Ding, Bicknell Roy, Corbett Xde Kevin D: **Proteomics-based screening of the endothelial heparansulfate interactome reveals that C-type lectin 14a (CLEC14A) is a heparin-binding protein**. *J Biol Chem* 2020, **295**:2804–2821.
 - Vuorio J, Vattulainen I, Martinez-Seara H: **Atomistic fingerprint of hyaluronan-CD44 binding**. *PLoS Comput Biol* 2017, **13**.
 - Samsonov SA, Gehrcke JP, Pisabarro MT: **Flexibility and explicit solvent in molecular-dynamics-based docking of protein-glycosaminoglycan systems**. *J Chem Inf Model* 2014, **54**:582–592.
 - Samsonov SA, Zacharias M, Beauchene IC: **Modeling large protein-glycosaminoglycan complexes using a fragment-based approach**. *J Comput Chem* 2019, **40**:1429–1439.
 - Siebenmorgen T, Engelhard M, Zacharias M: **Prediction of protein-protein complexes using replica exchange with repulsive scaling**. *J Comput Chem* 2020, **41**:1436–1447.
- A novel repulsive scaling replica exchange molecular dynamics simulations protocol, where Lennard-Jones parameters were adjusted in different replicas, was applied to a set of protein-protein complexes. This technique allowed effective protein-protein docking thanks to the improved search of the conformational space.
- Maszota-Zieleniak M, Marcisz M, Kogut MM, Siebenmorgen T, Zacharias M, Samsonov SA: **Evaluation of replica exchange with repulsive scaling approach for docking glycosaminoglycans**. *J Comput Chem* 2021, **42**:1040–1053.

