

Faculty of Chemistry, University of Gdańsk

M.S. Mateusz Piotr Marcisz

# Development of novel computational approaches for glycosaminoglycans

Supervisor:

PhD Sergey Samsonov

Department of Theoretical Chemistry

Laboratory of Molecular Modeling

Gdańsk 2023

## **Abstract**

Glycosaminoglycans (GAGs) are linear anionic periodic polysaccharides that play a crucial role in various biologically relevant functions within the extracellular matrix. Through their interactions with proteins, GAGs mediate processes such as cell proliferation, cancer development, inflammation and the onset of neurodegenerative diseases. Experimental approaches suffer from difficulties in the investigations of protein-GAG systems because of the complex nature of the GAGs. Computational studies proved to be helpful in addressing some of the challenges faced by experimental approaches. Nonetheless, GAGs have not received as much attention from the computational community as other biomolecule classes, leading to a lack of modeling tools specifically designed for their theoretical analysis. As a result, researchers, for example, must rely on existing docking software developed mainly for small drug molecules that differ significantly from GAGs in terms of their basic physico-chemical properties. It would be of great importance to develop and implement new tools that allow computational GAG community to study these biologically relevant molecules with the ease, precision and accuracy similar to the ones in computational studies of other groups of biomolecules. If the mentioned issues become solved theoretical approaches can not only complement the experimental studies but also successfully investigate areas that are not yet accessible for the experimental research.

The main goals of this PhD thesis were to develop new computational tools for GAG-containing molecular systems and examination of GAG interactions using computational approaches. To achieve these goals a set of theoretical approaches were designed and applied to specific biologically relevant systems involving GAG molecules.

First, the analysis and revision of currently available molecular docking tools was performed. This allowed for developing new approaches targeting analysis of GAG interactions. One of them is a coarse-grained model representing GAG monosaccharide units that has been developed to make the analysis more accessible for less experienced researchers in the computational field. In this approach, a ready-to-use script was provided for a user to elongate a GAG molecule in the analyzed complex. Then, a novel Molecular Dynamics (MD) based docking tool named RS-REMD (Repulsive Scaling Replica Exchange Molecular Dynamics) has been implemented for studies of protein-GAG systems. In this method, van der Waals radii are being scaled in each consecutive replica which allows for a faster sampling of the system.

Next, improvements regarding the use of explicit water model have been implemented further enhancing accuracy of the RS-REMD technique. In the second part of the PhD thesis, interactions of GAG with proteins have been investigated. In order to accomplish this, a representative nonredundant dataset of protein-GAG complexes has been analyzed. The effect of length of the GAGs on binding to protein has been studied complemented by comprehensive analysis of technical computational aspects regarding the performance of the MD simulations. Then, in the investigation of GAG influence on the APRIL (A proliferation-inducing ligand) protein and its receptors - TACI (Transmembrane activator and CAML interactor) and BCMA (B-cell maturation antigen) - new molecular mechanism of APRIL-receptor complex forming facilitated by GAG binding was proposed. At last, the role of water in GAG MD simulations has been studied. The influence of particular solvent models on highly sulfated GAGs has been described. It was shown that TIP5P and OPC water models performed essentially better than widely the used TIP3P model.

This PhD thesis presents data that expand the general understanding of GAG-related systems. Additionally, novel computational techniques were designed for GAG molecular docking approaches, in particular, and GAG system computational analysis strategies, in general. The results show the importance and high potential of theoretical approaches as powerful tools in studying protein-GAG interactions.

## Streszczenie

Glikozoaminoglikany (GAG) to liniowe, ujemnie naładowane, periodyczne polisacharydy, które pełnią kluczową funkcję w biologicznie istotnych procesach zachodzących w macierzy pozakomórkowej. Poprzez oddziaływania z białkami, GAGi pośredniczą w procesach takich jak proliferacja komórek, onkogeneza, stan zapalny i rozwój chorób neurodegeneracyjnych. Metody eksperymentalne okazują się często niewystarczające, aby dokładnie zbadać układy białko-GAG ze względu na złożoną naturę GAGów. W takich przypadkach pomocne są metody komputerowe, które bardzo dobrze radzą sobie z wyzwaniem, przed którym stawiane są badania układów zawierających GAGi. Niemniej jednak społeczność naukowców teoretycznych nie poświęca GAGom tak dużo uwagi jak innym klasom biomolekuł, co skutkuje brakiem specjalistycznych narzędzi zaprojektowanych do ich badań. W rezultacie badacze muszą korzystać z istniejących programów, opracowanych głównie do dokowania małowładczystekowych leków, które znacząco różnią się od GAGów pod względem ich podstawowych właściwości fizykochemicznych. Rozwój i wdrożenie nowych narzędzi, które umożliwią badanie tych biologicznie istotnych cząsteczek z łatwością, precyzją i dokładnością podobną do tych stosowanych w badaniach obliczeniowych innych grup biomolekuł, ma więc ogromne znaczenie. Głównymi celami niniejszej pracy doktorskiej było opracowanie nowych narzędzi obliczeniowych dla systemów molekularnych zawierających GAGi oraz badanie oddziaływania GAGów z biomolekułami za pomocą metod obliczeniowych. Aby osiągnąć te cele, zaprojektowane zostały nowe metody teoretyczne, które następnie zastosowano do istotnych biologicznie kompleksów białek z GAGami. W pierwszym etapie prac przeprowadzona została rewizja dostępnych narzędzi dokowania molekularnego. Pozwoliło to na opracowanie nowych metod dostosowanych do układów zawierających GAGi. Jedną z nich jest model gruboziarnisty, w którym jednostki monosacharydowe GAGów są reprezentowane jako kuliste centra oddziaływań. Opracowanie takiego modelu umożliwia prowadzenie badań obliczeniowych dla wydłużonych cząsteczek GAGów w analizowanych kompleksach. Kolejnym narzędziem, które może zostać wykorzystane do dokowania GAGów, jest opracowana w ramach badań doktorskich metoda oparta na dynamice molekularnej (MD) z wymianą replik RS-REMD (ang. Repulsive Scaling Replica Exchange Molecular Dynamics). W tej metodzie, w każdej kolejnej replice skalowane są promienie van der Waalsa, co umożliwia szybsze i bardziej efektywne próbkowanie przestrzeni konformacyjnej badanego układu. Następnie dokładność metody RS-REMD została zwiększona poprzez wprowadzenie

do niej jawnego modelu wody. W kolejnej części pracy doktorskiej przeprowadzone zostały badania oddziaływań GAGów z białkami. Korzystając z reprezentatywnego zbioru kompleksów białko-GAG zbadany został wpływ długości GAGów na ich możliwości wiązania się z białkiem oraz wydajność symulacji MD w zależności od różnych parametrów obliczeniowych. Podczas prowadzenia badań nad wpływem GAGów na strukturę białka APRIL (ang. A proliferation-inducing ligand) oraz jego receptorów – TACI (ang. Transmembrane activator and CAML interactor) oraz BCMA (ang. B-cell maturation antigen) – zaproponowany został nowy mechanizm molekularny tworzenia kompleksu APRIL-receptor inicjowany wiązaniem się GAGów. W ostatnim etapie pracy zbadana została rola różnych modeli wody w symulacjach komputerowych. Opisany został wpływ konkretnych modeli rozpuszczalnika na silnie usiarczanowane GAGi. Udowodniono także, że zastosowanie modeli wody TIP5P i OPC pozwala na uzyskanie dużo dokładniejszych wyników niż przy użyciu powszechnie stosowanego modelu TIP3P.

Niniejsza praca doktorska zawiera dane, które znacząco poszerzają dostępną do tej pory wiedzę na temat kompleksów zawierających GAGi. Ponadto, opracowane zostały nowe metody obliczeniowe służące do dokowania molekularnego oraz analizy obliczeniowej systemów zawierających GAGi. Wyniki zawarte w niniejszej pracy pokazują wysokie znaczenie i potencjał metod teoretycznych w badaniach cząsteczek GAGów.

## Table of contents

1. Introduction .....	8
1.1. Glycosaminoglycans .....	8
1.2. The role of water in GAG-containing systems.....	14
2. Methods used in GAG-related studies .....	16
2.1. Experimental approaches for GAG-containing systems .....	16
2.1.1. Nuclear magnetic resonance.....	16
2.1.2. Electron microscopy.....	18
2.1.3. X-ray diffraction.....	19
2.1.4. Surface plasmon resonance .....	19
2.1.5. Bio-layer interferometry.....	20
2.1.6. Capillary electrophoresis.....	21
2.1.7. Circular dichroism.....	21
2.1.8. Isothermal titration calorimetry.....	22
2.2. Computational approaches for GAG-containing systems.....	23
2.2.1. Electrostatic potential calculations.....	23
2.2.2. Molecular docking.....	24
2.2.3. Strategies and application of GAG molecular docking.....	29
2.2.4. Molecular Dynamics .....	31
2.2.5. Coarse-grained approaches.....	33
2.2.6. Molecular dynamics of GAG-containing systems .....	33
2.2.7. Free Energy Calculations .....	38
2.2.8 Water models and their role in molecular dynamics.....	41
2.2.9. Water models in computational GAG studies.....	44
3. Goals of the PhD .....	45
4. Summary of the research papers used in this thesis .....	46
D1. Evaluation of replica exchange with repulsive scaling approach for docking glycosaminoglycans .....	46
D2. Explicit solvent repulsive scaling replica exchange molecular dynamics (RS-REMD) in molecular modeling of protein-glycosaminoglycan complexes.....	49
D3. Modeling Protein-Glycosaminoglycan Complexes: Does the Size Matter?.....	50
D4. Further analyses of APRIL/APRIL-receptor/glycosaminoglycan interactions by biochemical assays linked to computational studies .....	52
D5. Advanced Molecular Dynamics Approaches to Model a Tertiary Complex APRIL/TACI with Long Glycosaminoglycans.....	54

D6. Solvent models benchmark for molecular dynamics of glycosaminoglycans.....	56
D7. Modeling glycosaminoglycan-protein complexes.....	57
5. Conclusions .....	59
Other publications in my PhD studies.....	62
List of abbreviations.....	63
List of Figures and Tables .....	66
Figures:.....	66
References .....	67

# 1. Introduction

This PhD thesis is divided into five chapters: 1. Introduction; 2. Methods used in glycosaminoglycan-related studies; 3. Goals of the PhD; 4. Summary of the research papers used in this thesis; 5. Conclusions. In the introduction, current knowledge regarding glycosaminoglycans and the role of the water in glycosaminoglycan-containing systems is discussed. Both physiochemical properties and biological functions of glycosaminoglycans are highlighted by corresponding scientific findings reported in research papers and reviews. Later, in the second chapter, both experimental and theoretical methodology used to study glycosaminoglycan containing systems is presented. Further, the aims of PhD research described in this thesis are explained. In the fourth chapter, scientific papers published within my PhD studies are summarized and their results are presented. In the fifth chapter, conclusions corresponding to the set goals are drawn.

## 1.1. Glycosaminoglycans

Cells express numerous molecules that perform diverse biological functions. One class of these essential molecules are glycosaminoglycans (GAGs), a particular type of biopolymers. They are long linear periodic negatively charged polysaccharides. They are the key players in numerous biologically relevant processes in the extracellular matrix. By interacting with a variety of proteins they influence proteins' functions.<sup>1</sup> GAGs are crucial for processes such as cell proliferation,<sup>2</sup> tissue regeneration,<sup>3-5</sup> neuroplasticity,<sup>6,7</sup> cell maturation,<sup>8,9</sup> cell signaling<sup>10</sup> and angiogenesis.<sup>11</sup> They also affect pathological processes such as cancer development<sup>11-13</sup> inflammatory response<sup>13</sup>, infection,<sup>10,14</sup> cardiovascular diseases<sup>15</sup> and Parkinson's and Alzheimer's diseases<sup>16</sup> or infamous Sars-Cov-2.<sup>17-20</sup> Additionally, by attracting numerous molecules in the connective tissues and thus allowing loadbearing capabilities, GAGs take part in the osmotic swelling response.<sup>21</sup>

All this is possible due to their intramolecular interactions with growth factors (GFs),<sup>22-24</sup> chemokines,<sup>25-27</sup> integrins,<sup>28</sup> morphogenes<sup>29</sup> and lipoproteins.<sup>30</sup> Many of these protein-GAG interactions are considered non-specific.<sup>31,32</sup> In some cases though, they can be highly specific<sup>33</sup> or selective.<sup>34</sup> The GAG-protein interactions are predominantly electrostatic driven and occur between sulfate and carboxyl groups of the GAG units and positively charged lysine, arginine,

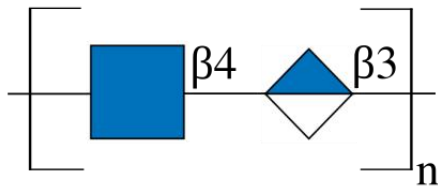


and sometimes histidine residues on protein surfaces.<sup>35</sup> GAG binding often represents “multipose binding”,<sup>36–39</sup> which means that GAG molecules may bind to multiple regions of the protein or they may exhibit multiple binding patterns/positions in the same region that correspond to almost identical binding strength.

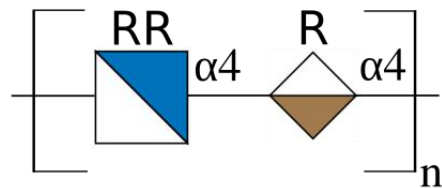
GAGs are built of repeating disaccharide units made up of aminosugar and uronic acid or galactose.<sup>40</sup> They are synthesized and sulfated in the Golgi apparatus.<sup>41</sup> They may exhibit different sulfation patterns that alter their binding properties and conformational characteristics.<sup>40,42</sup> Depending on their sulfation pattern and saccharide composition they may display 408 disaccharide unit variants.<sup>43</sup> 202 of these combinations are expressed in mammal cells.<sup>1,44</sup>

There are several groups of GAGs: hyaluronic acid (HA), heparan sulfate (HS), heparin (HP), chondroitin sulfate (CS), dermatan sulfate (DS) and keratan sulfate (KS). Most popular form of graphical representation of GAG molecule is SNFG (Symbol Nomenclature for Glycans).<sup>45</sup> All the types of GAGs are shown in Figure 1 using SNFG representation. Additionally, basic properties regarding repetitive unit charge, composition and size of GAG molecules have been listed in Table 1.

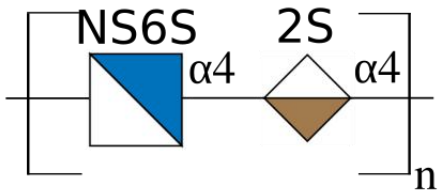
## Hyaluronic Acid



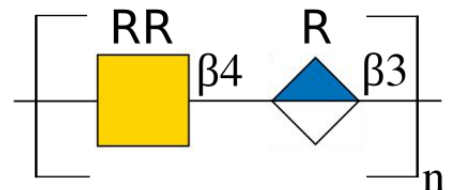
## Heparan Sulfate



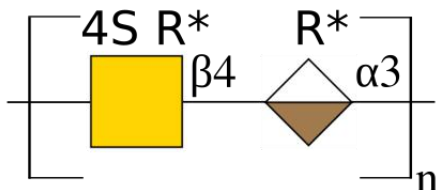
## Heparin



## Chondroitin Sulfate



## Dermatan Sulfate



## Keratan Sulfate

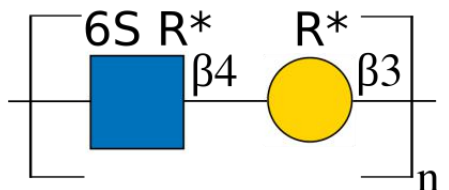


Figure 1. Particular GAG types (represented in SNFG representation), their composition and sulfation patterns. \*Depending on the classification DS and KS can be defined as groups of chemically different molecules (according to such classification DS and KS could be sulfated in 3 different positions) or a specific molecule (in case of DS: GalNAc4Sβ1-4IdoA and KS: GlcNAc6Sβ1-3Gal).

All the GAGs with the exception of hyaluronic acid are bound to proteins forming proteoglycans, a subclass of glycoproteins. GAGs can be linked to proteins via N- and O-glycosylation, one of the most ubiquitous post-translational modifications that occur in about 10% of the protein structures deposited in the Protein Data Bank (PDB).<sup>46</sup> The carbohydrate percentage in the glycoprotein usually varies between low 1% and high 80% of its total weight.<sup>47</sup> In turn, proteoglycans, a heavily glycosylated subclass of glycoproteins, can contain

more than 95% of a carbohydrate part, and therefore, their properties bear a much closer resemblance to polysaccharides than to proteins.<sup>41</sup>

Table 1. Characteristics of GAG molecules: charge per disaccharide unit, composition, bond types and size of molecule. “dp” stands for the degree of polymerization. \* Depending on the classification DS and KS can be considered as groups of molecules (therefore, the charge would be either -2 or -3 depending on the sulfation) or a specific molecule (in case of DS: GalNAc4Sβ1-4IdoA and KS: GlcNAc6Sβ1-3Gal).

Name	Charge per dp <sup>2</sup>	GAG composition and glycosidic bond type	Size of the GAG (dp)
Hyaluronic acid (HA)	-1	GlcNAc (β1→4) GlcA(β1→3)	15- 3*10 <sup>4</sup>
Heparan sulfate (HS)	-1 to -4	GlcNS (α1→4) IdoA(α 1→4)	50-400
Heparin (HP)	-4	GlcNS (α 1→4) IdoA (α 1→4)	20-100
Chondroitin sulfate (CS)	-2 or -3	GalNAc (β 1→4) GlcA (β 1→3)	80-200
Dermatan sulfate (DS)	-2*	GalNAc (β 1→4) IdoA (α 1→3)	100-400
Keratan sulfate (KS)	-2*	GlcNAc (β 1→3) Gal (β 1→4)	10-70

HA is the only GAG that is not covalently bound to the protein as a part of proteoglycan. Despite displaying a fairly simple structure with repeating *N*-acetylgalactosamine (GalNAc) and glucuronic acid (GlcA) without any sulfation its biological function is very complex.<sup>48,49</sup> HA takes part in the response to injury and inflammation, cancer formation, resistance, and cell migration.<sup>50</sup> It is a hydrophilic molecule that can bind the water of weight of a thousand times of its own and form an extended random coil structure in an aqueous environment.<sup>50</sup> Due to its water-binding abilities HA surrounding chondrocytes establishes a biomechanical structure protecting cartilage.<sup>51</sup> It also plays a role in controlling ovulation via maintaining proper

structures of the cumulus ECM.<sup>52</sup> In the endothelium HA prevents inflammatory cells from interacting with endothelial cells by maintaining a proper nonadhesive surface.<sup>53</sup> HA often acts as an insulation or mechanical buffer that blocks high-molecular weight (HMW) components from reaching the cell's surface without blocking the diffusion of small nutrients or electrolyte molecules.<sup>54</sup> Size of the HA molecules may vary and thus influence their properties. It has been shown that high molecular weight (HMW) hyaluronic acid of mass higher than 1 MDa or around  $dp2.5 \times 10^3$  (dp means degree of polymerization) has anti-inflammatory activity,<sup>55</sup> it inhibits phagocytosis,<sup>56</sup> elastase release,<sup>57</sup> and respiratory burst activity.<sup>58</sup> At the same time low-molecular weight (LMW) hyaluronic acid of around 5-20 kDa/dp16-60, increases expression of proinflammatory chemokines and iNOS (Nitric oxide synthase) in macrophages.<sup>59,60</sup> High levels of LMW HA can be observed during myocardial infarction,<sup>61</sup> arthritis<sup>62</sup> and transplant rejection.<sup>63</sup> Additionally, the metabolism of the HA is of a peculiar order: around one-third of the HA undergoes turnover daily which is one of the highest turnover rates among molecules in ECM.<sup>64,65</sup> Medical and therapeutic application of HA involves visco-supplementation in arthritis, pain control, tissue augmentation, surgical adhesions, and wound healing.<sup>66,67</sup> It also may serve as a drug delivery vector or filler in skin wrinkling.<sup>50,68</sup>