30. Gandhi NS, Mancera RL: **Free energy calculations of glycosaminoglycan–protein interactions.** *Glycobiology* 2009, **19**:1103–1115.
31. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA: **Cheatham 3rd TE: calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models.** *Acc Chem Res* 2000, **33**:889–897.
32. Genheden S, Ryde U: **The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities.** *Expert Opin Drug Discov* 2015, **10**:449–461.
33. Huai Z, Yang H, Li X, Sun Z: **SAMPL7 TrimerTrip host–guest binding affinities from extensive alchemical and end-point free energy calculations.** *J Comput Mol Des* 2020, **35**:117–129.
34. Speciale I, Duncan GA, Unione L, Agarkova IV, Garozzo D, Jimenez-Barbero J, Lin S, Lowary TL, Molinaro A, Noel E, Laugieri ME, Tonetti MG, Van Etten JL, De Castro C: **The N-glycan structures of the antigenic variants of chlorovirus PBCV-1 major capsid protein help to identify the virus-encoded glycosyltransferases.** *J Biol Chem* 2019 Apr 5, **294**:5688–5699, <https://doi.org/10.1074/jbc.RA118.007182>. Epub 2019 Feb 8. PMID: 30737276; PMCID: PMC6462530.
35. Chazeirat T, Denamur S, Bojarski KK, Andrault PM, Sizaret D, Zhang F, Saidi A, Tardieu M, Linhardt RJ, Labarthe F, Brömme D, Samsonov SA, Lalmanach G, Lecaille F: **The abnormal accumulation of heparan sulfate in patients with mucopolysaccharidosis prevents the elastolytic activity of cathepsin V.** *Carbohydr Polym* 2021 Feb 1, **253**:117261, <https://doi.org/10.1016/j.carbpol.2020.117261>. Epub 2020 Oct 20. PMID: 33278943.
36. Marcisz M, Huard B, Lipska AG, Samsonov SA: **Further analyses of APRIL/APRIL-receptor/glycosaminoglycan interactions by biochemical assays linked to computational studies.** *Glycobiology* 2021, **31**:772–786.
- In this study, which combined *in silico* and *in vitro* approaches, the interactions between APRIL protein and a set of short GAG oligosaccharides were characterized. Methodological aspects of molecular dynamics simulations and free energy calculations using MM-GBSA and LIE protocols were rigorously analyzed and discussed.
37. Kogut MM, Maszota-Zieleniak M, Marcisz M, Samsonov SA: **Computational insights into the role of calcium ions in protein-glycosaminoglycan systems.** *Phys Chem Chem Phys* 2021, **23**:3519–3530.
- A combination of molecular docking, conventional and conformational sampling enhanced molecular dynamics and free energy calculation approaches were applied to model and analyze the interactions in annexin-calcium ions-heparin systems in detail. The computational data were thoroughly examined and discussed in the context of the structural data previously reported by the crystallographic studies.
38. Bojarski KK, Becher J, Riemer T, Lemmritzer K, Moller S, Schiller J, Schnabelrauch M, Samsonov SA: **Synthesis and in silico characterization of artificially phosphorylated glycosaminoglycans.** *J Mol Struct* 2019, **1197**:401–416.
39. Jafari S, Ryde U, Irani M: **Two-substrate glyoxalase I mechanism: a quantum mechanics/molecular mechanics study.** *Inorg Chem* 2020, **60**:303–314.
40. Woodhead NE, Long WF, Williamson FB: **Binding of zinc ions to heparin. Analysis by equilibrium dialysis suggests the occurrence of two, entropy-driven, processes.** *Biochem J* 1986, **237**:281–284.
41. Stevic I, Parmar N, Paredes N, Berry LR, Chan AKC: **Binding of heparin to metals.** *Cell Biochem Biophys* 2011, **59**:171–178.
42. Multhaupt G, Bush AI, Pollwein P, Masters CL: **Interaction between the zinc (II) and the heparin binding site of the Alzheimer's disease beta A4 amyloid precursor protein (APP).** *FEBS Lett* 1994, **335**:151–154.
43. Eckert R, Raag H: **Zinc ions promote the interaction between heparin and heparin cofactor II.** *FEBS Lett* 2003, **541**:121–125.
44. Han Q, Fu Y, Zhou H, He Y, Lou Y: **Contributions of Zn (II)-binding to the structural stability of endostatin.** *FEBS Lett* 2007, **581**:3027–3032.
45. Zhang F, Liang X, Beaudet JM, Lee Y, Linhardt RJ: **The effects of metal ions on heparin/heparin sulfate-protein interactions.** *J Bio Technol Res* 2014, **1**:1–7.
46. Uciechowska-Kaczmarzyk U, Babik S, Zsila F, Bojarski KK, Beke-Somfai T, Samsonov SA: **Molecular dynamics-based model of VEGF-A and its heparin interactions.** *J Mol Graph Model* 2018, **82**:157–166.
47. Potthoff J, Bojarski KK, Kohut G, Lipska AG, Liwo A, Kessler E, Ricard-Blum S, Samsonov SA: **Analysis of procollagen C-Proteinase enhancer-1/glycosaminoglycan binding sites and of the potential role of calcium ions in the interaction.** *Int J Mol Sci* 2019, **20**:5021.
- The effect of GAG length, type and sulfation pattern as well as the role of calcium ions were analyzed for C-Proteinase enhancer-1/GAG interactions system. The study represents one of the first attempts to consider ions in the protein-GAG systems explicitly. The computational data were used for the interpretation of the experimental binding assays.
48. Shao C, Zhang F, Kemp MM, Linhardt RJ, Waisman DM, Head JF, Seaton BA: **Crystallographic analysis of calcium-dependent heparin binding to annexin A2.** *J Biol Chem* 2006, **281**:31689–31695.
49. Capila I, Hernaiz MJ, Mo YD, Mealy TR, Campos B, Deadman JR, R. Linhardt RJ, Seaton BA: **Annexin V–heparin oligosaccharide complex suggests heparan sulfate–mediated assembly on cell surfaces.** *Structure* 2001, **9**:57–64.
50. Perez S, Bonnardelet F, Lisacek F, Imbert A, Ricard Blum S, Makashova O: **GAG-DB, the new interface of the three-dimensional landscape of glycosaminoglycans.** *Biomolecules* 2020, **10**:1660.
- An open-source curated database of protein-GAG structures was build using the data obtained from X-ray single-crystal diffraction, X-ray fiber diffractometry, solution NMR spectroscopy, and scattering data often associated with molecular modeling approaches.
51. Guvench O, Whitmore EK: **Sulfation and calcium favor compact conformations of chondroitin in aqueous solutions.** *ACS Omega* 2021, **6**:13204–13217.
- All-atomic simulations of the two types of chondroitin sulfate in the presence and absence of calcium ions have been performed. The effects of the GAG sulfation and ions on structural/dynamic properties of the analyzed GAGs have been rigorously discussed.
5. Saxena A, Sept D: **Multisite ion models that improve coordination and free energy calculations in molecular dynamics simulations.** *J Chem Theor Comput* 2013, **9**:3538–3542.
53. Aqvist J, Warshel A: **Free energy relationships in metalloenzyme-catalyzed reactions. Calculations of the effects of metal ion substitutions in staphyloboccal nuclease.** *J Am Chem Soc* 1990, **112**:2860–2868.
54. Aqvist J, Warshel AJ: **Computer simulation of the initial proton transfer step in human carbonic anhydrase.** *J Mol Biol* 1992, **224**:7–14.
55. Oelschlaeger P, Klahn M, Beard WA, Wilson SH, Warshel AJ: **Magnesium-cationic dummy atom molecules enhance representation of DNA polymerase β in molecular dynamics simulations: improved accuracy in studies of structural features and mutational effects.** *Mol Biol* 2007, **366**:687.
56. Masetti M, Musiani F, Bernetti M, Falchi F, Cavalli A, Ciurli S, Recanatini M: **Development of a multisite model for Ni(II) ion in solution from thermodynamic and kinetic data.** *J Comput Chem* 2017, **38**:1834–1843.
57. Lemke T, Edte M, Gebauer D, Peter C: **Three reasons why aspartic acid and glutamic acid sequences have a surprisingly different influence on mineralization.** *J Phys Chem B* 2021, **125**:10335–10343.
58. Kulp JL, Blumenthal SN, Wang Q, Bryan RL, Guarnieri F: **A fragment-based approach to the SAMPL3 challenge.** *J Comput Aided Mol Des* 2012, **26**:83–594.
59. Blum A, Bottcher J, Dorr S, Heine A, Klebe G, Diederich WE: **Two solutions for the same problem: multiple binding modes of pyrrolidine-based HIV-1 protease inhibitors.** *J Mol Biol* 2011, **410**:745–755.

10 Protein-Carbohydrate complexes and glycosylation

60. Gushchina LV, Gabdulkhakov AG, Nikonov SV, Filimonov VV: **High-resolution crystal structure of spectrin SH3 domain fused with a proline-rich peptide.** *J Biomol Struct Dyn* 2011, **29**:485–495.
61. Atkovska K, Samsonov SA, Paszkowski-Rogacz M, Pisabarro MT: **Multipose binding in molecular docking.** *Int J Mol Sci* 2014, **15**:2622–2645.
62. Rother S, Samsonov SA, Hofmann T, Blaszkiewicz J, Köhling S, Schnabelrauch M, Möller S, Rademann J, Kalkhof S, von Bergen M: **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.
63. Penk A, Baumann L, Huster D, Samsonov SA: **NMR and molecular modeling reveal specificity of the interactions between CXCL14 and glycosaminoglycans.** *Glycobiology* 2019, **29**:715–725.
*
Complementary NMR and modeling study of CXCL14 protein interactions with a set of oligomeric GAGs was performed. The CXCL14 epitope for heparin suggests a binding pose distinguishable from the other investigated GAGs investigated. Additionally, this observation of the interaction specificity was supported by computational methods that included molecular docking, molecular dynamics and free energy calculations.
64. Bojarski KK, Samsonov SA: **Role of oligosaccharide chain polarity in protein-glycosaminoglycan interactions.** *J Chem Inf Model* 2020, **61**:455–466.

**STATEMENTS CONFIRMING THE AUTHORSHIP OF
PUBLICATIONS INCLUDED IN THIS PHD THESIS**

M.Chem. Małgorzata M. Kogut-Günthel
University of Gdańsk, Faculty of Chemistry
Department of Theoretical Chemistry
Wita Stwosza 63, Gdańsk, Poland
malgorzata.kogut@phdstud.ug.edu.pl

Gdańsk, 20.11.2022

STATEMENT

I declare that I am a co-author of the following publications included in the PhD thesis:

Computational insights into the calcium ions role in protein-glycosaminoglycan systems.