HS is a diverse group of GAGs of which 48 dp2 unit variants are expressed in mammals.<sup>44</sup> It is built of repeating units of uronic acid and aminosugar that vary in the number and position of sulfate groups attached to these units.<sup>69,70</sup> HS is linked to the protein core that can be large (up to 500 kDa) and thus heavily affecting the function of the formed heparan sulfate proteoglycan (HSPG).<sup>71</sup> The two main protein groups that HS binds to are syndecans and GPI-anchored glypicans. Rarely, HS can also be connected to neuropilin, betaglycan, and CD44.<sup>30,72</sup> HS can be partially desulfated by endo-6-sulfatases or Sulfs, unlike other GAGs whose sulfation pattern rather does not change after synthesis.<sup>73,74</sup> Additionally, HS interacts with a variety of proteins and very often those interactions are largely electrostatic mediated.<sup>75</sup> Moreover, HS showcase GAG's multipurpose binding properties very often, therefore, may adapt several energetically comparable binding positions and conformations to better fit their protein partner.<sup>76</sup> It was also proven that one unique fold, motif, or amino acid pattern that could define HS binding site does not exist.<sup>30,77</sup> HS binding may influence function of proteins belonging to the groups such as growth factors, morphogens and migration factors.<sup>71</sup> HS is responsible for the bFGF (basic Fibroblast Growth Factor)-mediated fibroblast growth and myoblast differentiation.<sup>78</sup> It is also often required for their activity as HS-FGF interaction is necessary for FGF binding to its high affinity receptor.<sup>79</sup> Another example of the HS's important part in controlling GFs is their role in influencing hepatocyte growth factor expression and activity which in turn affect multiple

myeloma carcinogenesis.<sup>80</sup> HS also participates in cell signaling as it was proposed that HS is pivotal for the midkine's neurite growth activity.<sup>81,82</sup> Another big group of proteins that are heavily influenced by HS is chemokines which control transplant rejection, inflammation, autoimmunity and wound healing through their interactions with HS.<sup>83-87</sup>

HP which could be categorized as a particular highly sulfated case of HS with a clearly defined sulfation pattern is often regarded as an individual GAG class due to its specific nature, complex interactions with its binding partners and a multitude of cellular functions controlled by protein-HP interactions. Due to its particularly high and diverse biological relevance, HP is the most investigated GAG as about half of the GAG studies include HP.<sup>88</sup> HP is crucial for the proper mast cell granules cytokine and protease content and thus for their function.<sup>89-91</sup> HP and its mimetics are widely used in medical treatment as anticoagulants to prevent the development of thrombosis.<sup>92-94</sup> Besides its anticoagulant activities HP also regulates metabolism by many diverse mechanisms and can be used as therapeutic in acute pancreatitis caused by hypertriglyceridemia.<sup>95,96</sup> HP is also used in cancer treatment,<sup>12,97</sup> inflammatory diseases,<sup>98,99</sup> diabetic associated complications<sup>100</sup> and neurodegenerative disorders.<sup>101,102</sup> Last but probably one of the most meaningful roles of HP nowadays is its role in COVID-19.<sup>17,19,20,103</sup> It was shown that LMW HP is a potent drug in the fight with coronavirus and could decrease mortality through anti-viral, anti-inflammatory and anti-coagulant activities.<sup>104-108</sup>

CS, another GAG class, depending on the number of sulfates, sulfation pattern, and uronic acid epimerization can be distinguished into seven different CS groups: nonsulfated, trisulfated, CS-A (chondroitin-4-sulfate), CS-B (chondroitin-2,4-sulfate) also known as DS, CS-C (chondroitin-6-sulfate), CS-D (chondroitin-2,6-sulfate) and CS-E (chondroitin-4,6-sulfate).<sup>109</sup> The length of CS, similarly to DS, varies in the range of dp80-200 which is significantly lower than the length of HA.<sup>110,111</sup> CS is abundant in blood vessels, tendons, ligaments and skin.<sup>112</sup> CS plays diverse therapeutical roles, e.g., may be used as a drug for osteoarthritis,<sup>113,114</sup> immune-mediated responses,<sup>115,116</sup> inflammation.<sup>117,118</sup> In chondrocytes, it regulates the nuclear transactivation of NF-kappaB and reduce nitroprusside-induced apoptosis.<sup>119</sup> CS is also responsible for the development, cell progression and pathogenesis of the central nervous system (CNS).<sup>7,120</sup> Similarly to other GAGs CS of different molecular weights varies in its biological effects, e.g., HMW CS induces synoviocyte growth and proliferation more effectively than its LWM counterpart.<sup>121</sup> It was also shown that not only the size of CS but also its sulfation pattern matters as different CS variants differently affect FGF-2's binding properties to its receptor.<sup>122</sup>

DS is very similar to CS in its function and structure. In fact, DS used to be named CS-B and classified as a particular variant of CS. The unique feature of the DS is an iduronic acid (IdoA) in place of the GlcA. The length of the DS is reported to be between dp100-400.<sup>123</sup> Its function is very diverse and overlaps with the functions of other GAGs, e.g., DS also has anticoagulant activity, however, the mechanism where it is involved seems to differ from the mechanisms where other GAGs are involved.<sup>124</sup> DS was also shown to play an important role cardiovascular disease.<sup>125</sup> Additionally, DS was found in the capillary endothelial cells in angiogenesis associated with inflammation.<sup>126</sup> Its role in angiogenesis was also shown in cancer-related studies where removal of DS resulted in the reduction of metastasis.<sup>127</sup> Moreover, multiple studies reveal the role of DS in cancer development.<sup>128-131</sup> There have been also multiple disorders reported to be associated with the disruption of DS synthesis.<sup>132-134</sup> Besides this, DS is important for homeostasis,<sup>135</sup> tissue development,<sup>136</sup> cell migration,<sup>137</sup> wound healing,<sup>138</sup> and assembly of ECM.<sup>123</sup>

KS is built of *N*-acetyl- $\beta$ -glucosamine (GlcNAc) and  $\beta$ -galactose (Gal) as it is the only GAG that does not consist of any acidic unit, either iduronic or glucuronic acid.<sup>139</sup> KS located in the cartilaginous tissues may also contain some fucose residues, *N*-acetylneuraminic acid and sialic acid<sup>140</sup> and thus is distinguished as an KS-II.<sup>141</sup> KS in brain was named KS-III as it also differs from the regular KS by containing mannose units.<sup>142</sup> Although KS is the least investigated GAG its role and significance have been lately described in several studies.<sup>141,143</sup> KS can bind growth factors and morphogenes and thus is involved in corneal development and healing.<sup>144</sup> It has been also shown that KS takes part in cell signaling.<sup>145</sup> KS plays a role in tissue hydration and weight bearing.<sup>146</sup> KS forming a proteoglycan within osteoadherin may regulate apoptosis and growth of the osteoblast cells<sup>147,148</sup> as well as their differentiation and mineralization.<sup>149</sup> It also plays a role in tooth<sup>150</sup> and cornea<sup>151</sup> development as well as arteriogenesis<sup>152</sup> and atrioventricular canal formation.<sup>153</sup>

## **1.2. The role of water in GAG-containing systems**

Water is essential for all living organisms and is pivotal for very heterogeneous processes such as proton transport or protein folding. It is also immensely important in GAG-containing molecular systems. It was proposed that protein-GAG interplay can be stabilized by solvent molecules which function as structural water that help in maintaining proper interactions.<sup>34,154-</sup>

<sup>157</sup> In contrast to protein-protein interfaces in protein-GAG interfaces there are around three times more water molecules and about half of the protein-GAG contacts are mediated by water.<sup>158</sup> Another argument for the inestimable role of water is the importance of electrostatic interactions in GAG systems.<sup>159</sup> Relevance of dynamic behavior of the water surrounding of GAGs on the proper saccharide unit's conformation was reported in multiple studies,<sup>160-162</sup> e.g., water's role is important in the binding of GAGs to hydroxyapatite surfaces.<sup>163</sup> It was also proposed that the presence of water is necessary for the GAG's pivotal role in sustaining the toughness of the bone tissue.<sup>164</sup> Some studies show the importance of the desolvation penalties in the GAG binding affinity suggesting how crucial the solvation component and thus water is in protein-GAG interactions.<sup>165,166</sup> Water also mediates the contacts of GAGs and phospholipids in the synovial fluid.<sup>167</sup> In the study of Sarkar *et al.* it was shown that water is essential for the proper binding site conformation and is involved in the recognition of the ligand.<sup>166</sup> Van Dam and coworkers showed the importance of water in the structural properties of hyaluronic acid.<sup>168</sup> There are several studies analyzing protein-GAG complexes that illustrate that both water bridging and non-bridging interactions are necessary for GAG recognition.<sup>154,156,169,170</sup> In the work of Jana *et al.* it was reported that the hydration shell of a protein-GAG complex is essential not only for facilitating the recognition process but also for preserving the complex's structure.<sup>154</sup> Authors also claim that reduction in water density in the protein-GAG complex is a crucial factor in facilitating the recognition between the protein and the carbohydrate, as it involves the removal of water molecules from their binding residues. In the study of Gandhi and Mancera, the authors reported that 2-O-sulfate group's interactions of IdoA residues with protein are mainly mediated by water.<sup>170</sup>

## **2. Methods used in GAG-related studies**

GAGs have been investigated using a multitude of different experimental and theoretical approaches. This relatively high interest in GAG molecules among other carbohydrates or their derivatives may be caused by their high biological relevance, ubiquitous presence in the eukaryotic cells as well as curious, complex and complicated high structural variety. Thus, obtaining more information regarding GAGs' structure and properties could potentially elucidate their interaction with their pivotal protein targets. Therefore, identification of GAG binding sites, recognition of protein motifs specific to GAGs and explanation of protein structural changes or oligomerization events induced by GAGs would help in understanding many key processes ongoing in our organisms.

### **2.1. Experimental approaches for GAG-containing systems**

#### **2.1.1. Nuclear magnetic resonance**

Nuclear magnetic resonance (NMR) is a widely used technique to study GAG systems.<sup>32,171,172</sup> Only X-ray surpasses it in the amount of solved GAG structures. One of the main reasons for the popularity of this method in the GAG field is the high flexibility of the investigated molecules and therefore limited use of other techniques for structural investigations. In this spectroscopic method using electromagnetic waves in the range of radio waves nuclear spins of the investigated molecules are affected.<sup>173</sup> Afterwards, during the relaxation of the spin the nucleus emits specific electromagnetic waves corresponding to specific energy states that can be detected and analyzed. Every nucleus can be characterized by the energy spectrum that is determined by its chemical environment. Therefore, using NMR one can solve the structure of the investigated molecule by analyzing the number and position of chemical shifts (a resonant frequency of a nucleus that is relative to a standard in a magnetic field).<sup>174</sup> NMR phenomenon is used in different NMR-based techniques such as saturation transferred difference, chemical shift perturbation, or transferred nuclear Overhauser effect. Additionally, many state-of-art methods including solid state NMR, isotopic enrichment and paramagnetic labeling of GAG ligands or WaterLOGSY (Water–ligand observed via gradient spectroscopy) have been used in GAG studies. Among those methods, the most popular one is the analysis of chemical shift



perturbation (CSP).<sup>175</sup> In protein-GAG studies CSP is usually used with <sup>15</sup>N-labeled proteins allowing for the identification of GAG binding sites.<sup>172</sup> This technique was successfully applied to characterize the binding site and interactions of Interleukin-8 (IL-8) with GAGs.<sup>176</sup> Arguably the biggest contribution of the NMR to the GAG field was solving the structure of different GAG variants. For example, knowledge of the heparin dp12 structure (ID PDB: 1HPN<sup>177,178</sup>) was a huge leap in the GAG structural studies and also allowed computational approaches to not only study the unbound HP molecule using this experimental structure as a reference but also to model HP in the complexes with proteins where the interactions were confirmed but the experimental structure was unavailable. Another example of GAG's solved structure using NMR is CS: in the work of Sattelle *et al.*<sup>179</sup> structural model of the dp6 CS (PDB id: 2KQO<sup>179,180</sup>) was proposed and it was shown that 4-sulfation has little effect on its backbone conformation. NMR spectroscopy was used also to solve the structures of HA (PDB ID: 2BVK<sup>181,182</sup>) showing that HA folds into a contracted left-handed 4-fold helix.<sup>181</sup> In the study of García-Mayoral *et al.* this technique was used to solve the structure of Eosinophil cationic protein (ECP) – HP trimer complex.<sup>183</sup> Amino acid residues significantly contributing to the GAG binding have been analyzed and it was shown that <sup>2</sup>S<sub>0</sub> IdoA ring conformation is preferred by protein for the GAG recognition. Nuclear Overhauser effects (NOEs) analysis was successfully applied to characterize conformations of the IdoA rings in the HP disaccharides. It was shown that IdoA is flexible and exists in the form of equilibrium between <sup>1</sup>C<sub>4</sub>, <sup>4</sup>C<sub>1</sub> and <sup>2</sup>S<sub>O</sub> conformers in the free state.<sup>184</sup> NOEs were also used in a similar study of Hricovini *et al.*<sup>185</sup>, where the HP/HS-mimicking synthetic saccharides were used to unravel the importance of the IdoA ring conformation on the biological activity of the antithrombin III showing that <sup>2</sup>S<sub>O</sub> conformation of IdoA was dominant. Another widely used approach in the GAG studies by NMR GAG is the use of paramagnetic labels.<sup>186–188</sup> This technique allows for the detection of weak protein-GAG interactions and straightforward determination of the orientation and position of the ligand on the protein surface.<sup>171</sup> One example of the research conducted using paramagnetic labels is the work of Moure *et al.*<sup>189</sup> where <sup>15</sup>N labeled amide nitrogens of lysine residues were used allowing for the detection of the binding site of HP and determination of the atomic level model of the complex.

### 2.1.2. Electron microscopy

Electron microscopy (EM) is a relatively new technique if it comes to solving molecules' structures with high resolution. Nevertheless, lately, this technique began rapidly surging in popularity. The idea behind this technique is based on the fact that the wavelength of the electron is tens/hundreds of thousands of times shorter than visible light wavelength which technically allows for that much greater resolution in comparison to the optical microscopy. In the EM the source of illumination is the beam of the electrons, which are shot by an electron emitter (e.g., tungsten filament cathode) and then accelerated by an anode. Next, this beam of electrons possessing very high energy pass through investigated sample and hit the detector producing electric potential maps similarly to X-ray-derived electron density maps. EM can be categorized depending on many aspects. The most popular variants are transmission electron microscopy (TEM), scanning electron microscopy (SEM) and cryogenic electron microscopy (CryoEM). Although the big interest in EM emerged lately, coinciding with the award of the Noble prize in chemistry for the development of CryoEM in 2017, already in the year 1990 EM was proven to be useful in GAG studies as used for the analysis of the shapes of different GAGs as it showed many structural features of the KS, DS, CS and HA.<sup>190</sup> The EM technique developed specially to improve its resolution is CryoEM. Although this technique is still developing and not at its full potential, CryoEM allows to combine some of the best features from both NMR and crystallography. Similarly to NMR spectroscopy, this method allows for the analysis of more natural conformation of the investigated molecule in contrast to what happens in crystals used in X-ray crystallography. At the same time, CryoEM allows for studying a much bigger protein-GAG complex than NMR. Even though this technique still has much room for improvement, especially if it comes to the resolution of the obtained structures, it already produced promising results, e.g., the work of Maloney *et al.*<sup>191</sup> shows the structure of the complex of hyaluronic acid synthase (HAS) with the HA and complemented by MD simulations allowed for the identification of residues interacting with HA inside of the HAS channel unraveling the mechanism of HA biosynthesis. This work demonstrates the high potential of the CryoEM method and introduced a new competitor in GAG structural biology field. Although very promising, CryoEM is not the only option in the EM arsenal as SEM was also successfully applied in studies of GAGs. This method described in detail by Koutsakis *et al.* allows for the observing of the cellular appearance in 2D and 3D cultures and thus studying the effect of the added GAGs on the cell morphology.<sup>192</sup> SEM technique was also applied for the investigation of collagen fibrils.<sup>193,194</sup>

### 2.1.3. X-ray diffraction

X-ray crystallography is a widely used technique for the determination of the position and arrangement of atoms in the crystal as it provides access to high resolution up to 0.5 Å. This technique in contrast to NMR and CryoEM is less affected by the size of the studied molecule. That however does not mean crystallography is without any flaws and disadvantages. First of all, the investigated molecule must form crystals. Secondly, if the molecule has flexible regions it may prevent solving the structure of these regions. Lastly, but probably most importantly obtained structure is “a static image” of the molecule which may not be biologically active. In this technique crystalline structure of the investigated sample causes the incoming X-rays to diffract at specific angles. Those diffracted X-rays then hit the detector and form a specific scattering pattern. Afterward using Fourier transformation one can obtain three-dimensional electron density maps from two-dimensional diffraction pictures which then are used to solve the structure of investigated molecules. X-ray crystallography has been the undisputed king of structural biology for many years. According to the RCSB PDB (Research Collaboratory for Structural Bioinformatics Protein Data Bank), it has been used to solve more than 170 thousand structures which is about 12 times more than NMR or EM.<sup>195</sup> This technique was also very successful in the field of structural biology of systems involving GAGs. For example, the structure of DS (dp4 and dp6), dp4 CS and dp4 HA in the complex with AC lyase has been solved using X-ray crystallography.<sup>196</sup> There are other numerous examples of the use of X-ray diffraction in the GAG studies.<sup>197–202</sup> The database for the GAGs called MatrixDB<sup>203</sup> created by Chautard *et al.*<sup>204</sup> and the GAGDB database<sup>205</sup> created by Perez *et al.*<sup>206</sup> indicate that X-ray diffraction is the most popular method of solving the structures of GAG complex and report 117 and 113 solved complexes’ structures, respectively. It shows the importance and power of X-ray methods in the GAG field, especially when coupled with computational methods. Although MD approaches tend to yield the results that agree with X-ray results in general, there may be a divergence between them.<sup>207</sup>

### 2.1.4. Surface plasmon resonance

Surface plasmon resonance (SPR) spectroscopy is a technique that allows for label-free detection of non-covalent interactions noninvasively and in real-time. SPR is a phenomenon that involves light-induced excitement of electrons in the thin metal surface layer. The angle of

the reflected light depends on the refractive index of the environment (studied molecules) near the metal surface. If the light is constant, then the change in the reflective index (due to the presence of the bound molecule to the metal surface) obstructs the SPR phenomenon and thus can be detected as an absence of the signal.<sup>208</sup> Numerous scientific groups describe the applicability and value of this technique in GAG-related studies<sup>209–213</sup> Recently, Sandoval *et al.*<sup>214</sup> described the use of SPR for the detection of GAG-binding proteins. In this work, authors presented the workflow for both identification and also characterization of the membrane and extracellular proteins that may bind HS. The authors also present evidence to support the use of the technique for the prediction of the GAG binding site. SPR spectroscopy was also successfully used for the detection of the interaction between HS and SARS-CoV-2 Spike protein region binding domain.<sup>215</sup> In another study combination of SPR and mass spectrometry was shown as a promising method for probing protein-GAG interactions.<sup>216</sup> Similarly, in the work of Yu *et al.*<sup>217</sup> SPR was employed in the protocol for the structural characterization of GAGs and the detection of their interaction with proteins. Additionally, SPR spectroscopy was able to unravel the kinetic and structural aspects of MPXV (monkeypox virus) A29 protein interactions with GAGs.<sup>218</sup>

### **2.1.5. Bio-layer interferometry**

The concept of Bio-layer interferometry (BLI) is very similar to that of SPR spectroscopy. In BLI, there are also molecules immobilized on the biosensor, but this biosensor then is dipped in the solution with the analyte. Afterward, the interference patterns caused by the binding of the molecule on the biosensor and the molecules of interest are measured. The binding event measurement can provide the user with information on rate constants, reaction rates, or binding strength.<sup>219,220</sup> This technique has been reported to be extremely useful in obtaining HS/HP-protein interaction networks. The map of the interaction of HS and HP with 170 proteins has been presented in the book of Vallet *et al.*<sup>210</sup> In another study BLI (together with NMR and HPLC) was used for the characterization of HS-protein interactions and HS Biosynthesis Enzyme Activity.<sup>221</sup> Groner *et al.* showed that BLI can also be used in medicine-oriented studies to obtain sulfated virus-like nanoparticles to mimic HP anticoagulant activity.<sup>222</sup> BLI technique also allowed showcasing the importance of HS molecules on the intoxication effect of Cytotoxic Necrotizing Factor Y produced by *Yersinia pseudotuberculosis*.<sup>223</sup> To sum up, this

relatively new technique was already proven to be a highly useful tool for GAG investigations in numerous studies allowing for binding detection or kinetics.<sup>224–227</sup>