Małgorzata M. Kogut, Martyna Maszota-Zieleniak, Mateusz Marcisz, Sergey A. Samsonov.

Published in: Physical Chemistry Chemical Physics, 2021, 23: 3519–3530.

I contributed to the preparation of this manuscript by designing and conducting the research, analysing the data, writing, and revising the manuscript.

Percentage contribution: 80%

Impact of calcium ions on the structural and dynamic properties of heparin oligosaccharides by computational analysis

Małgorzata M. Kogut, Annemarie Danielsson, Sylvie Ricard-Blum, Sergey A. Samsonov.

Published in Computational Biology and Chemistry, 2022, 99:107727.

I contributed to the preparation of this manuscript by conducting the research, analysing the data, and writing and revising the manuscript.

Percentage contribution: 80%

Affinity and putative entrance mechanisms of alkyl sulfates into the β -CD cavity

Małgorzata M. Kogut, Ola Grabowska, Dariusz Wyrzykowski, Sergey A. Samsonov.

Published in Journal of Molecular Liquids, 2022, 364:119978.

I contributed to the preparation of this manuscript by conducting computational research, analysing the data, and writing and revising the manuscript.

Percentage contribution: 50%


.....
(signature)

Physicochemical nature of sodium dodecyl sulfate interactions with bovine serum albumin revealed by interdisciplinary approaches.

Aleksandra Tesmar, Małgorzata M. Kogut, Krzysztof Żamojć, Ola Grabowska, Katarzyna Chmur, Sergey A. Samsonov, Joanna Makowska, Dariusz Wyrzykowski, Lech Chmurzyński.

Published in: Journal of Molecular Liquids, 2021, 340:117185.

I contributed to the preparation of this manuscript by conducting computational research, analysing the data, and writing and revising the manuscript.

Percentage contribution: 30%

Effect of Tetrphenylborate on Physicochemical Properties of Bovine Serum Albumin.

Ola Grabowska, Małgorzata M. Kogut, Krzysztof Żamojć, Sergey A. Samsonov, Joanna Makowska, Aleksandra Tesmar, Katarzyna Chmur, Dariusz Wyrzykowski, Lech Chmurzyński.

Published in: Molecules, 2021, 26(21):6565.

I contributed to the preparation of this manuscript by conducting computational research, analysing the data, and writing and revising the manuscript.

Percentage contribution: 20%

Induced Circular Dichroism as a Tool to Monitor the Displacement of Ligands Between Albumins.

Luiza de Carvalho Bertozo, Małgorzata M. Kogut, Martyna Maszota-Zieleniak, Sergey A. Samsonov, Valdecir F. Ximenes.

Published in: Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2022, 278:121374.

I contributed to the preparation of this manuscript by conducting computational research, analysing the data, and writing and revising the manuscript.

Percentage contribution: 40%


(signature)

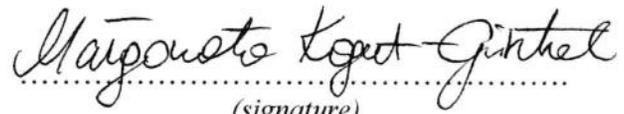
Modeling glycosaminoglycan-protein complexes.

Małgorzata M. Kogut, Mateusz Marcisz, Sergey A. Samsonov.

Published in: Current Opinion in Structural Biology, 2022, 73:102332.

I contributed to the preparation of this manuscript by researching the literature, writing, and revising the manuscript.

Percentage contribution: 40%


(signature)

Dr hab. Sergey A. Samsonov, prof. UG
University of Gdańsk, Faculty of Chemistry
Department of Theoretical Chemistry
Wita Stwosza 63, Gdańsk, Poland
sergey.samsonov@ug.edu.pl

Gdańsk, 20.11.2022

STATEMENT

I declare that I am a co-author of the following publications included in the PhD thesis:

Computational insights into the calcium ions role in protein-glycosaminoglycan systems.

Małgorzata M. Kogut, Martyna Maszota-Zieleniak, Mateusz Marcisz, Sergey A. Samsonov.

Published in: Physical Chemistry Chemical Physics, 2021, 23: 3519–3530.

I contributed to the preparation of this manuscript by designing and supervising the research, data analysis, and writing and revising the manuscript.