### 2.1.6. Capillary electrophoresis

Capillary electrophoresis (CE) is a molecule separation method that uses electric current for separation. In this particular version of electrophoresis capillaries of submillimeter diameter are used. This technique, dependent on the conditions and types of equipment can be divided into affinity capillary electrophoresis (ACE), capillary electrokinetic chromatography (CEC), gel electrophoresis (CGE), capillary zone electrophoresis (CZE), frontal analysis continuous capillary electrophoresis (FACCE) and others.<sup>228</sup> In the work of Heegaard *et al.* use of the ACE method allowed for the identification of the amino acid sequence responsible for HP binding to the serum amyloid P component (SAP).<sup>229</sup> The same group investigated also the SAP-HP binding properties and the differences in the binding with and without the presence of the divalent metal ions.<sup>230</sup> Similar studies were also conducted by other groups.<sup>231,232</sup> In one of them, CE allowed for calculations of HP binding affinities to different forms of Cleaved beta 2-microglobulin revealing high differences in the binding potential to HP.<sup>233</sup> Another variant of electrophoresis, CZE was used to assess measured dissociation constants between the HP and SAP-derived synthetic peptides.<sup>234,235</sup> One more example of successful use of CE, this time the FACCE variant, was showcased by Hattori *et al.* in the work where the researchers were able to elucidate the nature of the interaction between HP and BSA.<sup>236</sup> Overall, CE has been shown to help better describe interactions of GAGs and their binding partners in numerous studies.<sup>217,237–241</sup>

### 2.1.7. Circular dichroism

Circular dichroism (CD) spectroscopy is a technique in which optically active molecules (chiral molecules which rotate light) are being investigated in terms of their ability to absorb differently left and right-polarized light. Typically, this method is used for the investigation of proteins and nucleic acids, their structural properties, the stoichiometry of complexes involving these chiral molecules and their conformational changes upon binding ligands.<sup>242</sup> This technique has been

reported to be also an excellent tool for studying saccharides, including GAGs.<sup>243,244</sup> Matsuo *et al.* have used vacuum-ultraviolet circular dichroism (VUVCD) to study in detail the structural features of GAG molecules (including CS, HA, HP), to unravel characteristics of GAGs' constituent functional groups (sulfate, carboxyl, hydroxyl, etc.).<sup>245</sup> CD spectroscopy is also useful for the investigation of GAG's interactions,<sup>246,247</sup> e.g., Stone *et al.* used this technique to prove HP binding to antithrombin and to investigate these interactions. They presented that HP facilitates factor Xa-antithrombin interactions.<sup>248</sup> CD (assisted by MD) was able to characterize conformational changes undergoing in Glial-cell-line-derived neurotrophic factor (GDNF) protein upon binding of different GAG molecules (HS, HP, HA).<sup>249</sup> It was shown that GAGs bind strongly to GDNF and induce the formation of the alpha-helical structure of the N-terminal part of the protein and thus activate the anti-neurodegenerative properties of GDNF.<sup>249</sup> Zsila in his study involving the use of CD proved that CS and HP can bind to berenil and pentamidine, anti-microbial drugs.<sup>250-252</sup> Similar study was conducted by Stanley *et al.* where antimalarial drugs were reported to bind HP and thus prevent GAGs from taking part in prion protein conformational disorders by competitive binding.<sup>253</sup> CD was proven multiple times to be a great technique for the identification of GAG-protein interactions, especially when complemented by molecular dynamics.<sup>254-257</sup>

### **2.1.8. Isothermal titration calorimetry**

The isothermal titration calorimetry (ITC) technique is an excellent choice for the determination of thermodynamic parameters of complexes' interactions in solution by measuring heat. It can provide quantitative information regarding binding affinity and stoichiometry of the analyzed complexes. The measured heat reflects the change of enthalpy. ITC experiments are performed at different temperatures that allows for the calculation of entropy and Gibbs free energy.<sup>258</sup> In the ITC study of Sepuru *et al.* the stoichiometry of the complex between chemokine CXCL5 and GAG was determined. ITC also assisted NMR and MD simulations with the characterization of the interactions of the complex.<sup>156</sup> In the work of Dutta *et al.* the use of ITC for the determination of thermodynamic parameters of protein-GAG systems has been thoroughly described.<sup>259</sup> In another study mucoadhesion properties of the GAG-based nanogels were investigated helping the development of new drug delivery options.<sup>260</sup> Recently Malicka *et al.* studied HP binding to lysozyme in the presence or lack of potassium glutamate.

Researchers were able to provide free energy of binding of this complex and describe the hydration effect.<sup>261</sup> ITC method also helped to further understand the anticoagulant properties of the HP molecule.<sup>262</sup> Overall, this technique showed to be capable of providing various information regarding thermodynamic parameters and interactions, especially when coupled with other experimental and theoretical approaches.<sup>34,263–268</sup>

## **2.2. Computational approaches for GAG-containing systems**

Due to the specific nature of GAGs as their flexibility, size, high charge, periodicity and multitpose binding propensity, they are very challenging to work with using exclusively experimental techniques. To complement experimental studies theoretical approaches can shine a light on the issue and tackle the problems outside of the reach of current experimental techniques.

### **2.2.1. Electrostatic potential calculations**

Electrostatic interactions could be crucial for governing the binding of charged molecules. Computational methods can serve as useful tools to assess these interactions. It was already proven a long time ago that electrostatic potential calculations are able to help understand and predict molecular binding properties.<sup>269,270</sup> GAG involving complexes are an example of systems dominated by electrostatic interactions.<sup>31,88,255,271–273</sup> Therefore, assessing electrostatic potentials can be a very useful method of preliminary investigation of the molecules potentially being part of GAG systems. It was proven that simply predicting proteins motifs consisting of positively charged amino acids can be helpful in predicting GAG binding sites.<sup>274</sup> Additionally, it has been shown that positive electrostatic potential formed by patches of protein residues are promising binding region for GAGs.<sup>275,276</sup> The most recognized method for calculating such potentials is Poisson-Boltzmann surface area (PBSA) method which was proven useful in the production of potential protein's binding sites for GAGs.<sup>277,278</sup> In PBSA in order to model solvent-mediated electrostatic interactions implicit solvent models are used.<sup>279</sup> Solutes represented in an all-atom model containing point charges produce an electrostatic field, both

in the solute and solvent region, which can be computed by solving the Poisson-Boltzmann equation<sup>280</sup>

Equation 1

$$\nabla^*[\varepsilon(r)\nabla\varphi(r)] = -4\pi\rho(r) - 4\pi\lambda(r) \sum_i z_i c_i \exp\left(\frac{-z_i\varphi(r)}{k_B T}\right)$$

where  $\varepsilon(r)$  represents the dielectric constant,  $\varphi(r)$  is the electrostatic potential,  $\rho(r)$  denotes the solute charge,  $\lambda(r)$  is the Stern layer masking function,  $z_i$  represents the charge of ion type  $i$ ,  $c_i$  is the bulk number density of ion type  $i$  far from the solute,  $k_B$  denotes the Boltzmann constant, while  $T$  is the temperature. This approach showed to be very successful in predicting potential binding sites for GAGs as presented in the work of Samsonov *et al.*<sup>281</sup>

### 2.2.2. Molecular docking

Biological processes rely on the interactions that occur between biomolecules. To comprehend these processes and to identify molecules that can be utilized as bioactive substances computational approaches serve as primary tools. Molecular docking has become an indispensable tool in drug discovery and studying molecular interactions since its appearance in the 1980s.<sup>282</sup> This technique is able to predict the preferred orientation of a molecule when it binds to another to create a stable complex.<sup>283</sup> The understanding of the preferred orientation can be utilized to estimate the binding affinity between two molecules. As a result, this technique has become crucial for the development of rational drug design protocols, which include structure-based virtual screening for the identification of new drug candidates and the comprehension of vital chemical components that govern protein-ligand interactions in significant biological targets.<sup>284,285</sup> Identification of such interactions/binding conformations can be divided into two steps: 1. exploring of vast conformational space that represents diverse potential binding modes, including both prediction of the potential binding site and binding pose of the ligand (placement); 2. assessment of the interaction energy related to each of the predicted binding conformation (scoring). Some docking tools can perform so-called blind docking which means that the whole surface of the receptor is considered in the docking simulation. However, most of the docking tools are not that powerful and are limited to



scanning only part of the receptor predefined by the user in advance. In order to define a potential docking site when literature data are not available one may use cavity detection programs such as: POCKET,<sup>286</sup> GRID,<sup>287</sup> SurfNet<sup>288</sup> or PASS.<sup>289</sup> In order to successfully predict binding poses by finding the global energy minimum search algorithms comprehensively scan the potential energy landscape. During the process of conformational search, the structural parameters of the ligands: dihedral, translational and rotational degrees of freedom, are systematically sampled. Various techniques are utilized by conformational search algorithms to accomplish this task.<sup>285</sup> A set of rules and parameters define those search algorithms that are utilized to predict potential conformations of the complex. Those rules and variables may vary depending on the goal and application of the algorithm, e.g., docking algorithms can be classified into two major groups based on the flexibility of ligand and receptor: rigid-body docking and flexible docking. These two sets employ distinct types of algorithms. The rigid-body docking approach takes into account essential geometric complementarities between a receptor and a ligand but does not account for the flexibility of the interacting partners. As a result, the specificity and accuracy of docking results are restricted, but despite the limitations they may be sometimes successful in the precise prediction of the binding site.<sup>290</sup> Although flexible docking, allowing for the conformational changes of the interactors, requires more computational power, its gains are more valuable. This method involves considering multiple potential conformations of either the ligand, the receptor, or both molecules simultaneously. Docking algorithms typically employ multiple conventional techniques for exploring conformational space. Specific search algorithms are utilized and can be divided into several classes depending on the conceptual idea behind and application area of the docking tool. They can be categorized based on the criteria used for classification:

1. Systematic search algorithms:

Systematic search algorithms are a group of algorithms that perform small modifications to the ligand structural parameters, leading to gradual changes in the conformation of the ligand. They aim to investigate all possible configurations of a molecule, taking into account bond rotations, angles and incremental increasing in the molecule's size.<sup>291</sup> Due to the vast number of possible conformations, systematic searches may encounter the issue of a combinatorial explosion which can be avoided with a proper alternative approach.<sup>292</sup>

- *Incremental construction:*

Fragmentation methods obtain ligand conformations by dividing the ligand of interest into fragments and incrementally constructing ligand conformations from these fragments. The initial ligand fragments are docked in the binding site one at a time and are gradually assembled increasing the size of the docked ligand. Alternatively, all fragments are docked into the binding site, and then they are linked covalently to form the final conformation of the full ligand.<sup>293–295</sup>

- *Distance geometry:*

The Distance Geometry algorithm utilizes intra- and intermolecular distances. It uses a reduced set of distance constraints often derived from the experiment and can handle a larger number of constraints. It accomplishes it by gradually adding additional fixed distances or reducing the number of already existing ones.<sup>296</sup>

- *Fast shape matching:*

Fast shape-matching algorithms rely on the geometric complementarity of the molecular surfaces of two molecules. Various algorithms are utilized to generate multiple structural conformations of ligand and receptor. Additionally, fast shape matching can anticipate the feasible conformations of the binding site to assess the geometric suitability of the ligand for the binding site topology.<sup>297,298</sup>

## 2. Stochastic or random search methods

Stochastic or random search techniques involve introducing random modifications to either a single ligand or a group of ligands, which are then assessed using a predetermined probability function, and the probabilistically favorable changes are accepted. To accomplish this, the algorithm produces sets of molecular conformations and explores a broad energy landscape. This approach prevents the final solution from getting stuck at a local energy minimum and enhances the likelihood of discovering a global minimum.<sup>299,300</sup>

- *Monte Carlo algorithm*

Monte Carlo algorithms generate initial configurations of the ligand in the active site, comprising a random ligand conformation, its rotational and translational state. This configuration is evaluated based on specific criteria (e.g., energy), and subsequently, small adjustments are made to produce new configurations. If the new configuration yields a higher score than the previous one, it is preserved. However, if it does not, the Metropolis criterion is employed to determine whether to accept or reject it.<sup>301,302</sup>

- *Genetic algorithm*

Genetic algorithms operate on the concept of population and biological evolution. The arrangement of the ligand and protein is specified by a series of parameters that define the ligand's translation, rotation and internal conformation in relation to the protein. The chromosome encodes parameters that are stochastically varied and evaluated through a fitness function. In molecular docking, the fitness value is determined by the total interaction energy between the protein and ligand. Afterward, a process known as the crossover is employed to randomly combine pairs of chromosomes and produces a new one that inherits the genes from either parent. Additionally, random mutations are produced if they provide better fitness value.<sup>303,304</sup>

#### *- Tabu search algorithm*

The Tabu search algorithm involves introducing several small random modifications to the current ligand configuration and ranking them based on the fitness function. The adjustment with the least number of rejected conformations (taboos) is accepted. This technique exhibited good precision by preventing the simulation from becoming ensnared in local minima and refraining from revisiting previously identified minimal energy conformations.<sup>305,306</sup>

As described above molecular docking serves as a tool to predict ligand poses within the binding site of the given receptor. Afterward, the assessment and ordering of predicted ligand conformations are accomplished using certain approximate mathematical functions referred to as scoring functions. Scoring functions have several crucial applications in molecular docking, namely: identifying the most probable ligand binding mode and predicting binding affinity. Additionally score functions can screen virtual databases to achieve mentioned goals. Those scoring functions can be divided into the following classes:

1. Force field-based scoring

In the development of force field (FF)-based scoring functions classical molecular mechanics FFs are being used that express the energy of the system as a sum of non-bonded (intermolecular) terms involved in molecular recognition. Force field methods could utilize a number of force field parameters. The idea behind this scoring function design is that the energy of binding can be described by a sum of individual uncorrelated terms.<sup>307,308</sup>

2. Empirical scoring

Empirical scoring functions comprise multiple energy terms, and the weights of these terms are determined using experimentally observed values obtained from regression analysis based on

experimental structures and binding energies. The simplicity of these energy terms makes the calculation of binding scores much faster than with force field scoring functions. However, despite their speed and direct estimation of binding affinities, this type of scoring schemes experiences major limitations, such as the penalty term for poor structures or high dependence on the position of the hydrogens.<sup>309-311</sup> Another drawback of this approach is the fact that the use of large training sets can lead to the inclusion of methodological artifacts caused by overtraining.<sup>309</sup> Additionally, it also shows a worse performance when applied to heterogeneous datasets as usually they are trained on homogenous sets.<sup>312</sup>

### 3. Knowledge-based scoring

Knowledge-based scoring functions are developed by analyzing statistical data on intermolecular interactions gathered from various databases. These functions are primarily designed to reproduce molecular structures rather than for predicting binding energies. Some of the popular implementations of such scoring functions include the Potential of Mean Force (PMF) and DrugScore, which rely on pairwise atomic potentials. These scoring functions aim to capture the binding effects implicitly, which could be challenging to model explicitly. This is in an opposition to empirical approaches.<sup>313,314</sup>

### 4. Machine-learning scoring

Similarly to knowledge-based and in contrast to classical FF scoring functions, these functions do not assume a pre-determined functional form for the relationship between binding affinity and the structural features that describe a protein-ligand complex. Instead, they infer the functional form directly from the available data. As a result, machine-learning scoring functions have the possibility to outperform classical scoring functions in predicting binding affinity for a wide range of protein-ligand complexes. However, one needs to be careful when using methods with such functions as although they look powerful, they may fail to properly predict the energies of the complexes that are not sufficiently represented in the available data.<sup>315-317</sup>

One more important way of picking particular poses after docking other than relying simply on the scoring function is clustering. It relies on the assumption that events which occur in clusters are not random. Docking involves a fundamental principle wherein clustering occurs due to molecules being steered towards a low free-energy attractor in the binding region thanks to long-range electrostatic and/or desolvation forces together with shorter-range van der Waals

forces. It was shown that utilization of optimal radius for the electrostatics and desolvation free energies enhances the differentiation of near-native complex structures even further.<sup>318</sup> There are several clustering strategies but most of them simply rely on using the algorithm to find the structures with the largest number of neighbors containing a given number of shared atoms within a specified radius.

In an ideal scenario, the combination of a search algorithm, scoring function and proper clustering procedure should produce a solution that closely matches the position of the natural ligand in regard to the receptor in the experiment. However, in practice, docking algorithms are evaluated based on their ability to reproduce known ligand conformations in an experimental structure within a specific margin of accuracy. They should be constructed with the aim to recognize the conformation closest to the experimental structure as the best solution. In the case of small drug molecules docking algorithms very often succeed in this task.

### **2.2.3. Strategies and application of GAG molecular docking**

Since the first computational investigation on protein-GAG interactions in 1989<sup>319</sup> where two consensus sequences for HP binding have been reported (–XBBXBX– and –XBBBXXBX– where X = hydrophobic residue and B = basic residue), there have been multiple molecular docking studies on GAG-containing systems. Docking tools used then may seem rudimentary to nowadays standards. Although molecular docking for small drug molecules and protein-protein docking is becoming better and better in performance there has not been any major leaps in docking of GAGs. The computational field of GAGs is simply lagging behind tools dedicated to other types of molecules as there have been very few attempts to fix this issue. This is because many researchers rely on docking programs that were not developed specifically for GAGs. Although there is an abundance of docking software, those optimized for other ligands usually do not perform satisfactorily when applied to GAG systems.<sup>320,321</sup> In the work of Uciechowska-Kaczmarzyk *et al.*<sup>320</sup> only a few of the fourteen compared docking tools yielded acceptable results. Additionally, some of those programs were not free of charge. Among the tested software, AutoDock3<sup>322</sup> seems to be the best and most widely used free-of-charge tool for GAG docking.<sup>320,323</sup> Nevertheless, it is still limited in many aspects, e.g., a major flaw of AD3 is the limitation in the number of torsional degrees of freedom that is represented by the number of

rotatable bonds in the ligand. For the long and flexible GAG molecules 32 torsional degrees of freedom allowed by the program are not sufficient to sample the ligand conformational space efficiently, which forces the user to manually picking of the most relevant ones. This also implies that docking results obtained for GAG molecules longer than dp8/10 are at least questionable in terms of the reliability and quality. Some docking programs particularly dedicated to GAG molecules have been based on AutoDock, such as Vina-Carb<sup>324</sup> and GlycoTorch Vina<sup>325</sup>. Both of these programs clearly outperformed AD Vina<sup>326</sup> and Glide<sup>327</sup> for the same protein-GAG dataset,<sup>325</sup> however neither of them took up the glove to challenge the AD3 and to directly compare the results with AD3 results.