Impact of calcium ions on the structural and dynamic properties of heparin oligosaccharides by computational analysis

Małgorzata M. Kogut, Annemarie Danielsson, Sylvie Ricard-Blum, Sergey A. Samsonov.

Published in Computational Biology and Chemistry, 2022, 99:107727.

I contributed to the preparation of this manuscript by designing and supervising the research, data analysis, and writing and revising the manuscript.

Modeling glycosaminoglycan-protein complexes.

Małgorzata M. Kogut, Mateusz Marcisz, Sergey A. Samsonov.

Published in: Current Opinion in Structural Biology, 2022, 73:102332.

I contributed to the preparation of this manuscript by researching the literature, writing, and revising the manuscript.

.....
Sergey Samsonov 
(signature)

Affinity and putative entrance mechanisms of alkyl sulfates into the β -CD cavity

Małgorzata M. Kogut, Ola Grabowska, Dariusz Wyrzykowski, Sergey A. Samsonov.

Published in Journal of Molecular Liquids, 2022, 364:119978.

I contributed to the preparation of this manuscript by designing and supervising the theoretical part of the research, data analysis, and writing and revising the manuscript.

Physicochemical nature of sodium dodecyl sulfate interactions with bovine serum albumin revealed by interdisciplinary approaches.

Aleksandra Tesmar, Małgorzata M. Kogut, Krzysztof Żamojć, Ola Grabowska, Katarzyna Chmur, Sergey A. Samsonov, Joanna Makowska, Dariusz Wyrzykowski, Lech Chmurzyński.

Published in: Journal of Molecular Liquids, 2021, 340:117185.

I contributed to the preparation of this manuscript by designing and supervising the theoretical part of the research, data analysis, and writing and revising the manuscript.

Effect of Tetraphenylborate on Physicochemical Properties of Bovine Serum Albumin.

Ola Grabowska, Małgorzata M. Kogut, Krzysztof Żamojć, Sergey A. Samsonov, Joanna Makowska, Aleksandra Tesmar, Katarzyna Chmur, Dariusz Wyrzykowski, Lech Chmurzyński.

Published in: Molecules, 2021, 26(21):6565.

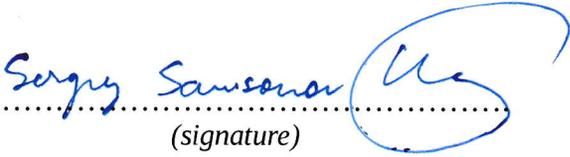
I contributed to the preparation of this manuscript by designing and supervising the theoretical part of the research, data analysis, and writing and revising the manuscript.

Induced Circular Dichroism as a Tool to Monitor the Displacement of Ligands Between Albumins.

Luiza de Carvalho Bertozo, Małgorzata M. Kogut, Martyna Maszota-Zieleniak, Sergey A. Samsonov, Valdecir F. Ximenes.

Published in: Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2022, 278:121374.

I contributed to the preparation of this manuscript by designing and supervising the theoretical part of the research, data analysis, and writing and revising the manuscript.


.....
(signature)

dr Martyna Maszota-Zieleniak
University of Gdańsk, Faculty of Chemistry
Department of Theoretical Chemistry
Wita Stwosza 63, Gdańsk, Poland
m.maszota.zieleniak@gmail.com

Gdańsk, 01.12.2022

STATEMENT

I declare that I am a co-author of the following publications included in the PhD thesis:

Computational insights into the calcium ions role in protein-glycosaminoglycan systems.

Małgorzata M. Kogut, Martyna Maszota-Zieleniak, Mateusz Marcisz, Sergey A. Samsonov.

Published in: Physical Chemistry Chemical Physics, 2021, 23: 3519–3530.

I contributed to the preparation of this manuscript by providing data, writing, and revising the manuscript.

Induced Circular Dichroism as a Tool to Monitor the Displacement of Ligands Between Albumins.

Luiza de Carvalho Bertozo, Małgorzata M. Kogut, Martyna Maszota-Zieleniak, Sergey A. Samsonov, Valdecir F. Ximenes.

Published in: Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2022, 278:121374.

I contributed to the preparation of this manuscript through methodology and investigation.