Some of the researchers may not be experienced with computational approaches and docking software thus user-friendly servers, such as ClusPro, HADDOCK, or SwissDock, may be very helpful.<sup>328-331</sup> Although their advantage is rather general accessibility than the quality (server programs usually do not offer the level of adjustable options as the regular docking software) they proved to be valuable tools, especially for someone without the expertise or with time-constrained ability to learn the use of standard docking tools. For example, ClusPro in the important study on SARS-CoV-2 helped in describing interactions between SARS-CoV-2 spike protein and HS.<sup>18</sup> HADDOCK showed its potential in a GAG-related study where interactions between heparin and CXCL-8 were analyzed.<sup>38</sup> As an alternative method to the use of a docking software could be the manual placement of the given molecule near the anticipated binding site. Then the obtained structure could be refined by the regular MD simulations in order to find potential binding poses.<sup>214,332</sup>

Lately, a few entirely different approaches emerged to overcome the obstacles experienced by using conventional docking tools for GAGs. One of them is a method called Dynamic Molecular Docking (DMD)<sup>333</sup> which involves using a steered molecular dynamics approach to move a GAG molecule placed somewhere in the bulk solvent towards its binding site on the protein surface. This is achieved by using Jarzyński equality,<sup>334</sup> which describes free energy difference between two states of a system, that applies an additional distance-dependent potential between chosen groups of atoms in the protein and GAG molecule. DMD allows for full flexibility of both GAG molecules independently of their length and receptor. One significant drawback of this method is the need for *a priori* knowledge of the binding site, which is not always plausible. Additionally, the method is computationally expensive due to the size of the periodic boundary box required for the starting intermolecular configuration with the use of an explicit solvent model that is recommended with this approach. One more tool that is potentially useful for docking longer GAG molecules is the fragment-based approach.<sup>335</sup>

This method involves sampling the protein's surface by docking trimeric fragments of a GAG, which afterward are assembled into longer chains based on their overlaps. This approach allows for the docking of longer GAG molecules. Unfortunately, there is a major drawback in case when a GAG binding site is located near negatively charged amino acid residues. While positively charged residues energetically make up for the mentioned inconvenience in case of longer GAG molecules, short 3-residue fragments may be avoiding patches of negative electrostatic potential preventing a continuous assembly of a long GAG. This would simply result in unsuccessful docking if the docking site is in close proximity or involves several negatively charged residues. There are other tools that could potentially overcome all the issues featured in the above-mentioned techniques. RS-REMD is a novel MD-based method initially designed for protein-protein docking.<sup>336-338</sup> In RS-REMD approach, the van der Waals radii increase in different replicas, which physically corresponds to the increase of the size of the atoms without affecting other types of interactions in the system. This approach facilitates a thorough and robust search for binding sites and poses on the protein surface while allowing flexibility of the both docked molecule and the receptor sidechains. When applied to protein-GAG systems it could potentially solve many difficulties characteristic of the GAG docking field.

To summarize, despite the fact of some recent advances in computational approaches dedicated to GAGs there is still big room for improvement, especially considering its potential significance in terms of the high biological relevance of GAGs.

#### **2.2.4. Molecular Dynamics**

Under molecular dynamics (MD), one, in general, understands theoretical chemistry and computational approaches that calculate structural and other physico-chemical properties of biological/chemical systems by analyzing the movement of particles within these systems which are usually accomplished by solving Newton's equations of motion.<sup>339,340</sup> There are, however, several types of approaches for describing the properties and movements of analyzed particles. The most accurate ones rely on quantum mechanics (QM). They calculate the changes in the system by approximating the related time-dependent Schrödinger equation. Nonetheless, there are some downsides to this approach and the biggest one is the computational cost of these calculations due to the extreme complexity and details of the analyzed systems that include the

information about the electronic structures explicitly. To overcome those challenges methods that use further approximations are needed to be employed, such as Monte Carlo (MC) or MD. In the MC method, a stochastic approach, random sampling is used to obtain statistical properties from an investigated system.<sup>340</sup> This approach, however, is not suited for analyzing the evolution of molecular systems in time. For describing complex processes at the atomistic level MD simulations are much more convenient than MC. In MD, specific force fields (FFs) defining interatomic potentials are used to calculate potential energies and forces between any number of given molecules.<sup>341</sup>

Equation 2

$$E(R) = \sum_{\text{bonds}} K_r (r - r_{eq})^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_{eq})^2 + \sum_{\text{dihedrals}} K_\phi (1 + \cos(n\phi - \gamma)) + \sum_{\text{impropers}} K_\omega (\omega - \omega_{eq})^2 + \sum_{i < j}^{\text{atoms}} \varepsilon_{ij} \left[ \left( \frac{r_{min}}{r_{ij}} \right)^{12} - 2 \left( \frac{r_{min}}{r_{ij}} \right)^6 \right] + \sum_{i < j}^{\text{atoms}} \frac{q_i q_j}{4\pi \varepsilon_0 r_{ij}}$$

where E is the function of the Cartesian coordinate set, while R specifies the position of particular atoms.  $K_r$ ,  $K_\theta$ ,  $K_\phi$ ,  $K_\omega$  are force for bonds, angles, dihedral and improper dihedral, respectively, r is bond length,  $\theta$  is the bond angle,  $\phi$  is the dihedral angle,  $\omega$  is the improper dihedral angle, eq describes equilibrium positions, the dihedral term is characterized by  $K_\phi$  multiplicity (n) and phase shift ( $\gamma$ ),  $\varepsilon_{ij}$  is Lennard-Jones well depth,  $r_{min}$  is the distance of potential minimum,  $r_{ij}$  is the distance between i and j atoms,  $q_i$  and  $q_j$  are the charges of i and j atoms, respectively and  $\varepsilon_{ij}$  is a dielectric constant.

Force field parameters and corresponding potential energy functions may be derived from QM calculations and experimental data. Examples of most commonly used FFs are CHARMM,<sup>342</sup> AMBER,<sup>342</sup> GROMOS,<sup>343</sup> OPLS,<sup>344</sup> MMFF.<sup>345</sup> Equations describing forces in mentioned force fields are solved simultaneously in small time steps (typically ~fs) keeping, for example, pressure/volume and temperature in the required range. The product of solving those equations are coordinates of investigated molecules and as a function of time, they represent the trajectory of the system. The evaluation of this trajectory may provide macroscopic properties and a description of dynamical processes occurring in the system.

All atom (AA) simulations have proven that with recent advances in computational hardware of personal computers and supercomputers, it is possible to simulate systems in the millisecond range.<sup>346</sup> Additionally, even the complexity of the simulated system has gone up drastically as now it is possible to observe crowding in cellular systems consisting of membrane, proteins, RNA, metabolites, ions, and water.<sup>347</sup> Although those achievements are of great value, AA MD



simulations are not always able to be performed for very complex systems, especially, when changes at a long timescale need to be observed. For the longer time range of simulations and the improved sampling of the different regions of the configurational space other approaches may provide with more reasonable computational cost.

### **2.2.5. Coarse-grained approaches**

AA simulations can provide a high level of molecular details of any given system. However, sometimes it could not be reasonable to simulate big systems in a long-time scale due to the requirement of colossal computer resources that are not available for most users. A trick to overcome this challenge is to apply a higher level of abstraction to the representation of investigated molecules. It can be done by representing a group of atoms as a united residue and therefore limiting the complexity of the system due to a lower number of interactions and degrees of freedom which leads to more extensive sampling and the ability to simply study the same system but in much longer scale at the same cost.<sup>348</sup> FFs using such models are described as united-atom or coarse-grained (CG) force fields depending on the level of particle abstraction. Those models by their nature tend to provide lower accuracy due to less complex description of the molecules. Nevertheless, such models may be sufficient when atomistic precision is a lesser concern and the main focus is the more global dynamic evolution of the studied system. The parametrization of CG models can follow two ways: top-down and bottom-up. While the latter approach is more physics-based and similar to the development of AA models, in the former experimental data are applied for parametrization.<sup>349</sup> MARTINI FF<sup>350–353</sup> is an example where two of those strategies are combined: nonbonded interactions rely on experimental findings whereas bonded interactions are derived from AA simulations. MARTINI can be applied to systems involving polysaccharides, proteins, and nucleic acids. Such universality is helpful in the examination of a complex system involving carbohydrates (glycoconjugates, functionalized glycomaterials) and the investigation of protein-carbohydrate interactions.<sup>354</sup> Other important examples of commonly used CG FFs are UNRES/UNICORN,<sup>355,356</sup> PRIMO,<sup>357</sup> SCORPION,<sup>358</sup> AWSEM,<sup>359</sup> SIRAH,<sup>360</sup> LAMMPS.<sup>361</sup>

### **2.2.6. Molecular dynamics of GAG-containing systems**

### 2.2.6a. AA simulations

Generating force fields for carbohydrates poses unique challenges. One challenge is the complexity of the tertiary structures of monosaccharides due to their exceptionally high number of chiral centers. Additionally, the molecular geometries and electrostatic landscapes of carbohydrates can be difficult to predict and model because of their electronic characteristics.<sup>362</sup> Development of such force fields goes back to 1968 when the first studies on aldopyranose conformation were conducted.<sup>363</sup> Although around the 80s some of the general FFs for carbohydrates started to emerge,<sup>364–366</sup> there was a need to wait another 10 years for the introduction of first specific carbohydrate FFs implemented in CHARMM, GROMOS and AMBER, which are perhaps the most popular FFs nowadays.<sup>367–369</sup> Moreover, an additional couple of years had to pass before the appearance of more specialized carbohydrate FFs such as GLYCAM<sup>370</sup> that could be considered a breakthrough in the carbohydrate modeling field. Although mentioned GLYCAM, CHARMM, and GROMOS FFs differ in some aspects they usually yield similar/comparable results and have a lot in common.<sup>371</sup> Taking into account the complexity of interactions involving GAG molecules AA MD simulations are the most popular choice as they should provide more accurate results at the atomistic level.

*Simulations with GROMOS:* In 1998 Kaufman *et al.* modeled dp2 and dp3 hyaluronic acid molecules in order to better understand their hydration effect.<sup>372</sup> In this seemingly simple study initial coordinates were taken from the crystal structure, and dynamic reorientation of water molecules near the HA polar groups was observed. One of the first studies on HP using GROMOS was conducted by Veril *et al.*<sup>373</sup> It was a quite burdensome task as at first authors had to derive HP topology with the PRODRG program<sup>374</sup> from the NMR structure (PDB ID: 1HPN).<sup>177</sup> The topology then had to be modified using S—N bond length values from sulfonamide groups of GlcN residues, and the charge values needed to be calculated using the QM approach. Such obtained parameters optimized for HP provided good agreement with the experimental data in terms of average torsions of inter-glycosidic linkages. Another example of the use of GROMOS is the study of Gandhi and Mancera who tried to reproduce the puckering of IdoA2S<sup>375</sup> They were able to show the transition from the skewed boat to the chair conformation in the presence of SPC/E water while not observing a reverse, chair to boat, transition. Cipla and coworkers applied QM and MD simulations to study configurations of CSc and discovered that the use of standard parameters resulted in nonphysical behavior of sulfates such as disorientation and deformation of geometry. Luckily, reparametrization using charges derived from QM improved sulfate groups' behavior and enabled their MD simulations in SPC water.<sup>376</sup>

*Simulations with CHARMM:* CHARMM was used already in 1991 by Scoot *et al.* to successfully investigate structural characteristics of the HA and CS molecules simulated in water and *in vacuo*.<sup>377</sup> Later on Almond and coworkers have analyzed conformational preferences of the HA molecule.<sup>378–380</sup> For this they compared two different CHARMM FFs and performed MD simulations using the TIP3P water model. Additionally, the authors also compared fiber diffraction data with their MD results. X-ray fiber diffraction data were also used by Wiegler *et al.* to study polar interactions CSc<sup>381</sup> The same group in the work of Kaufmann and coworkers conducted almost an identical study, but this time for CSa tetrasaccharide.<sup>382</sup> In 2000 an excellent work of Almond and Sheehan examined the conformational preferences of CS and DS molecules. The study showed similarities and differences between the data obtained from MD simulation in explicit solvent and from fiber diffraction.<sup>207</sup> Ng. *et al.* used CHARMM to study the rigidity and flexibility of the tetrasaccharide linker in proteoglycans in the microsecond scale.<sup>383</sup>

*Simulations with GLYCAM/AMBER:* GLYCAM FF series is by far the most used FF for the simulations of GAG molecules.<sup>371,384</sup> This may be due to the fact that this FF dedicated specifically to carbohydrates right from the beginning in GLYCAM\_93<sup>370</sup> offered a consistent parameter set obtained from geometry optimization by Gaussian 90/92 at restricted Hartree-Fock (RHF). Moreover, it was compatible with the AMBER FFs which allowed for simulations of protein-carbohydrate systems. A few years later, an updated version of GLYCAM was published.<sup>385</sup> After another couple of years, it was followed by GLYCAM\_06,<sup>386</sup> a current version of this FF family. GLYCAM\_06 has solved some problems of his predecessors, e.g., the new version fixed the issue that when compared with crystal structures the first and second hydration shell properties were not reproduced appropriately as experimental rates of diffusion and radial pair distribution between the TIP3P water model and hydroxyl groups were not reflected properly in MD simulations. Already in 1993, general AMBER FF was used by Forster and Mulloy to simulate and understand IdoA ring puckering.<sup>387</sup> GLYCAM has been also used in the MD investigation of the behavior of a library of HP hexasaccharides in the work of Muñoz-García and co-workers<sup>388,389</sup> The same group investigated the effect of the sulfate group substitutions in the conformational equilibrium of iduronate in HP mimicking dp3.<sup>390</sup> In the work of Rodríguez-Carvajal *et al.* conformational behavior of CS was investigated and compared with experimental data.<sup>391</sup> In this work, the researchers calculated the adiabatic energy maps for each of the constituting disaccharides of CS and their variants. In 2014 Samsonov *et al.* using MD and MD-steered techniques showed the importance of explicit solvent in local molecular docking as well as investigated ligand and receptor flexibility in the GAG systems.<sup>333</sup> In the same year the same group using three different techniques, QM, MD

and NMR, characterized conformational properties of GAG monosaccharides. Among others, they compared QM and MD results for *gg/gt/tg* conformations (explained in Figure 2) distribution for GlcNAc and GalNAc rings. In 2016 GLYCAM FF was evaluated in terms of parameters used for the modeling of GAGs.<sup>392</sup> As an outcome new set of parameters has been added to GLYCAM for GAGs containing protonated GlcA and IdoA residues and sulfate groups. Hsieh *et al.* in another study of eight dp6 HS containing differently sulfated GlcN rings revealed the impact of sulfation patterns on the conformation of neighboring IdoA residues in HS molecules.<sup>393</sup> Lately, Potthoff *et al.* revealed the potential influence of the GAG length on their binding to proteins.<sup>394</sup> In this work, authors analyzed Procollagen C-Proteinase Enhancer-1, its GAG binding site and the role of calcium ions in the GAG-protein interactions. The MD simulations performed with GLYCAM\_06 also helped to unravel the multipose binding phenomenon in a microsecond-scale investigation of HP interaction with FGF1 protein.<sup>271</sup> Uciechowska-Kaczmarzyk *et al.* investigated the effect of HP binding on the Vascular Endothelial Growth Factor (VEGF) conformation. It was revealed that HP stabilizes the structure of the VEGF protein in a particular conformation and after the binding to HP this stable complex is able to interact with the VEGF receptor, which is crucial for cell signaling.<sup>254</sup> In another study Ruiz-Gómez *et al.* showed the importance of GAG influence on the enzymatic activity of Matrix metalloproteinases (MPPs) and their complex formations and proved that sulfated hyaluronic acid support the alignment of both MPP2 and MPP3 in fibrillar-like structures.<sup>395</sup> In a recent study Bojarski *et al.* discovered the role of HP and HS maturation of procathepsin B maturation. Authors claim that GAGs are crucial for preserving the appropriate active site conformation, which in turn is required for enzymatic cleavage/maturation.<sup>396</sup>

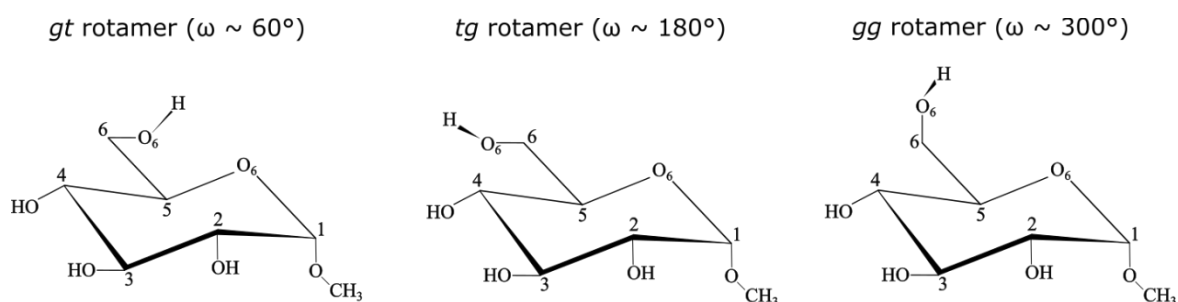


Figure 2. The *gauche-trans* (*gt*), *trans-gauche* (*tg*) and *gauche-gauche* (*gg*) rotamers of methyl  $\alpha$ -D-glucopyranoside. Rotamers are characterized by  $O_6-C_6-C_5-O_5$  torsion angle ( $\omega$  angle).

### 2.2.6b. CG simulations

CG approaches are significantly less demanding in terms of computational resources as the high computational costs of AA GAG simulations are the result of numerous degrees of freedom which require complex sampling schemes. This can be overcome by grouping atoms together into so-called pseudo-atoms and thus reducing the number of degrees of freedom.<sup>397</sup> Molinero and Goddard laid the foundations for the development of CG FFs for GAGs when they proposed their FF for the malto-oligosaccharides.<sup>398</sup> The first published GAG CG study was published by Bathe *et al.* one year later.<sup>399</sup> Authors proposed the CG model for the CS and HA molecules and analyzed their characteristics as a function of pH and ionic strength. Sattelle *et al.* in two studies further improved CG models by emphasizing the importance of ring puckering in CG models and implanting it into their model. Not only they added this feature in addition to inter-glycosidic torsions but also employed two distinct energy functions for puckering motions and inter-glycosidic linkages.<sup>400,401</sup> Another noteworthy approach has been proposed by Samsonov *et al.* where authors have created a method that involves 28 pseudo-atoms representing functional groups based on 17 distinct GAG residues. They have shown that such a set of CG parameters worked well with the AMBER FF. In that work parameters for torsional angles, bonds and virtual bonds of pseudo-atoms were taken from AA FFs, and then restraints were used to reproduce equilibrium values. The investigated GAG set was composed of HP, HS, desulfated HS, CS, DS, HA and their sulfated derivatives which covers all GAG classes with the exception of KS.<sup>277</sup> Around the same time of the works of Samsonov *et al.* and Sattelle *et al.*, Kolesnikov and coauthors proposed a field-theoretic approach for CG modeling of GAGs in aqueous solutions with the addition of varying salt concentrations. This study utilized experimental osmotic pressure data to compute a degree of dissociation, static structure factor, radius of gyration and persistence length for semi-flexible polyelectrolyte chains in the salt solutions. By using adjustable parameters, this straightforward theoretical method demonstrated consistency with the experimental studies.<sup>402</sup> There are also some CG models that focus on carbohydrates but did not implement GAG parameters yet. One example is MARTINI which allowed for the simulation of crystalline cellulose fibers.<sup>403</sup> Another example could be SUGRES-1P<sup>356,404</sup> where parameters for at least some of the GAG molecules should be implemented in short. One more example of potentially useful models in the future is the method proposed by Zhang *et al.* where authors proposed the use of Monte Carlo simulation combined with the torsion-angle MD for a range of biomolecules, including glycoproteins.<sup>405</sup>

### 2.2.7. Free Energy Calculations

Various important biological processes rely on the thermodynamic properties of receptor-ligand binding. Reliable and accurate calculation of binding affinities is, therefore, of immense importance in computational biophysics and drug discovery. Knowledge of binding strength and the contribution of each individual residue participating in the binding process can help in a better understanding of many biological processes and vastly improve drug design. The process of receptor-ligand binding is intricate and involves several factors that influence the binding free energy. These factors encompass the direct interaction between the receptor and the ligand, the desolvation process, the energy strain of the receptor and ligand as they both adopt particular binding conformations as well as the alteration in the configurational entropies of the receptor and ligand during binding. There are several strategies and approaches for calculating binding free energy and although some are more accurate and reliable all of them have a niche to fill. For example, fast end-point techniques, such as empirical scoring functions, are mainly utilized for virtual screening purposes. Then, there are methods based on implicit solvent models such as molecular mechanics/Poisson-Boltzmann surface area (MM/PBSA) or molecular mechanics/generalized Born surface area (MM/GBSA) which are more accurate but also require more computational time. Finally, there are also more rigorous methods like thermodynamic integration (TI) or free energy perturbation (FEP). These methods employ MD simulations with the explicit solvent and are based on statistically mechanically rigorous postprocessing of the MD trajectories.<sup>406-409</sup> The theory of FEP has been already proposed around 70 years ago.<sup>410</sup> Since that time not that much has changed in the implementation of the approach and there has been a relatively small number of improved protocols based on FEP.<sup>411</sup> The basis FEP method proposed in 1954 for rigorous calculation of relative free energy is based on the Nakajima-Zwanzig equation (Equations 3 and 4). The following formula is used by the FEP method to compute the difference in free energy between two systems that have the same number of particles but different potential energies:

Equation 3

$$\Delta F = F_1 - F_0 = -kT \ln \left( \frac{Q_1}{Q_0} \right) = -kT \langle \exp(-\beta \Delta U) \rangle_0$$

where  $\beta = 1/kT$ ,  $\Delta U = U_1(x) - U_0(x)$  is the difference in internal energy, the average is over the ensemble of the initial state that corresponds to the system with potential energy  $U_0(x)$ .  $Q$  denotes the partition function  $Q_i = \int d\Gamma \exp(-\beta U_i)$ . Additionally, the free energy difference could also be expressed as:

Equation 4

$$\Delta F = F_1 - F_0 = kT \ln \left( \frac{Q_0}{Q_1} \right) = kT \langle \exp(\beta \Delta U) \rangle_0$$

in terms of an average over the ensemble of the final state. What is important when simulations are performed for both states (initial state  $U_0(x)$  and final state  $U_1(x)$ ), they might provide significantly different estimates of the free energy difference between those systems. To address this issue method called the Bennett Acceptance Ratio (BAR) is implemented.<sup>412</sup> It combines data from simulations for both states in order to yield a possibly best estimate of the free energy difference and to minimize the variance.