.....
(signature)

mgr Mateusz Marcisz
University of Gdańsk, Faculty of Chemistry
Department of Theoretical Chemistry
Wita Stwosza 63, Gdańsk, Poland
mateusz.marcisz@gmail.com

Gdańsk, 01.12.2022

STATEMENT

I declare that I am a co-author of the following publications included in the PhD thesis:

Computational insights into the calcium ions role in protein-glycosaminoglycan systems.

Małgorzata M. Kogut, Martyna Maszota-Zieleniak, Mateusz Marcisz, Sergey A. Samsonov.

Published in: Physical Chemistry Chemical Physics, 2021, 23: 3519–3530.

I contributed to the preparation of this manuscript by providing data, writing, and revising the manuscript.

Modeling glycosaminoglycan-protein complexes.

Małgorzata M. Kogut, Mateusz Marcisz, Sergey A. Samsonov.

Published in: Current Opinion in Structural Biology, 2022, 73:102332.

I contributed to the preparation of this manuscript by researching the literature, writing, and revising the manuscript.



.....
(signature)

mgr Annemarie Danielsson
University of Gdańsk, Faculty of Chemistry
Department of Theoretical Chemistry
Wita Stwosza 63, Gdańsk, Poland
annemarie.danielsson@gmail.com

Gdańsk, 01.12.2022

STATEMENT

I declare that I am a co-author of the following publications included in the PhD thesis:

Impact of calcium ions on the structural and dynamic properties of heparin oligosaccharides by computational analysis.

Małgorzata M. Kogut, Annemarie Danielsson, Sylvie Ricard-Blum, Sergey A. Samsonov.

Published in Computational Biology and Chemistry, 2022, 99:107727.

I contributed to the preparation of this manuscript through formal analysis, investigation, methodology, and visualisation.


.....
(signature)

TO WHOM IT MAY CONCERN

STATEMENT

I declare that I am a co-author of the following publications included in Małgorzata Kogut 's PhD thesis **Impact of calcium ions on the structural and dynamic properties of heparin oligosaccharides by computational analysis** authored by Małgorzata M. Kogut, Annemarie Danielsson, **Sylvie Ricard-Blum**, and Sergey A. Samsonov and published in Computational Biology and Chemistry 2022, 99: 107727.

I contributed to the preparation of this article through conceptualization, writing the original draft, writing review & editing.

Villeurbanne, December 1st, 2022

Professor Sylvie RICARD-BLUM



dr hab. Dariusz Wyrzykowski, prof. UG
University of Gdańsk, Faculty of Chemistry
Department of General and Inorganic Chemistry
Wita Stwosza 63, Gdańsk, Poland
dariusz.wyrzykowski@ug.edu.pl

Gdańsk, 01.12.2022

STATEMENT

I declare that I am a co-author of the following publications included in the PhD thesis:

Affinity and putative entrance mechanisms of alkyl sulfates into the β -CD cavity

Małgorzata M. Kogut, Ola Grabowska, Dariusz Wyrzykowski, Sergey A. Samsonov.

Published in Journal of Molecular Liquids, 2022, 364:119978.

I contributed to the preparation of this manuscript through conceptualisation, investigation, methodology, project administration, supervision, writing – original draft, writing – review & editing.

Physicochemical nature of sodium dodecyl sulfate interactions with bovine serum albumin revealed by interdisciplinary approaches.

Aleksandra Tesmar, Małgorzata M. Kogut, Krzysztof Żamojć, Ola Grabowska, Katarzyna Chmur, Sergey A. Samsonov, Joanna Makowska, Dariusz Wyrzykowski, Lech Chmurzyński.

Published in: Journal of Molecular Liquids, 2021, 340:117185.

I contributed to the preparation of this manuscript through conceptualisation, project administration, writing - review & editing.

Effect of Tetraphenylborate on Physicochemical Properties of Bovine Serum Albumin.

Ola Grabowska, Małgorzata M. Kogut, Krzysztof Żamojć, Sergey A. Samsonov, Joanna Makowska, Aleksandra Tesmar, Katarzyna Chmur, Dariusz Wyrzykowski, Lech Chmurzyński.

Published in: Molecules, 2021, 26(21):6565.