Another useful way of calculating binding free energy is MM/PBSA and its approximation MM/GBSA. Those two related methods are arguably the most used in the GAG-related studies due to their satisfactory accuracy and relatively low computational cost. MM/PBSA was originally described by Kollman and coworkers.<sup>413</sup> This method has been modified many times for different purposes but there is no agreement regarding the details of the best performing protocols as it is heavily dependent on the analyzed system.<sup>308</sup> Free energy state in this method is estimated from the following equation:<sup>413,414</sup>

Equation 5

$$G = E_{bond} + E_{el} + E_{vdW} + G_{pol} + G_{npol} - TS$$

where  $E_{bond} + E_{el} + E_{vdW}$  MM energy terms from bonded interactions (dihedral, angle and bond), electrostatic interactions and van der Waals interactions,  $G_{pol}$  is polar and  $G_{npol}$  is the non-polar contribution to the solvation free energy.  $G_{el}$  term can be calculated using PB equation:

Equation 6

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

where  $\varphi(r)$  is the electrostatic potential in solvent (sol) and vacuum (vac), while  $q_i$  is a charge. In the generalized Born (GB) model, it is expressed as:

Equation 7

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

which naturally approximates MM/PBSA approach to MM/GBSA. In the Equation 7  $r_{ij}$  is the distance between  $i$  and  $j$  atoms, the  $R_i$  are the so-called effective Born radii,  $q_i q_j$  is a charge of atoms  $i$  and  $j$ ,  $\epsilon_w$  is dielectric constant,  $f^{GB}()$  is a certain smooth function of its arguments.

$G_{npol}$  in Equation 5 is a non-polar term that is taken from a linear relation to the solvent accessible surface area (SASA). The TS term is the absolute temperature (T) multiplied by entropy (S) that could be estimated by a normal-mode (NM) analysis of vibrational frequencies or QH. The vibrational frequencies of normal modes can be verified at local energy minima on the potential energy surface using the NM method.<sup>415</sup> Alternatively, the quasi-harmonic (QH) method can be employed, which involves approximating the eigenvalues of the mass-weighted covariance matrix computed from each member of the ensemble as frequencies of global and orthogonal movements.<sup>416</sup>

An even faster end-point alternative method to MM/GBSA is Linear Interaction Energy (LIE) that was developed by Åqvist and coworkers in 1994<sup>417</sup> and was derived from Linear Response Approximation (LRA)<sup>418</sup> in order to compute electrostatic contributions of the binding affinity. To obtain  $\Delta G_{bind}$ , the obtained van der Waals and electrostatic energies are scaled by parameters  $\alpha$  and  $\beta$ :

Equation 8

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

where  $\gamma$  is a constant free energy term, while  $\alpha$  and  $\beta$  are experimentally assessed parameters that for each system should be optimized independently. One may also use default values, e.g., originally  $\beta$  was set to 0.5 for charged ligands, but this approach may result in less reliable results.



Very popular FEP, MM/GBSA(/PBSA), and LIE are not the only methods that are used for energy calculations of various biological systems. Another excellent and particularly rigorous approach is PMF that could be determined by using umbrella sampling (US) or by employing the Jarzyński equality during steered MD simulations.<sup>419,420</sup> PMF calculates the free energy landscape of a system by integrating the potential energy of the system over all possible configurations and provides information on the relative stability of different conformations or states of the system.<sup>421</sup>

Some studies suggest using QM (quantum mechanics) calculations to replace MM approaches or combining those two into QM/MM methods.<sup>308,422</sup> In QM/MM-PBSA approach ligand could be treated with QM and the rest of the system using MM which theoretically should improve the assessment of the binding energy.<sup>423</sup> This methodology however is not widely used in GAG-related studies. In this kind of studies probably MM/PBSA and MM/GBSA are the most popular ones, e.g., it was used to calculate binding free energies in the GAG system involving VEGF-A,<sup>254</sup> TIMP-3,<sup>39</sup> PCPE-1,<sup>394</sup> IL-8,<sup>424</sup> FGF-2,<sup>425</sup> thioflavin T,<sup>251</sup> ellipticine,<sup>426</sup> BMP-2,<sup>427</sup> sclerostin<sup>428</sup> and many others. Despite its very low computational cost LIE has not been as popular method as MM/PBSA to study GAG complexes. However, some studies reported its use as a helpful tool<sup>272,429-431</sup> Although powerful, FEP has not been almost used at all to study systems involving GAGs as only very few studies implemented this method.<sup>166,432</sup>

## 2.2.8 Water models and their role in molecular dynamics

### *Implicit water models*

Implicit models are less complex and accurate than explicit solvents.<sup>433</sup> However, they are still useful, particularly for analyzing large systems with multiple proteins where computational efficiency is crucial. Implicit solvent models due to their lower complexity can drastically speed up MD simulations up to 100 times.<sup>434</sup> It happens by approximating solvent as a continuum and therefore limiting the number of atoms/molecules in the system. Additionally, sampling of the conformational space is much faster in the case of implicit solvents.<sup>434-437</sup> Nowadays GB models are the most popular among implicit models as they are similarly accurate but less time-consuming than Poisson-Boltzman methods.<sup>438</sup> In the AMBER suite used in studies described in this PhD thesis there are several GB models (IGB=1,<sup>439</sup> 2,<sup>440</sup> 5,<sup>441</sup> 7,<sup>442</sup> 8<sup>443</sup> ) implemented and dedicated to different types of systems and purposes. One needs to be aware that GB water

models are not compatible with polarizable force fields.<sup>444</sup> Among their limitations, implicit solvent models can overestimate the strength of salt bridges,<sup>445,446</sup> and yield secondary structure distributions that deviate considerably from those produced by the same protein parameters in explicit solvent. The GB model, in particular, tends to favor excessive  $\alpha$ -helix content.<sup>447,448</sup>

#### *Explicit water models*

Explicit water models provide more accurate solvent characteristics and description of interactions between solvent and solute during molecular mechanics and MD simulations. Those types of water models can be categorized depending on whether the model is flexible or rigid, by the number of so-called sites which describe the number of interaction points within a water molecule, by considering polarization effects. Most of the water models used in MD simulations of biological systems are rigid. Examples of such water models are TIP3P, SPC, TIP4P, OPC, or TIP5P, where van der Waals interactions using Lennard-Jones potential between oxygen atoms, charges on the oxygen and hydrogen atoms (and pseudo-atoms on 4- and 5-site models) describe electrostatic interactions while the bond lengths and angles are fixed.<sup>449</sup> In flexible models, additional potentials with coupling terms are introduced to account for flexibility.<sup>450,451</sup> Polarizable water models are able to reproduce the effect of polarization, therefore they improve screening of interactions that depend on the local environment.<sup>452</sup> Lastly, the number of interaction points in water models used in MD studies may vary from two to six, with the emphasis that 2- and 6-site water models are used only for very specific cases (e.g., 6-site water model may be used for the study of water-ice systems<sup>453</sup>). 3-site water model has 3 interaction points (oxygen and two hydrogens atoms), 2-site models have a hydrogen-like site that averages both hydrogen atoms, the 4-site water model introduces the “M” pseudo-atom with an additional negative charge, 5-site models have two negatively charged “L” dummy atoms that mimic lone pairs of the oxygen atom, while 6-site water model has both “M” and “L” pseudo-atoms. Widely used 3-, 4- and 5-site models have been presented in more detail in Figure 3 and Table 2.

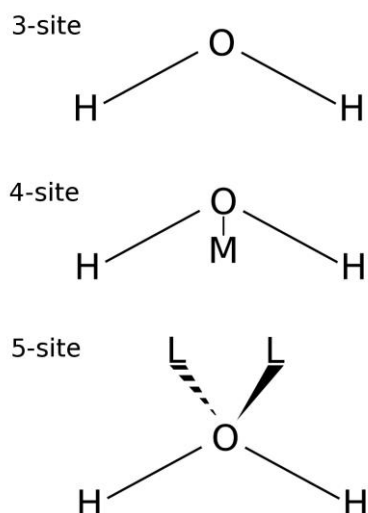


Figure 3. Three types of most commonly used atomistic water models in MD simulations. “M” in the 4-site model represents a pseudo-atom with a negative charge. “L” dummy atoms in 5-site represent lone pairs of valence electrons.

Table 2. Parameters of the selected most used water models.

	TIP3P	SPC/E	TIP4P	TIP4PE <sub>w</sub>	OPC	TIP5P
$q_O$	-0.834	-0.8476	0	0	0	0
$q_H$	+0.417	+0.4238	+0.52	+0.52422	+0.6791	+0.241
$q_M$	N/A	N/A	-1.04	1.04844	-1.3582	N/A
$q_L$	N/A	N/A	N/A	N/A	N/A	-0.241
$r_{OH}$ [Å]	0.9572	0.8724	0.9572	0.9572	0.8724	0.9572
$r_{OM}$ [Å]	N/A	N/A	0.15	0.125	0.1594	N/A
$r_{OL}$ [Å]	N/A	N/A	N/A	N/A	N/A	0.7
$^{\circ}HOH$ [deg]	104.52	109.47	104.52	104.52	103.6	104.52
$^{\circ}LOL$ [deg]	N/A	N/A	N/A	N/A	N/A	109.47

$q_O$  is the charge of the oxygen,  $q_H$  is the charge of hydrogen,  $q_M$  is the charge of M pseudo-atom,  $q_L$  is the charge of electron lone pairs,  $r_{OH}$  is the distance between oxygen and hydrogen,  $r_{OM}$  is the distance between oxygen and pseudo-atom M,  $r_{OL}$  is the distance between oxygen and electron lone pairs,  $^{\circ}HOH$  is the angle between hydrogen atoms and oxygen,  $^{\circ}LOL$  is the angle between electron lone pairs and oxygen. N/A = not applicable.

### 2.2.9. Water models in computational GAG studies

TIP3P is the most popular water model used not only in GAG computational studies but also in other MD-related studies which is reflected in 37949 citations of the original work of Jorgensen *et al.*<sup>454,455</sup> This water model proved a multitude of times to perform properly and is now established as a go-to solvent model in MD involving GAG systems.<sup>169,456–459</sup> There are several reasons for the wide use of the TIP3P water model. The most important is probably its low computational cost when compared to the models with more interaction sites like TIP4P or TIP5P. Considering this 3-site models represent a compromise between precision and cost. Extreme examples of favoring cost over precision are implicit water models. Although they are not as accurate as explicit solvents<sup>433</sup> they are still valuable, especially considering large systems with multiple proteins where computational cost is immensely important. However, nowadays with easier access to high-performance computing, these models become less used. Moreover, GB models may fail to reproduce structural features as in the case of *de novo* designed peptides.<sup>460</sup> It was shown in multiple studies that the use of an explicit water model may improve the overall quality of MD simulations in docking protocols.<sup>169,207,273,380,461,462</sup> It had to be admitted that not only implicit solvents are far from perfection but also 3-site models have their limits, and the use of more complex models may improve MD quality.<sup>463,464</sup> The importance of water in GAG involving systems have been shown in the study of Sarkar *et al.*<sup>166</sup> where it was proven that water dominates specific protein-GAG interactions. Additionally, it was shown that the use of a proper solvent model is important in reproducing proper ring puckering in GAGs.<sup>375</sup> In the study of Neamtu *et al.* influence of different water models on the properties of CS was shown.<sup>465</sup> In that work, the TIP3P model favored intra-molecular H-bonds, while TIP4P and TIP5P water models disfavored them. On the other hand, the TIP5P model yielded a higher number of water bridges along the dp8 CS4 than the other models. It was also shown that the second layer of hydration was better represented in the case of the TIP5P water model. Despite the impact of the water model choice on the reliability of computational studies, there has been very little discussion on this matter as there was only one attempt to partially cover this topic involving GAG system.<sup>373</sup>

### 3. Goals of the PhD

The main goals of my PhD research consisted of the development of new computational approaches for GAG-containing molecular systems and the examination of GAG interactions using them. To achieve these goals various theoretical tools were applied to specific biologically relevant systems involving GAG molecules. Additionally, currently available tools were comprehensively analyzed to better understand the needs of the GAG research community in terms of the development of novel theoretical protocols for both docking and MD data analysis.

This thesis is divided into several parts to better address the complexity and diversity of the investigated issues:

#### 1. Development of novel approaches for GAGs:

- Analysis of currently available docking tools and creation of new ones.
- Design of new approaches for analysis of GAG-containing systems.
- Critical evaluation of the developed MD-based protocols.

#### 2. Analysis of interactions of GAG with proteins in terms of:

- Effect of GAGs binding on protein structural properties.
- Effect of GAGs binding on the protein function.

#### 3. Analysis of the role of water in GAG computational studies:

- Characterization of the water models used in GAG-related studies.
- Elucidation of the effect of the water models on the local and global parameters of GAGs.

## **4. Summary of the research papers used in this thesis**

The full versions of the research papers that make up this PhD thesis are located at the end of the thesis. This section contains only an essential summary of these publications. From here on the letter “D” followed by the number corresponds to the research paper included in the PhD thesis.

### **D1. Evaluation of replica exchange with repulsive scaling approach for docking glycosaminoglycans**

GAGs are extremely hard to work with using both experimental and computational tools due to their specific nature (periodicity, flexibility, length, high charge, multipose binding phenomenon). There has been very little development in the area of docking programs designed specifically for GAGs. Moreover, programs that are currently available usually do not perform at the desired level. In two recent comprehensive studies, the performance of the docking tools was evaluated. In the first study, the evaluation of 6 docking software was conducted on the dataset of all protein-GAG structures available at that particular time in the PDB. In the second study, eight other software were tested using the dataset of 28 complexes with GAG molecules longer than dp3. The results suggested that while several approaches correctly predicted the placement of ligand binding poses, they often poorly assigned scores to the docking poses.<sup>281,320</sup> In order to solve the issue of the lack of GAG-specific docking tools and the dubious performance of the conventional ones, the newly published Replica Exchange Molecular Dynamics with Repulsive Scaling (RS-REMD)<sup>336</sup> was applied to the set of experimentally solved 21 protein-GAG complexes with the length of GAGs ranging from dp5 to dp7. RS-REMD is an MD-based docking tool involving Replica Exchange. Although conventional MD approaches are expected to extensively sample the conformational space of the ligand on the receptor's surface, this process is impractical and costly in reality, as molecular systems often get stuck in local minima in the free energy landscape.<sup>466</sup> To address this issue, biased potentials as implemented in Hamiltonian Replica Exchange approaches can be utilized.<sup>467</sup> The RS-REMD method introduces an extra potential by increasing the effective pairwise van der Waals radii, while leaving other types of inter- and intramolecular interactions unchanged. This

approach has been proven successful for protein-protein docking.<sup>336</sup> Protein-GAG complexes exhibit highly robust electrostatic interactions, which impede dissociation or significant alterations to the GAG binding pose during conventional MD simulations due to the formation of strong charge-charge interactions. To overcome this challenge and to prevent getting trapped in local minima, this method could be especially useful for such systems. Additionally, increasing the van der Waals radii in MD replicas would enable the ligand to explore the protein surface more thoroughly. RS-REMD when applied to 21 protein-GAG systems successfully predicted binding sites in 19 of them. The time required for the simulation convergence varied between 1 to 220 ns (16 of them converged in less than 30 ns), depending on the specific complex being studied. In the majority of cases, the prediction of binding poses was accurate regarding GAG orientation within the binding site, as evidenced by the acceptable differences in the structural ensembles between obtained solutions and the corresponding experimental structures from the PDBs. In the two cases, where the RS-REMD approach failed, binding sites were located in enzymatic pockets. This happened due to the fact that the REMD-RS simulations involve replicas with larger van der Waals radii in each consecutive replica which results in higher volumes of the ligand and receptor atoms than those in the unmodified force field. This is the reason why this method faces difficulty in achieving effective docking in the pocket- or groove-type binding sites. Fortunately, this should not be a vital problem since the vast majority of the GAG binding sites are located at the surface of the protein and only in the case of GAG processing enzymes those sites are found in enzymatic pockets.

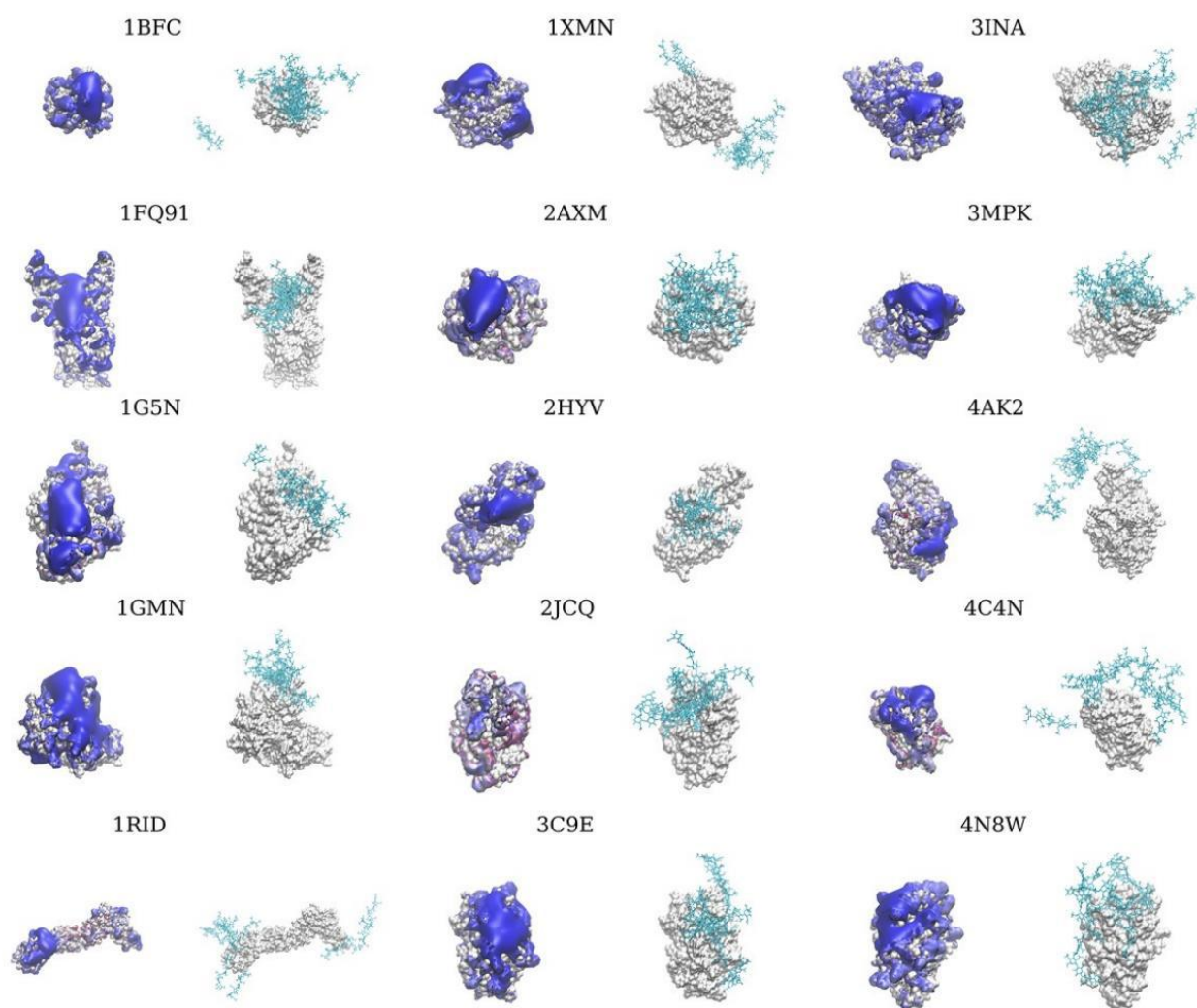


Figure D1. Comparison of positive electrostatic potential isosurfaces obtained using PBSA approach from AMBER suite indicating potential GAG binding sites with the results from the RS-REMD docking (10 energetically best docked poses represented in cyan sticks). Positive electrostatic potential is colored in blue (1BFC:  $5 \text{ kcal/mol} \cdot e^{-1}$ , 1FQ9:  $3 \text{ kcal/mol} \cdot e^{-1}$ , 1GA5N:  $4 \text{ kcal/mol} \cdot e^{-1}$ , 1GAMN:  $4 \text{ kcal/mol} \cdot e^{-1}$ , 1RID:  $2 \text{ kcal/mol} \cdot e^{-1}$ , 1XMN:  $2 \text{ kcal/mol} \cdot e^{-1}$ , 2GXM:  $3 \text{ kcal/mol} \cdot e^{-1}$ , 2HYV:  $5 \text{ kcal/mol} \cdot e^{-1}$ , 2JCQ:  $1 \text{ kcal/mol} \cdot e^{-1}$ , 3C9E:  $5 \text{ kcal/mol} \cdot e^{-1}$ , 3ING:  $4 \text{ kcal/mol} \cdot e^{-1}$ , 3MPK:  $4 \text{ kcal/mol} \cdot e^{-1}$ , 4GK2:  $2 \text{ kcal/mol} \cdot e^{-1}$ , 4C4N:  $1 \text{ kcal/mol} \cdot e^{-1}$ , 4N8W:  $5 \text{ kcal/mol} \cdot e^{-1}$ ). Results correlate well with each other and experimentally obtained binding sites.