I contributed to the preparation of this manuscript through conceptualisation, project administration, writing - review & editing.


.....
(signature)

mgr Ola Grabowska
University of Gdańsk, Faculty of Chemistry
Department of General and Inorganic Chemistry
Wita Stwosza 63, Gdańsk, Poland
ola.grabowska@ug.edu.pl

Gdańsk, 01.12.2022

STATEMENT

I declare that I am a co-author of the following publications included in the PhD thesis:

Affinity and putative entrance mechanisms of alkyl sulfates into the β -CD cavity

Małgorzata M. Kogut, Ola Grabowska, Dariusz Wyrzykowski, Sergey A. Samsonov.

Published in Journal of Molecular Liquids, 2022, 364:119978.

I contributed to the preparation of this manuscript through formal analysis, investigation, methodology, visualisation, writing – original draft, writing – review & editing.

Physicochemical nature of sodium dodecyl sulfate interactions with bovine serum albumin revealed by interdisciplinary approaches.

Aleksandra Tesmar, Małgorzata M. Kogut, Krzysztof Żamojć, Ola Grabowska, Katarzyna Chmur, Sergey A. Samsonov, Joanna Makowska, Dariusz Wyrzykowski, Lech Chmurzyński.

Published in: Journal of Molecular Liquids, 2021, 340:117185.

I contributed to the preparation of this manuscript through investigation and resources.

Effect of Tetrphenylborate on Physicochemical Properties of Bovine Serum Albumin.

Ola Grabowska, Małgorzata M. Kogut, Krzysztof Żamojć, Sergey A. Samsonov, Joanna Makowska, Aleksandra Tesmar, Katarzyna Chmur, Dariusz Wyrzykowski, Lech Chmurzyński.

Published in: Molecules, 2021, 26(21):6565.

I contributed to the preparation of this manuscript through formal analysis, investigation, methodology, visualisation, writing – original draft, writing – review & editing.


.....
(signature)

dr Aleksandra Tesmar
University of Gdańsk, Faculty of Chemistry
Department of General and Inorganic Chemistry
Wita Stwosza 63, Gdańsk, Poland
aleksandra.tesmar@ug.edu.pl

Gdańsk, 01.12.2022

STATEMENT

I declare that I am a co-author of the following publications included in the PhD thesis:

Physicochemical nature of sodium dodecyl sulfate interactions with bovine serum albumin revealed by interdisciplinary approaches.

Aleksandra Tesmar, Małgorzata M. Kogut, Krzysztof Żamojć, Ola Grabowska, Katarzyna Chmur, Sergey A. Samsonov, Joanna Makowska, Dariusz Wyrzykowski, Lech Chmurzyński.

Published in: Journal of Molecular Liquids, 2021, 340:117185.

I contributed to the preparation of this manuscript through conceptualisation and data curation.

Effect of Tetraphenylborate on Physicochemical Properties of Bovine Serum Albumin.

Ola Grabowska, Małgorzata M. Kogut, Krzysztof Żamojć, Sergey A. Samsonov, Joanna Makowska, Aleksandra Tesmar, Katarzyna Chmur, Dariusz Wyrzykowski, Lech Chmurzyński.

Published in: Molecules, 2021, 26(21):6565.

I contributed to the preparation of this manuscript through conceptualisation and data curation.

.....*A. Tesmar*.....
(signature)

dr hab. Joanna Makowska
University of Gdańsk, Faculty of Chemistry
Department of General and Inorganic Chemistry
Wita Stwosza 63, Gdańsk, Poland
joanna.makowska@ug.edu.pl

Gdańsk, 01.12.2022

STATEMENT

I declare that I am a co-author of the following publications included in the PhD thesis:

Physicochemical nature of sodium dodecyl sulfate interactions with bovine serum albumin revealed by interdisciplinary approaches.

Aleksandra Tesmar, Małgorzata M. Kogut, Krzysztof Żamojć, Ola Grabowska, Katarzyna Chmur, Sergey A. Samsonov, Joanna Makowska, Dariusz Wyrzykowski, Lech Chmurzyński.

Published in: Journal of Molecular Liquids, 2021, 340:117185.