## **D2. Explicit solvent repulsive scaling replica exchange molecular dynamics (RS-REMD) in molecular modeling of protein-glycosaminoglycan complexes**

As the recently proposed RS-REMD method proved to be a useful tool for GAG docking, it still utilized an implicit solvent model, which could be inferior to the explicit ones.<sup>468</sup> Changing the implicit solvent to a more accurate explicit solvent model could be of great benefit given how important water is in the GAG-containing systems. The quantity of water molecules in protein-GAG complexes is approximately three times greater than in protein-protein complexes, based on the currently available experimental structures.<sup>469</sup> Roughly 50% of GAG-protein interactions involve mediation by water, making precise modeling of water-mediated interactions crucial for the accurate representation of electrostatic interactions in protein-GAG complexes.<sup>169</sup> Several studies have demonstrated that incorporating an explicit water model can enhance the quality of molecular docking in general.<sup>470–473</sup> Moreover, using an explicit water model in MD simulations and docking procedures has the potential to provide significant benefits in their performance.<sup>207,273,380,461</sup> Some studies have shown that water molecules bridge protein-GAG interactions and potentially function as structural water to facilitate molecular recognition and stabilizing these interactions.<sup>34,154–157</sup> Furthermore, the TIP3P water model is widely utilized and accepted in protein-GAG investigations, having proven effective in this class of systems.<sup>456–458</sup> Therefore, there are reasons to expect that an explicit solvent model, which is more sophisticated, can better describe interactions between a protein and GAG molecules. In this work, an improved RS-REMD protocol including an explicit water model was proposed as explicit solvent is expected to be superior in terms of quality of docking due to providing more realistic interactions between solvent and solutes.<sup>169</sup> To test the new protocol three complexes were chosen: Acidic Fibroblast Growth Factor (FGF1) with HP dp6, Basic Fibroblast Growth Factor (FGF-2) with HP dp6 and Antithrombin III (ATIII) with HP dp8. These complexes were rigorously analyzed and compared with the implicit version of the RS-REMD in terms of docking quality. Significant improvements were observed in the docking performance of the method: RMSatd (root mean squared atom type deviation) decreased from  $6.7 \pm 5.3$  Å,  $5.4 \pm 1.3$  Å and  $10.9 \pm 8.1$  Å to  $5.4 \pm 1.1$  Å,  $4.4 \pm 0.5$  Å and  $2.5 \pm 0.2$  Å respectively giving improvement by 1.3 Å, 1.0 Å and 8.4 Å for the analyzed systems when comparing the implicit and explicit RS-REMD protocols. Ranking by MM/GBSA free energies was proposed as a method of binding pose evaluation. (Figure D2) Additionally, the implementation of the

explicit solvent did not increase the wall-time of docking simulations in comparison to the use of the implicit solvent protocols.

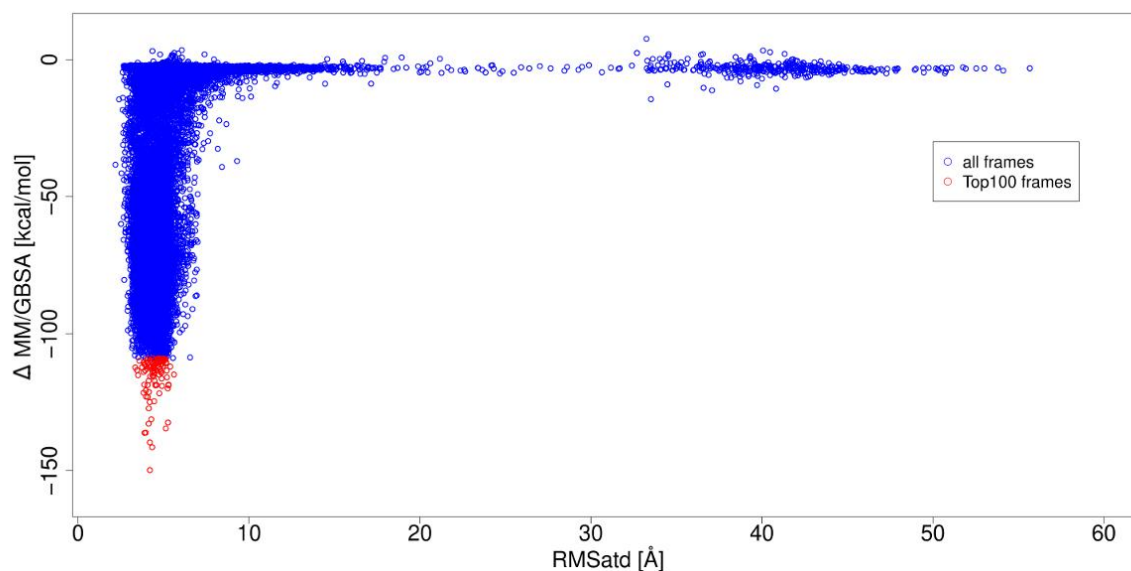


Figure D2. MM/GBSA binding free energies in the protein-GAG complex 2AXM and their correlation to RMSatd values of the GAG molecule (with crystal structure as a reference). RMSatd is a root-mean-square of atomic distances pairing up atoms of the same type that are closest to each other in space. This graph shows that the best solutions from docking represent low RMSatd values, thus confirming ranking by so calculated energy score as a viable option.

### D3. Modeling Protein-Glycosaminoglycan Complexes: Does the Size Matter?

Due to the intricate structure of GAGs, which are long periodic linear negatively charged polysaccharides, docking these molecules has been a difficult task. While standard docking tools like AutoDock3 have been effective in docking GAGs of up to hexameric length, they encounter difficulties in accurately docking longer GAGs. Other docking methods, which are typically designed for smaller ligands, face similar limitations. Additionally, more powerful and advanced docking approaches can prove to be difficult for inexperienced users who lack expertise with complex *in silico* methodologies. To solve this issue, several user-friendly ideas were implemented in this study, which was divided into several parts. Initially, the MM/PBSA

and MM/GBSA methods were employed to determine the binding free energies in a dataset of experimental protein-GAG structures. The dataset for AA and CG GAGs, modeled using previously obtained CG parameters that portray several chemical moieties of the GAG as separate beads, was compared.<sup>277</sup> Afterwards, the general applicability of these free energy calculation approaches for a CG GAG model was justified. Next, a new and significantly simplified CG model of GAG was introduced. Based on AA simulations, this model represents each GAG monosaccharide unit with a single pseudo-atom, which replaces parts of the GAG that do not establish van der Waals contacts with the receptor. These systems with CG components were then simulated, and the disparities between the resulting free binding energies in AA and CG simulations were examined. It was reported that the difference between obtained AA and mixed AA/CG binding energies values is 5.6%. This value is significantly lower than the average difference between the binding energy of dp6 and dp16 GAG which is on average 24% in the investigated complexes. It proved that the use of a mixed AA/CG model is a viable approach for representing protein-GAG systems in terms of their thermodynamic properties. The final objective was to develop a model that permits the calculation of the free binding energy for a GAG of a specified length, without the requirement to explicitly simulate the full-length GAG, using Coulomb or Hückel models of electrostatics. Unfortunately, the influence of the addition of each subsequent ring varied significantly for different systems. Therefore, without additional calibration, it was not possible to assess the energies of any given protein-GAG system. There was also an attempt to analyze the interactions between these GAGs and the protein using only one further modified CG bead to represent the whole elongated segment, e.g., one pseudo-atom for 10 GAG residues. This approach did not yield results as convincing as the regular mixed AA/CG approach and therefore it was not recommended by the authors. All the scripts needed for free energy calculations and elongation of the GAG molecules are included in the supplementary materials of this research paper.

## MIXED MODEL

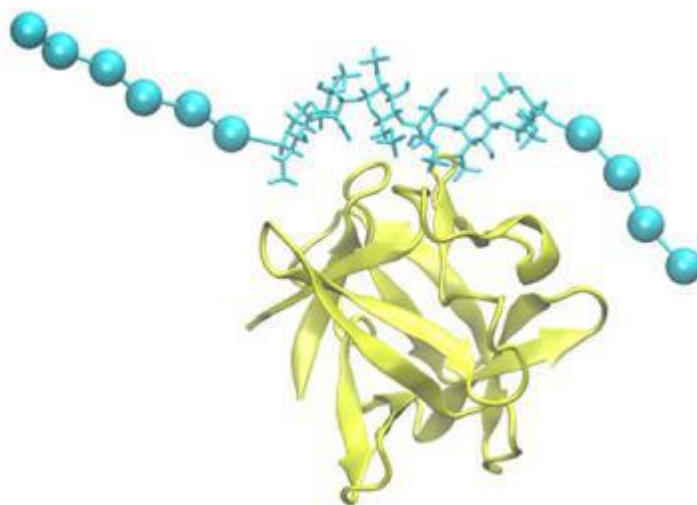


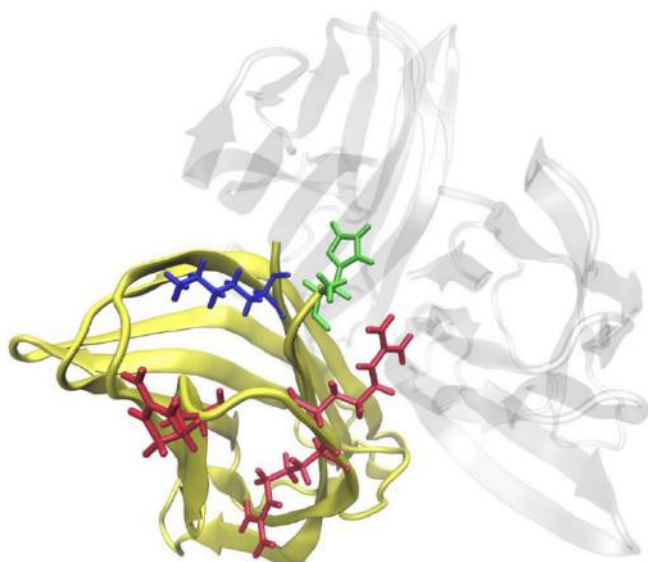
Figure D3. Graphical representation of mixed AA-CG GAG model where AA dp6 HP was docked using AutoDock3 and then using script HP was elongated with additional 10 residues represented as pseudo-atoms that are designed to mimic electrostatic properties of AA residues.

### **D4. Further analyses of APRIL/APRIL-receptor/glycosaminoglycan interactions by biochemical assays linked to computational studies**

APRIL protein (A proliferation-inducing ligand), a member of the tumor necrosis factor superfamily, possesses distinctive features in comparison to its other members. Firstly, it can bind to GAGs through its positively charged N-terminus. Secondly, one of its signaling receptors, the transmembrane activator and CAML interactor (TACI), has been observed to bind GAGs as well. APRIL-GAG binding allows for its oligomerization and therefore to efficiently signal into specific target cells.<sup>474</sup> The main cellular target for APRIL are the antibody-producing plasmacytes.<sup>475</sup> Two isoforms of the APRIL protein were investigated: a full-length wild type of the APRIL and an N-truncated version for which experimental structure is available. In the case of the full-length protein missing the N-terminal part was modeled using the UNRES CG approach and then refined into the AA model. By utilizing experimental biochemical evidence and computational studies with an APRIL deletion mutant, it was shown that APRIL's interactions with GAGs are not limited only to its N-terminus, but some lysine and arginine residues in the C-termini and the loops near N-termini are involved in GAG binding (Figure D4). *In silico* analysis confirmed that APRIL prefers to interact with HP

followed by CS-E while binding with CS-C is less likely or even unlikely. Additionally, both computational and experimental approaches did not report HS binding to the TACI protein. The computational findings, in conjunction with the experimental results, provided atomistic insights into the relevant tertiary complex made up of the protein, its receptor and a GAG. Both computational and experimental approaches agree on the fact that N-termini of the APRIL protein are crucial for the binding of GAG molecules which is reflected in the fact that the truncated version binds GAGs weaker. Moreover, a comprehensive analysis of the free energy calculations was conducted using various techniques to thoroughly evaluate the computational methodologies employed. An optimal number of repeats of MD simulations has been obtained followed by an investigation regarding the minimal length of the MD simulations in GAG-related systems. It was shown that sometimes running multiple short (10-20ns) MD simulations can be more effective than running a few very long ones. Moreover, this approach did not sacrifice accuracy and no potential lack of convergence was reported. Additionally, the docking scoring procedure was checked. It was shown that the selection of the docking poses by clustering was an effective way of picking the most relevant/representative poses from data generated by AutoDock3 and that the investigated poses showed the same distribution for the free energy values as the best 50 results (best poses) from docking.

**Truncated APRIL**



**Full-length APRIL**

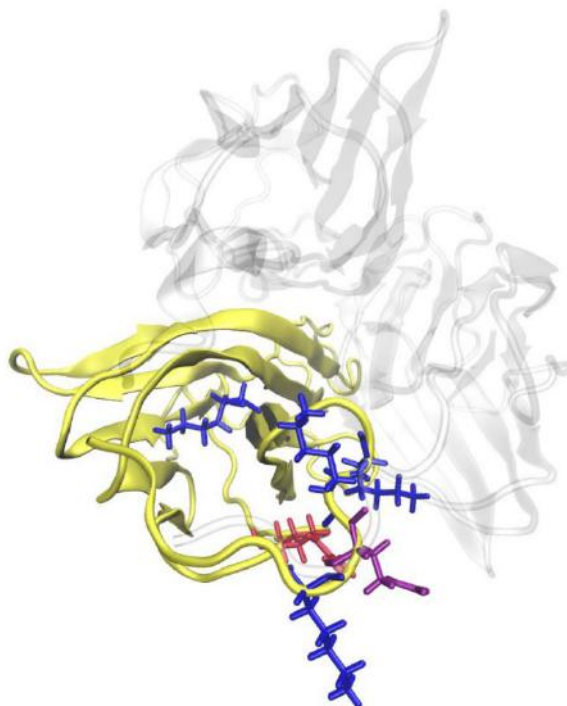


Figure D4. Residues with the highest contribution to GAG binding for both truncated (left) and wild type (right) APRIL protein. Amino acid residues were colored as follows: histidine – green, arginine – red, lysine – blue, glutamine – purple. This shows the major contribution of arginine and lysine to the GAG binding. For the truncated version of the protein N-terminal part consisting of several lysine residues is missing, thus some exposed arginine residues make up for the lacking lysine residues and contribute to the GAG binding instead.

## **D5. Advanced Molecular Dynamics Approaches to Model a Tertiary Complex APRIL/TACI with Long Glycosaminoglycans**

Conventional molecular docking encounters significant challenges when attempting to dock GAGs longer than dp6-10. This study, for the first time, utilized all-atomic repulsive-scaling Hamiltonian replica exchange molecular dynamics (RS-REMD) to dock dp24 and dp48 GAGs. This novel methodology is based on using MD replicas in which van der Waals radii are scaled which allows for faster sampling of potential binding sites. In this work, long GAG molecules were docked to APRIL protein prebound to its receptors: TACI (transmembrane activator and calcium modulator and cyclophilin ligand interactor) and BCMA (B cell maturation antigen). APRIL, a member of the TNF superfamily, has been demonstrated to bind to GAGs, which are believed to facilitate its oligomerization and enable its function in cell signaling.<sup>476-480</sup> The region on APRIL's surface that binds to GAGs is located near the protein's N-terminus, along with a series of positively charged lysine residues. Although this GAG binding is considered to facilitate APRIL's binding to its receptors (BCMA and TACI),<sup>474,481</sup> it has been found that BCMA does not bind to heparan sulfate.<sup>481,482</sup> The interaction between GAGs and TACI is somewhat complicated and even contradictory. While some studies indicate that TACI does bind to proteoglycans,<sup>481-483</sup> others suggest that TACI does not bind to HSPG<sup>484</sup> or deem such binding to be unlikely.<sup>478</sup> The docking and MD investigation revealed stronger binding between heparin and the APRIL-TACI complex was observed than in the case of heparin and APRIL-BCMA complex. This is reflected not only in the binding free energies but also in the number of contacts between molecules in the mentioned complexes. (Figure D5) In all cases, the GAG molecules initially bound to the APRIL GAG binding site, and only afterward interact with the receptors with the chains elongated from this binding site. Researchers were unable to unravel how TACI, which is spatially located far away from the GAG in the case of HP dp24, could significantly affect GAG binding by the APRIL protein. This phenomenon may be attributed to

the long-range electrostatic effect that is present in this highly charged system. These findings suggest that TACI's interaction with GAGs may be enhanced by its binding to APRIL, while this interaction was not observed in the absence of APRIL. Additionally, the performance of Autodock3 and RS-REMD method in docking long GAG molecules was compared. Given that RS-REMD provided superior results for a representative protein-GAG dataset without requiring additional computational resources, we are confident that the RS-REMD method is significantly more effective for docking long GAG molecules.