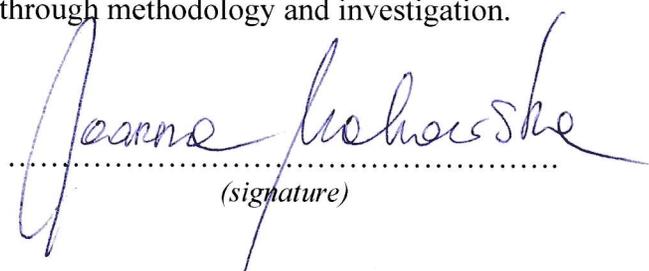
I contributed to the preparation of this manuscript through methodology and investigation.

Effect of Tetraphenylborate on Physicochemical Properties of Bovine Serum Albumin.

Ola Grabowska, Małgorzata M. Kogut, Krzysztof Żamojć, Sergey A. Samsonov, Joanna Makowska, Aleksandra Tesmar, Katarzyna Chmur, Dariusz Wyrzykowski, Lech Chmurzyński.

Published in: Molecules, 2021, 26(21):6565.

I contributed to the preparation of this manuscript through methodology and investigation.


.....
(signature)

dr inż. Krzysztof Żamojć
University of Gdańsk, Faculty of Chemistry
Department of General and Inorganic Chemistry
Wita Stwosza 63, Gdańsk, Poland
krzysztof.zamojc@ug.edu.pl

Gdańsk, 01.12.2022

STATEMENT

I declare that I am a co-author of the following publications included in the PhD thesis:

Physicochemical nature of sodium dodecyl sulfate interactions with bovine serum albumin revealed by interdisciplinary approaches.

Aleksandra Tesmar, Małgorzata M. Kogut, Krzysztof Żamojć, Ola Grabowska, Katarzyna Chmur, Sergey A. Samsonov, Joanna Makowska, Dariusz Wyrzykowski, Lech Chmurzyński.

Published in: Journal of Molecular Liquids, 2021, 340:117185.

I contributed to the preparation of this manuscript through methodology and investigation.

Effect of Tetraphenylborate on Physicochemical Properties of Bovine Serum Albumin.

Ola Grabowska, Małgorzata M. Kogut, Krzysztof Żamojć, Sergey A. Samsonov, Joanna Makowska, Aleksandra Tesmar, Katarzyna Chmur, Dariusz Wyrzykowski, Lech Chmurzyński.

Published in: Molecules, 2021, 26(21):6565.

I contributed to the preparation of this manuscript through methodology and investigation.


.....
(signature)

Prof. Valdecir F. Ximenes
Sao Paulo State University
Department of Chemistry
Sao Paulo, Brasil
valdecir.ximenes@unesp.br

Sao Paulo, 01.12.2022

STATEMENT

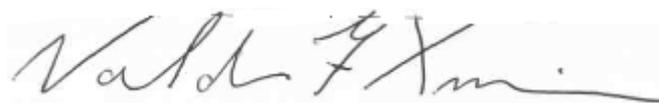
I declare that I am a co-author of the following publications included in the PhD thesis:

Induced Circular Dichroism as a Tool to Monitor the Displacement of Ligands Between Albumins.

Luiza de Carvalho Bertozo, Małgorzata M. Kogut, Martyna Maszota-Zieleniak, Sergey A. Samsonov, Valdecir F. Ximenes.

Published in: Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2022, 278:121374.

I contributed to the preparation of this manuscript through conceptualisation, supervision, resources, methodology, investigation, data curation, writing – original draft, writing – review & editing.



.....
(signature)

Luiza de Carvalho Bertozo
Sao Paulo State University
Department of Chemistry
Sao Paulo, Brasil
luiza.bertoza@unesp.br

Sao Paulo, 01.12.2022

STATEMENT

I declare that I am a co-author of the following publications included in the PhD thesis:

Induced Circular Dichroism as a Tool to Monitor the Displacement of Ligands Between Albumins.

Luiza de Carvalho Bertozo, Małgorzata M. Kogut, Martyna Maszota-Zieleniak, Sergey A. Samsonov, Valdecir F. Ximenes.

Published in: Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2022, 278:121374.

I contributed to the preparation of this manuscript through conceptualisation, methodology and investigation.

Luiza de Carvalho Bertozo