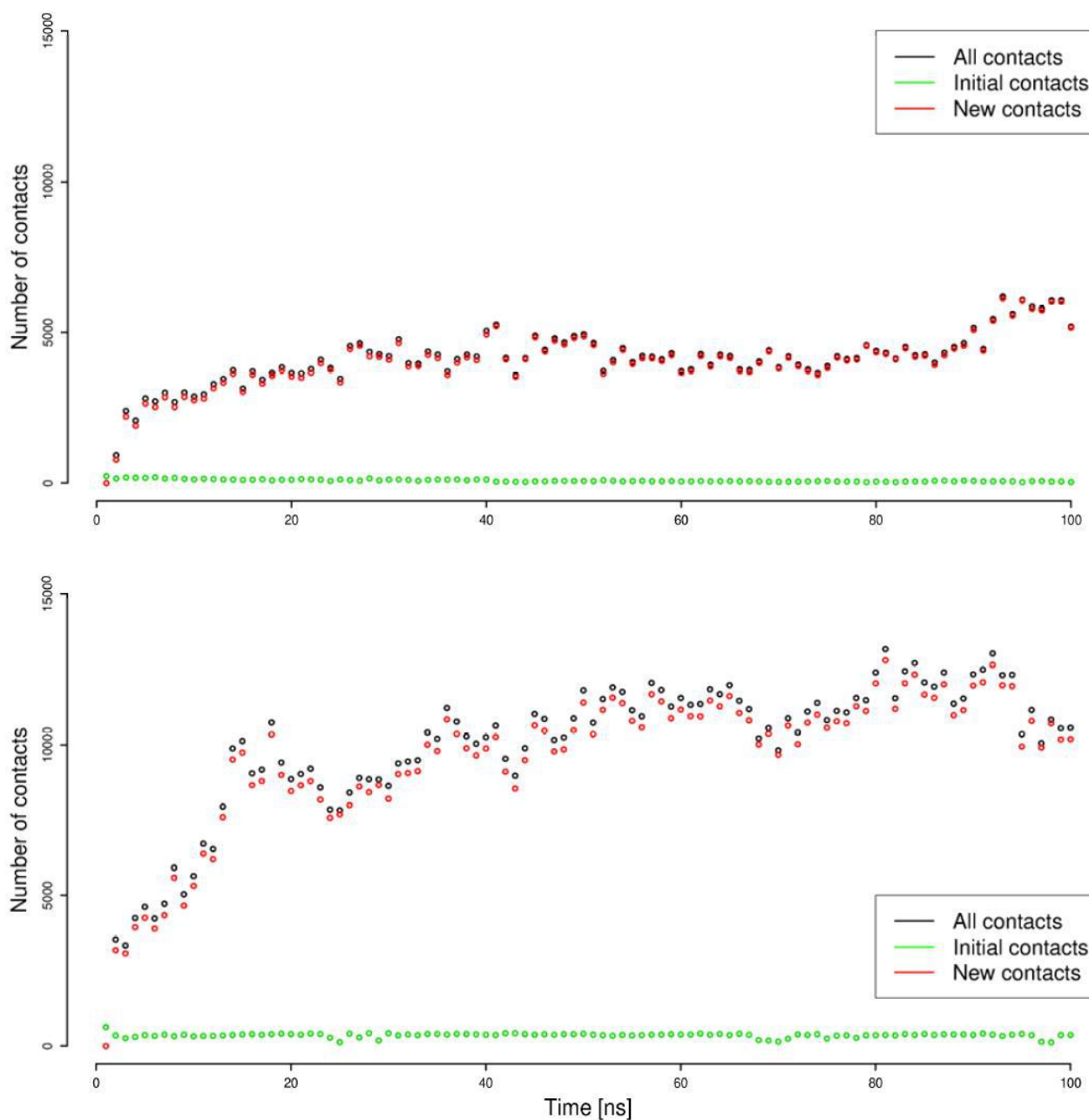


Figure D5. Predicted number of APRIL-BCMA-HP dp48 (top panel) and APRIL-TACI-HP dp48 (bottom panel) contacts. Figure showing more contacts in case of TACI containing complex.

## D6. Solvent models benchmark for molecular dynamics of glycosaminoglycans

There have been almost no studies and scarcely any discussion in the GAG scene regarding solvent models in MD simulations of GAGs. Most GAG-related MD studies utilize the TIP3P water model as it has been widely accepted in the GAG field and has demonstrated its effectiveness in protein-GAG systems, as well as in MD studies of biomolecular systems in general.<sup>456-458</sup> The primary factor driving the prevalent utilization of the TIP3P and other 3-site water models is their lower computational expense in comparison to 4- or 5-site solvent models. However, many explicit and implicit water models, which could be useful in computational GAG research, have been thus overlooked. Up to this day, no comprehensive comparative analysis has been performed on water models that could address the question of which model offers advantages in systems containing GAG molecules. This topic holds great significance because approximately half of the protein-GAG residue contacts in the PDB are mediated by water, and the number of water molecules in protein-GAG interfaces is about 10 times greater than in protein-protein interfaces.<sup>158</sup> The aim of this work was to evaluate the different properties of the HP molecule in various implicit and explicit water models to conclude which of them performs the best in the MD simulation of GAG molecules with the reference to the available experimental data. To achieve this, 5  $\mu$ s MD simulations were performed involving HP, HS and CS decamers together with the following water models: implicit IGB = 1, 2, 5, 7, 8 and explicit TIP3P, SPC/E, TIP4P, TIP4PEw, OPC, TIP5P. As a measure of consistency, five separate 200 ns MD simulations were conducted for each configuration to ensure convergence of the analyzed parameters. The resulting trajectories were utilized to evaluate HP properties, including end-to-end distance, volume, a radius of gyration, ring puckering, hydrogen bonds and dihedral angles. (Table D1) There were significant variations observed in the MD-based molecular descriptors of HP between the groups of implicit and explicit water models employed in the simulations. The IGB=7 model occasionally induced anomalous behavior in the HP structure, resulting in the observation of improper glycosidic linkage populations. However,



this could be expected since this particular solvent model is not compatible with nucleic acids and GAG resembles them by being also long and highly negatively charged. In general, implicit solvent models fail to even qualitatively match experimental results regarding the HP ring puckering. Therefore, unless restraints are applied to maintain the rings in specific pucker conformations that reproduce experimental data, employing implicit models for GAG simulations is not recommended. While explicit water models well reproduced experimental findings regarding local parameters of GAGs (dihedral angles and ring puckering), many of them failed in the case of global parameters (end-to-end distance, a radius of gyration, volume, or the number of hydrogen bonds formed). In the case of the TIP3P water model, abnormal GAG chain curvature and end-to-end distance were observed. Only TIP5P and OPC models provided satisfactory results in these terms. This study shows how significant impact solvent models may have on the modeled GAG molecules in MD simulations and how much work still needs to be done to further test which solvent model is the best choice for MD studies of GAG molecules.

Table D1. HP descriptors obtained from 5  $\mu$ s MD simulations. \*End-to-end distance. \*\*Atomic fluctuation. \*\*\* Radius of gyration. \*\*\*\*Atomic fluctuations and RMSD are compared to the 1HPN structure.

	PDB 1HPN	IGB=1	IGB=2	IGB=5	IGB=7	IGB=8	TIP3P	SPC/E	TIP4P	TIP4PEw	OPC	TIP5P
Dist* [ $\text{\AA}$ ]	41.0	40.4 $\pm$ 2.9	41.6 $\pm$ 2.7	40.4 $\pm$ 3.1	38.1 $\pm$ 13.3	42.4 $\pm$ 2.2	16.1 $\pm$ 6.4	20.7 $\pm$ 7.6	17.3 $\pm$ 7.9	21.2 $\pm$ 6.0	28.4 $\pm$ 4.5	26.1 $\pm$ 4.7
Fluct** [ $\text{\AA}$ ]	N/A****	3.5 $\pm$ 0.7	3.4 $\pm$ 0.6	3.5 $\pm$ 0.8	4.8 $\pm$ 1.6	3.8 $\pm$ 0.4	4.8 $\pm$ 1.4	5.3 $\pm$ 1.6	4.6 $\pm$ 1.9	5.9 $\pm$ 1.3	4.5 $\pm$ 1.0	4.7 $\pm$ 1.2
RMSD [ $\text{\AA}$ ]	N/A****	4.2 $\pm$ 0.7	4.9 $\pm$ 0.7	4.0 $\pm$ 0.8	5.2 $\pm$ 2.9	5.9 $\pm$ 0.4	9.0 $\pm$ 1.5	7.4 $\pm$ 2.0	7.9 $\pm$ 2.2	8.7 $\pm$ 2.6	4.9 $\pm$ 1.4	5.7 $\pm$ 1.4
Radgyr*** [ $\text{\AA}$ ]	12.8	13.6 $\pm$ 0.4	14.1 $\pm$ 0.3	14.0 $\pm$ 0.4	13.0 $\pm$ 3.0	14.0 $\pm$ 0.3	8.8 $\pm$ 1.0	9.9 $\pm$ 1.5	9.0 $\pm$ 1.6	9.5 $\pm$ 1.6	12.4 $\pm$ 1.0	11.6 $\pm$ 1.2
Volume (MVEE) [ $\text{\AA}^3$ ]	5325	7974 $\pm$ 875	8079 $\pm$ 841	8311 $\pm$ 913	7042 $\pm$ 1873	7674 $\pm$ 653	6085 $\pm$ 787	6551 $\pm$ 1032	5853 $\pm$ 1140	6902 $\pm$ 779	7685 $\pm$ 875	7641 $\pm$ 890

## D7. Modeling glycosaminoglycan-protein complexes

In this review, we present recent advances in and the current state of computational studies related to GAGs. Four different topics are discussed: molecular docking of glycosaminoglycans, free energy calculations of protein-GAG complexes, the role of ions in protein-ion-GAG complexes and multipose binding in protein-GAG complexes. First, currently

available docking programs are compared. Additionally, we argue why this area of computational studies of GAGs is behind the ones dedicated to proteins and small drug molecules and what needs to be done in the future to improve the state-of-art of docking approaches for GAGs. In the free energy calculations of protein-GAG complexes chapter techniques dedicated to calculating binding energies of protein-GAG complexes are presented together with their drawbacks or benefits. In the next part, the role of ions in GAG-involving systems is discussed. Additionally, both experimental and theoretical approaches for studying ions within the GAG-containing systems are presented. Next, the mulipose binding phenomenon is explained. Then the implication and consequences of mulipose binding for the studies of GAGs are mentioned. At the end, we draw conclusions regarding the recent state-of-art and advances in the computational GAG field.

## 5. Conclusions

In this PhD thesis, the aim was to develop new approaches for molecular docking and analysis of GAG involving systems and to examine protein-GAG interactions. The conclusions are presented in this section separately regarding each subsequent part of the thesis.

### 1. Development of new tools for GAGs: [D1, D2, D3, D7]

First, the analysis and revision of currently available docking tools was performed. [D1, D2, D7] After the analysis of the present needs in the GAG computational community novel docking approach has been proposed and tested. [D1] In this work 21 protein-GAG complexes have been investigated using the novel docking technique RS-REMD. This Hamiltonian Replica Exchange approach has been proven successful as in 19 of 21 it allowed for the identification of the binding site and it failed only in two cases because of the technique's limitation to dock into the enzymes' cavity which in a majority of the protein-GAG complexes is not an issue. Next, a new upgraded version of the RS-REMD protocol with explicit solvent was developed. [D2] It was rigorously tested on 3 protein-GAG complexes and showed the superiority of the new protocol which included explicit water model TIP3P in contrast to the previous protocol where the implicit solvent was used. The inclusion of TIP3P as a water model provided better docking quality in terms of the similarity of docked poses when compared to experimental structures. Additionally, this new optimized protocol did not increase the need for computational resources and even slightly decreased it providing better performance. New approaches targeting the analysis of GAG interactions have been also developed during the PhD studies. [D3] Strong emphasis was put on making the analysis more accessible for less experienced researchers in the computational field. For this, a new coarse-grained model representing GAG monosaccharide units has been developed. In this approach, a ready-to-use script is provided for the user to elongate the GAG molecule in the studied complex. This allows for either examination of already solved experimental structures or for the use of user-friendly docking tools like AutoDock3 for studying more biologically relevant lengths of the GAG molecules. Additionally, a script was provided that allows for simple and straightforward evaluation of binding energies of the protein-GAG complex with the GAG length specified by the user by employing Coloumb and Hückel models of electrostatics.

### 2. Interactions of GAG with proteins: [D3, D4, D5, D7]

In order to deepen the knowledge regarding interactions of GAG molecules with proteins various complexes as acidic FGF with HP, basic FGF with HP, APRIL with HP, CSc and CSe have been analyzed. In this part of the work effect of the length of the GAG molecules on the binding to proteins was investigated. It was found that as it is rather expected the longer the GAG molecule the stronger the interactions but the effect is not linear. The bigger the elongated GAG molecule, the smaller is the favorable change in free binding energy from subsequent addition of saccharide rings. [D3] Additionally in the investigation of APRIL protein and its receptors (TACI and BCMA) a new mechanism of APRIL-receptor complex forming facilitated by GAG binding was proposed complementing the experimental. [D5] Moreover, an effect of different GAG molecules on binding to the APRIL protein has been studied and supported by the experiment, followed with investigation on the role of specific amino acid residues on the binding of various GAGs. [D4]

### **3. Role of water in GAG MD studies:** [D6]

To investigate the influence of the solvent models used in GAG computational studies, first, the analysis of the currently most used water models was conducted and it was found that the use of explicit water models may improve the quality of docking programs and MD simulations.<sup>456,457,459</sup> In fact, in the work conducted during my PhD it was found that explicit water models are superior to implicit ones if it comes to reproducing experimental findings regarding structural features of GAGs. [D6] Additionally it was found that TIP5P and OPC water models performed essentially better than widely used TIP3P in the case of highly charged GAGs like HP. TIP5P and OPC represented both local (glycosidic linkages, ring puckering) and global features (end-to-end distance, a radius of gyration, RMSD, atomic fluctuations, etc.) of the HP molecule the best. However, there were very few differences in the performance of TIP5P, OPC and TIP3P when it comes to the less charged GAGs as hyaluronic acid or chondroitin sulfate. This showed that the topic of solvent modeling in GAG-related studies is a complex one and further research is required to fully unravel the role of water in GAG computational studies.

The data summarized in this PhD thesis contributes to the widening of general knowledge on GAG involving systems. New computational approaches have been developed for molecular docking and free energy analysis of GAG-containing systems. I strongly believe that these new tools will help scientists further unravel the GAG properties and biological functions of those still not well characterized molecules. In addition to the development of new computational approaches a few protein-GAG systems have been analyzed, such as FGF-HP and APRIL-

HP/CS. A mechanism of potential GAG-mediated APRIL binding to its receptor TACI has been proposed. Finally, the role of water in computational GAG studies has been investigated revealing the influence of the usage of solvent models on the MD simulations of GAGs. This together highlights the relevance of my PhD studies, and the work mentioned in this PhD thesis will serve to broaden the knowledge of GAG involving systems and will facilitate the studies of other computational and experimental researchers in the field.

## Other publications in my PhD studies

1. Marcisz M., Maszota-Zieleniak M., Samsonov S.A. Repulsive Scaling Replica Exchange Molecular Dynamics in Modeling Protein-Glycosaminoglycan Complexes. *Nikos K. Karamanos (ed.), Proteoglycans: Methods and Protocols, Methods in Molecular Biology*. 2023. 2619. doi: 10.1007/978-1-0716-2946-8\_12.
2. 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. doi: 10.1039/d0cp05438k. Epub 2021 Jan 29. PMID: 33514968.
3. Antoniuk A, Biskupek I, Bojarski KK, Czaplewski C, Giełdoń A, Kogut M, Kogut MM, Krupa P, Lipska AG, Liwo A, Lubecka EA, Marcisz M, Maszota-Zieleniak M, Samsonov SA, Sieradzan AK, Ślusarz MJ, Ślusarz R, Wesołowski PA, Zięba K. Modeling protein structures with the coarse-grained UNRES force field in the CASP14 experiment. *J Mol Graph Model*. 2021. 108:108008. doi: 10.1016/j.jmglm.2021.108008. Epub 2021 Aug 17. PMID: 34419932.
4. Czajkowski, R., Marcisz, M. & Bartnik, P. Fast and reliable screening assay developed to preselect candidate Soft Rot Pectobacteriaceae Tn5 mutants showing resistance to bacteriophage infection. *Eur J Plant Pathol* 2019. 155:671–676. <https://doi.org/10.1007/s10658-019-01786-z>.

## List of abbreviations

Abbreviations listed in alphabetic order:

AA	All atom
ACE	Affinity capillary electrophoresis
AD3	Autodock3
APRIL	A proliferation-inducing ligand
ATIII	Antithrombin III
BAR	Bennett Acceptance Ratio
BCMA	B-cell maturation antigen
bFGF	Basic fibroblast growth factor
BLI	Bio-layer interferometry
BMP-2	Bone morphogenetic protein 2
CD	Cluster of differentiation
CD	Circular dichroism
CE	Capillary electrophoresis
CEC	Capillary electrokinetic chromatography
CG	Coarse-grained
CryoEM	Cryogenic electron microscopy
CSc	Chondroitin sulfate C
CSe	Chondroitin sulfate E
CSP	Chemical shift perturbation
CXCL5	C-X-C motif chemokine ligand 5
CZE	Capillary zone electrophoresis
DMD	Dynamic Molecular Docking
dp	Degree of polymerization
DS	Dermatan sulfate
ECM	Extracellular matrix
ECP	Eosinophil cationic protein
EM	Electron microscopy
FACCE	Frontal analysis continuous capillary electrophoresis
FEP	Free energy perturbation

FF	Force- field
FGF	Fibroblast growth Factor
GAG/GAGs	Glycosaminoglycan/s
Gal	Galactose
GalNAc	N-acetylgalactosamine
GB	Generalized Born
GDNF	Glial-cell-line-derived neurotrophic factor
GEC	Gel electrophoresis
GF	Growth Factor
GlcA	Glucuronic acid
GlcNAc	N-acetylglucosamine
GlcNS	N-sulfoglucosamine
GPI	Glycosylphosphatidylinositol
HA	Hyaluronic acid
HMW	High-molecular weight
HP	Heparin
HPLC	High-performance liquid chromatography
HS	Heparan sulfate
HSPG	Heparan sulfate proteoglycan
IdoA	Iduronic acid
IL-8	Interleukin-8
iNOS	Nitric oxide synthase
ITC	Isothermal titration calorimetry
KS	Keratan sulfate
LIE	Linear interaction energy
LMW	Low-molecular weight
LRA	Linear response approximation
MC	Monte Carlo
MD	Molecular Dynamics
MM/GBSA	Molecular mechanics/generalized Born surface area
MM/PBSA	Molecular mechanics/Poisson-Boltzmann surface area
MPP	Matrix metalloproteinase



MPXV	Monkeypox virus
NM	Normal mode
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
OPC	Optimal point charge (water model)
PB	Poisson-Boltzmann
PBSA	Poisson-Boltzmann surface area
PCPE-1	Procollagen C-proteinase enhancer-1
PDB	Protein Data Bank
PG	Proteoglycan
PMF	Potential of mean force
QH	Quasiharmonic
QM	Quantum mechanics
QM/MM	Quantum mechanics /molecular mechanics
RHF	Restricted Hartree-Fock
RMSatd	Root mean squared atom type deviation
RS-REMD	Repulsive Scaling Replica Exchange Molecular Dynamics
SASA	Solvent accessible surface area
SEM	Scanning electron microscopy
SNFG	Symbol nomenclature for glycans
SPC	Simple point-charge
SPC/E	Extended simple point charge
SPR	Surface plasmon resonance
TACI	Transmembrane activator and CAML interactor
TEM	Transmission electron miscopy
TI	Thermodynamic integration
TIMP-3	Tissue inhibitor of metalloproteinase 3
TIP	Transferable intermolecular potential (water model)
US	Umbrella sampling
VEGF	Vascular endothelial growth factor
VUVCD	Vacuum-ultraviolet circular dichroism
WaterLOGSY	Water-ligand observed via gradient spectroscopy

## List of Figures and Tables

### Figures:

Figure 1. Particular GAG types (represented in SNFG representation), their composition and sulfation patterns .....	10
Figure 2. The gauche-trans (gt), trans-gauche (tg) and gauche-gauche (gg) rotamers of methyl $\alpha$ -d-glucopyranoside.....	41
Figure 3. Three types of most commonly used atomistic water models in MD simulations. ..	48
Figure D1. Comparison of positive electrostatic potential isosurfaces .....	54
Figure D2. MM/GBSA binding free energies in the protein-GAG complex 2AXM .....	56
Figure D3. Graphical representation of mixed AA-CG GAG.....	58
Figure D4. Residues with the highest contribution to GAG binding .....	60
Figure D5. Predicted number of APRIL-BCMA-HP and APRIL-TACI contacts.....	62

### Tables:

Table 1. Characteristics of GAG molecules.....	11
Table 2. Charge and structural properties of the selected most used water models.....	48
Table D1. HP descriptors obtained from 5 $\mu$ s MD simulations. ....	64

## References

- (1) Vallet, S. D.; Clerc, O.; Ricard-Blum, S. Glycosaminoglycan–Protein Interactions: The First Draft of the Glycosaminoglycan Interactome. *Journal of Histochemistry and Cytochemistry*. 2020. <https://doi.org/10.1369/0022155420946403>.
- (2) Moustakas, A.; Souchelnyskiy, S.; Heldin, C. H. Smad Regulation in TGF- $\beta$  Signal Transduction. *J. Cell Sci*. 2001.
- (3) Paganini, C.; Costantini, R.; Superti-Furga, A.; Rossi, A. Bone and Connective Tissue Disorders Caused by Defects in Glycosaminoglycan Biosynthesis: A Panoramic View. *FEBS Journal*. 2019. <https://doi.org/10.1111/febs.14984>.
- (4) Salbach, J.; Rachner, T. D.; Rauner, M.; Hempel, U.; Anderegg, U.; Franz, S.; Simon, J. C.; Hofbauer, L. C. Regenerative Potential of Glycosaminoglycans for Skin and Bone. *Journal of Molecular Medicine*. 2012. <https://doi.org/10.1007/s00109-011-0843-2>.
- (5) Kwok, J. C. F.; Afshari, F.; García-Álías, G.; Fawcett, J. W. Proteoglycans in the Central Nervous System: Plasticity, Regeneration and Their Stimulation with Chondroitinase ABC. *Restorative Neurology and Neuroscience*. 2008.
- (6) Galtrey, C. M.; Kwok, J. C. F.; Carulli, D.; Rhodes, K. E.; Fawcett, J. W. Distribution and Synthesis of Extracellular Matrix Proteoglycans, Hyaluronan, Link Proteins and Tenascin-R in the Rat Spinal Cord. *Eur. J. Neurosci*. 2008. <https://doi.org/10.1111/j.1460-9568.2008.06108.x>.
- (7) Kwok, J. C. F.; Warren, P.; Fawcett, J. W. Chondroitin Sulfate: A Key Molecule in the Brain Matrix. *International Journal of Biochemistry and Cell Biology*. 2012. <https://doi.org/10.1016/j.biocel.2012.01.004>.
- (8) Marcisz, M.; Maszota-Zieleniak, M.; Huard, B.; Samsonov, S. A. Advanced Molecular Dynamics Approaches to Model a Tertiary Complex April/Taci with Long Glycosaminoglycans. *Biomolecules* 2021. <https://doi.org/10.3390/biom11091349>.
- (9) Fenton, S. E.; Dentine, M. R.; Ax, R. L. Modulation of Bovine Oocyte-Cumulus Cell Complex Maturation and Fertilization In Vitro by Glycosaminoglycans. *J. Dairy Sci*. 1993. [https://doi.org/10.3168/jds.S0022-0302\(93\)77393-2](https://doi.org/10.3168/jds.S0022-0302(93)77393-2).
- (10) 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. Proteoglycan Chemical Diversity Drives Multifunctional Cell Regulation and Therapeutics. *Chemical Reviews*. 2018. <https://doi.org/10.1021/acs.chemrev.8b00354>.
- (11) Risau, W. Mechanisms of Angiogenesis. *Nature*. 1997. <https://doi.org/10.1038/386671a0>.
- (12) Ma, S. N.; Mao, Z. X.; Wu, Y.; Liang, M. X.; Wang, D. D.; Chen, X.; Chang, P. an; Zhang, W.; Tang, J. H. The Anti-Cancer Properties of Heparin and Its Derivatives: A Review and Prospect. *Cell Adhesion and Migration*. 2020. <https://doi.org/10.1080/19336918.2020.1767489>.
- (13) Morla, S. Glycosaminoglycans and Glycosaminoglycan Mimetics in Cancer and Inflammation. *International Journal of Molecular Sciences*. 2019. <https://doi.org/10.3390/ijms20081963>.
- (14) Shi, D.; Sheng, A.; Chi, L. Glycosaminoglycan-Protein Interactions and Their Roles in Human Disease. *Frontiers in Molecular Biosciences*. 2021. <https://doi.org/10.3389/fmolb.2021.639666>.
- (15) Wight, T. N. A Role for Proteoglycans in Vascular Disease. *Matrix Biology*. 2018. <https://doi.org/10.1016/j.matbio.2018.02.019>.
- (16) Huynh, M. B.; Ouidja, M. O.; Chantepie, S.; Carpentier, G.; Maïza, A.; Zhang, G.; Vilares, J.; Raisman-Vozari, R.; Papy-Garcia, D. Glycosaminoglycans from Alzheimer’s Disease Hippocampus Have Altered Capacities to Bind and Regulate Growth Factors Activities and to Bind Tau. *PLoS One* 2019. <https://doi.org/10.1371/journal.pone.0209573>.
- (17) Paiardi, G.; Richter, S.; Oreste, P.; Urbinati, C.; Rusnati, M.; Wade, R. C. The Binding of Heparin to Spike Glycoprotein Inhibits SARS-CoV-2 Infection by Three Mechanisms. *J. Biol. Chem*. 2022. <https://doi.org/10.1016/j.jbc.2021.101507>.
- (18) Clausen, T. M.; Sandoval, D. R.; Spliid, C. B.; Pihl, J.; Perrett, H. R.; Painter, C. D.; Narayanan, A.; Majowicz, S. A.; Kwong, E. M.; McVicar, R. N.; Thacker, B. E.; Glass, C. A.; Yang, Z.; Torres, J. L.; Golden, G. J.; Bartels, P. L.; Porell, R. N.; Garretson, A. F.; Laubach, L.; Feldman, J.; Yin, X.; Pu, Y.; Hauser, B. M.; Caradonna, T. M.; Kellman, B. P.; Martino, C.; Gordts, P. L. S. M.; Chanda, S. K.; Schmidt, A. G.; Godula, K.; Leibel, S. L.; Jose, J.; Corbett, K. D.; Ward, A. B.; Carlin, A. F.; Esko, J. D. SARS-CoV-2 Infection Depends on Cellular Heparan Sulfate and ACE2. *Cell* 2020, 183 (4), 1043-1057.e15. <https://doi.org/10.1016/J.CELL.2020.09.033>.
- (19) Kim, S. Y.; Jin, W.; Sood, A.; Montgomery, D. W.; Grant, O. C.; Fuster, M. M.; Fu, L.; Dordick, J. S.; Woods, R. J.; Zhang, F.; Linhardt, R. J. Characterization of Heparin and Severe Acute Respiratory Syndrome-Related Coronavirus 2 (SARS-CoV-2) Spike Glycoprotein Binding Interactions. *Antiviral Res*. 2020. <https://doi.org/10.1016/j.antiviral.2020.104873>.
- (20) Liu, L.; Chopra, P.; Li, X.; Bouwman, K. M.; Tompkins, S. M.; Wolfert, M. A.; De Vries, R. P.; Boons, G. J. Heparan Sulfate Proteoglycans as Attachment Factor for SARS-CoV-2. *ACS Cent. Sci*. 2021. <https://doi.org/10.1021/acscentsci.1c00010>.
- (21) Baylon, E. G.; Levenston, M. E. Osmotic Swelling Responses Are Conserved Across Cartilaginous Tissues With Varied Sulfated-Glycosaminoglycan Contents. *J. Orthop. Res*. 2020. <https://doi.org/10.1002/jor.24521>.
- (22) Uciechowska-Kaczmarzyk, U.; Babik, S.; Zsila, F.; Bojarski, K. K.; Beke-Somfai, T.; Samsonov, S. A. Molecular Dynamics-Based Model of VEGF-A and Its Heparin Interactions. *J. Mol. Graph. Model*. 2018. <https://doi.org/10.1016/j.jmgm.2018.04.015>.
- (23) Faham, S.; Hileman, R. E.; Fromm, J. R.; Linhardt, R. J.; Rees, D. C. Heparin Structure and Interactions with Basic Fibroblast Growth Factor. *Science*. 1996. <https://doi.org/10.1126/science.271.5252.1116>.
- (24) DiGabriele, A. D.; Lax, I.; Chen, D. I.; Svahn, C. M.; Jaye, M.; Schlessinger, J.; Hendrickson, W. A. Structure of a Heparin-Linked Biologically Active Dimer of Fibroblast Growth Factor. *Nature* 1998. <https://doi.org/10.1038/317141>.
- (25) Derler, R.; Gesslbauer, B.; Weber, C.; Strutzmann, E.; Miller, I.; Kungl, A. Glycosaminoglycan-Mediated Downstream Signaling of CXCL8 Binding to Endothelial Cells. *Int. J. Mol. Sci*. 2017. <https://doi.org/10.3390/ijms18122605>.
- (26) Penk, A.; Baumann, L.; Huster, D.; Samsonov, S. A. NMR and Molecular Modeling Reveal Specificity of the Interactions between CXCL14 and Glycosaminoglycans. *Glycobiology* 2019. <https://doi.org/10.1093/glycob/cwz047>.
- (27) Nordsieck, K.; Baumann, L.; Hintze, V.; Pisabarro, M. T.; Schnabelrauch, M.; Beck-Sickingler, A. G.; Samsonov, S. A. The Effect of Interleukin-8 Truncations on Its Interactions with Glycosaminoglycans. *Biopolymers*. 2018. <https://doi.org/10.1002/bip.23103>.
- (28) Faye, C.; Moreau, C.; Chautard, E.; Jetne, R.; Fukai, N.; Ruggiero, F.; Humphries, M. J.; Olsen, B. R.; Ricard-Blum, S. Molecular Interplay between Endostatin, Integrins, and Heparan Sulfate. *J. Biol. Chem*. 2009. <https://doi.org/10.1074/jbc.M109.002840>.
- (29) Wigén, J.; Elovsson-Rendin, L.; Karlsson, L.; Tykesson, E.; Westergren-Thorsson, G. Glycosaminoglycans: A Link between Development and Regeneration in the Lung. *Stem Cells and Development*. 2019. <https://doi.org/10.1089/scd.2019.0009>.
- (30) Xu, D.; Esko, J. D. Demystifying Heparan Sulfate-Protein Interactions. *Annual Review of Biochemistry*. 2014. <https://doi.org/10.1146/annurev-biochem-060713-035314>.
- (31) Nagarajan, B.; Holmes, S. G.; Sankaranarayanan, N. V.; Desai, U. R. Molecular Dynamics Simulations to Understand Glycosaminoglycan Interactions in the Free- and Protein-Bound States. *Current Opinion in Structural Biology*. 2022. <https://doi.org/10.1016/j.sbi.2022.102356>.
- (32) Imberty, A.; Lortat-Jacob, H.; Pérez, S. Structural View of Glycosaminoglycan–Protein Interactions. *Carbohydr. Res*. 2007.

- <https://doi.org/10.1016/j.carres.2006.12.019>.
- (33) Petitou, M.; Casu, B.; Lindahl, U. 1976-1983, a Critical Period in the History of Heparin: The Discovery of the Antithrombin Binding Site. *Biochimie*. 2003. [https://doi.org/10.1016/S0300-9084\(03\)00078-6](https://doi.org/10.1016/S0300-9084(03)00078-6).
- (34) Sepuru, K. M.; Nagarajan, B.; Desai, U. R.; Rajarathnam, K. Structural Basis, Stoichiometry, and Thermodynamics of Binding of the Chemokines KC and MIP2 to the Glycosaminoglycan Heparin. *J. Biol. Chem.* 2018. <https://doi.org/10.1074/jbc.RA118.004866>.
- (35) Raman, R.; Sasisekharan, V.; Sasisekharan, R. Structural Insights into Biological Roles of Protein-Glycosaminoglycan Interactions. *Chemistry and Biology*. 2005. <https://doi.org/10.1016/j.chembiol.2004.11.020>.
- (36) Kogut, M. M.; Marcisz, M.; Samsonov, S. A. Modeling Glycosaminoglycan-Protein Complexes. *Current Opinion in Structural Biology*. 2022. <https://doi.org/10.1016/j.sbi.2022.102332>.
- (37) Bojarski, K. K.; Samsonov, S. A. Role of Oligosaccharide Chain Polarity in Protein-Glycosaminoglycan Interactions. *J. Chem. Inf. Model.* 2021. <https://doi.org/10.1021/acs.jcim.0c01402>.
- (38) Joseph, P. R. B.; Mosier, P. D.; Desai, U. R.; 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. <https://doi.org/10.1042/BJ20150059>.
- (39) Rother, S.; Samsonov, S. A.; Hofmann, T.; Blaszkiewicz, J.; Köhling, S.; Moeller, S.; Schnabelrauch, M.; Rademann, J.; Kalkhof, S.; von Bergen, M.; Pisabarro, M. T.; Scharnweber, D.; Hintze, V. 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. <https://doi.org/10.1016/j.actbio.2016.08.030>.
- (40) Varki, A. C. R. D. E. J. D. F. H. H. S. P. B. C. R. H. G. W. E. M.; E. *Essentials of Glycobiology, 3rd Edition*; 2015.
- (41) Sasarman, F.; Maftai, C.; Campeau, P. M.; Brunel-Guitton, C.; Mitchell, G. A.; Allard, P. Biosynthesis of Glycosaminoglycans: Associated Disorders and Biochemical Tests. *Journal of Inherited Metabolic Disease*. 2016. <https://doi.org/10.1007/s10545-015-9903-z>.
- (42) Habuchi, H.; Habuchi, O.; Kimata, K. Sulfation Pattern in Glycosaminoglycan: {Does} It Have a Code? *Glycoconj. J.* 2004. <https://doi.org/10.1023/B:GLYC.0000043747.87325.5e>.
- (43) Bu, C.; Jin, L. NMR Characterization of the Interactions Between Glycosaminoglycans and Proteins. *Frontiers in Molecular Biosciences*. 2021. <https://doi.org/10.3389/fmolb.2021.646808>.
- (44) Clerc, O.; Mariethoz, J.; Rivet, A.; Lisacek, F.; PCrossed Signrez, S. D.; Ricard-Blum, S. A Pipeline to Translate Glycosaminoglycan Sequences into 3D Models. Application to the Exploration of Glycosaminoglycan Conformational Space. *Glycobiology* 2019. <https://doi.org/10.1093/glycob/cwy084>.
- (45) Varki, A.; Cummings, R. D.; Aebi, M.; Packer, N. H.; Seeberger, P. H.; Esko, J. D.; Stanley, P.; Hart, G.; Darvill, A.; Kinoshita, T.; Prestegard, J. J.; Schnaar, R. L.; Freeze, H. H.; Marth, J. D.; Bertozzi, C. R.; Etzler, M. E.; Frank, M.; Vliegthart, J. F. G.; Lütteke, T.; Perez, S.; Bolton, E.; Paulson, J.; Kanehisa, M.; Toukach, P.; Aoki-Kinoshita, K. F.; Dell, A.; Narimatsu, H.; York, W.; Taniguchi, N.; Kornfeld, S. Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology* 2015, 25 (12), 1323–1324. <https://doi.org/10.1093/glycob/cwv091>.
- (46) Rao, R. M.; Dauchez, M.; Baud, S. How Molecular Modelling Can Better Broaden the Understanding of Glycosylations. *Current Opinion in Structural Biology*. 2022. <https://doi.org/10.1016/j.sbi.2022.102393>.
- (47) Cylwik, B.; Lipartowska, K.; Chrostek, L.; Gruszevska, E. Congenital Disorders of Glycosylation. Part II. Defects of Protein o-Glycosylation. *Acta Biochimica Polonica*. 2013. [https://doi.org/10.18388/abp.2013\\_1993](https://doi.org/10.18388/abp.2013_1993).
- (48) Karousou, E.; Misra, S.; Ghatak, S.; Dobra, K.; Götte, M.; Vigezzi, D.; Passi, A.; Karamanos, N. K.; Skandalis, S. S. Roles and Targeting of the HAS/Hyaluronan/CD44 Molecular System in Cancer. *Matrix Biology*. 2017. <https://doi.org/10.1016/j.matbio.2016.10.001>.
- (49) Johnson, P.; Arif, A. A.; Lee-Sayer, S. S. M.; Dong, Y. Hyaluronan and Its Interactions with Immune Cells in the Healthy and Inflamed Lung. *Frontiers in Immunology*. 2018. <https://doi.org/10.3389/fimmu.2018.02787>.
- (50) Garantziotis, S.; Savani, R. C. Hyaluronan Biology: A Complex Balancing Act of Structure, Function, Location and Context. *Matrix Biology*. 2019. <https://doi.org/10.1016/j.matbio.2019.02.002>.
- (51) Huang, Y.; Askew, E. B.; Knudson, C. B.; Knudson, W. CRISPR/Cas9 Knockout of HAS2 in Rat Chondrosarcoma Chondrocytes Demonstrates the Requirement of Hyaluronan for Aggrecan Retention. *Matrix Biol.* 2016. <https://doi.org/10.1016/j.matbio.2016.04.002>.
- (52) Nagyova, E. The Biological Role of Hyaluronan-Rich Oocyte-Cumulus Extracellular Matrix in Female Reproduction. *International Journal of Molecular Sciences*. 2018. <https://doi.org/10.3390/ijms19010183>.
- (53) Petrey, A. C.; Obery, D. R.; Kessler, S. P.; Flamion, B.; de la Motte, C. A. Hyaluronan Depolymerization by Megakaryocyte Hyaluronidase-2 Is Required for Thrombopoiesis. *Am. J. Pathol.* 2016. <https://doi.org/10.1016/j.ajpath.2016.05.004>.
- (54) Lee, G. M.; Johnstone, B.; Jacobson, K.; Caterson, B. The Dynamic Structure of the Pericellular Matrix on Living Cells. *J. Cell Biol.* 1993. <https://doi.org/10.1083/jcb.123.6.1899>.
- (55) Foley, J. P.; Lam, D.; Jiang, H.; Liao, J.; Cheong, N.; McDevitt, T. M.; Zaman, A.; Wright, J. R.; Savani, R. C. Toll-like Receptor 2 (TLR2), Transforming Growth Factor- $\beta$ , Hyaluronan (HA), and Receptor for HA-Mediated Motility (RHAMM) Are Required for Surfactant Protein  $\alpha$ -Stimulated Macrophage Chemotaxis. *J. Biol. Chem.* 2012. <https://doi.org/10.1074/jbc.M112.360982>.
- (56) Tamoto, K.; Nochi, H.; Tada, M.; Shimada, S.; Mori, Y.; Kataoka, S.; Suzuk, Y.; Nakamura, T. High-Molecular-Weight Hyaluronic Acids Inhibit Chemotaxis and Phagocytosis but Not Lysosomal Enzyme Release Induced by Receptor-Mediated Stimulations in Guinea Pig Phagocytes. *Microbiol. Immunol.* 1994. <https://doi.org/10.1111/j.1348-0421.1994.tb01746.x>.
- (57) Akatsuka, M.; Yamamoto, Y.; Tobetto, K.; Yasui, T.; Ando, T. Suppressive Effects of Hyaluronic Acid on Elastase Release from Rat Peritoneal Leucocytes. *J. Pharm. Pharmacol.* 1993. <https://doi.org/10.1111/j.2042-7158.1993.tb03693.x>.
- (58) Suzuki, Y.; Yamaguchi, T. Effects of Hyaluronic Acid on Macrophage Phagocytosis and Active Oxygen Release. *Agents Actions* 1993. <https://doi.org/10.1007/BF02027210>.
- (59) Østerholt, H. C. D.; Dannevig, I.; Wyckoff, M. H.; Liao, J.; Akgul, Y.; Ramgopal, M.; Mija, D. S.; Cheong, N.; Longoria, C.; Mahendroo, M.; Nakstad, B.; Saugstad, O. D.; Savani, R. C. Antioxidant Protects against Increases in Low Molecular Weight Hyaluronan and Inflammation in Asphyxiated Newborn Pigs Resuscitated with 100% Oxygen. *PLoS One* 2012. <https://doi.org/10.1371/journal.pone.0038839>.
- (60) McKee, C. M.; Penno, M. B.; Cowman, M.; Burdick, M. D.; Strieter, R. M.; Bao, C.; Noble, P. W. Hyaluronan (HA) Fragments Induce Chemokine Gene Expression in Alveolar Macrophages: The Role of HA Size and CD44. *J. Clin. Invest.* 1996. <https://doi.org/10.1172/JCI119054>.
- (61) Waldenström, A.; Martinussen, H. J.; Gerdin, B.; Hällgren, R. Accumulation of Hyaluronan and Tissue Edema in Experimental Myocardial Infarction. *J. Clin. Invest.* 1991. <https://doi.org/10.1172/JCI115475>.
- (62) Wells, A. F.; Klareskog, L.; Lindblad, S.; Laurent, T. C. Correlation between Increased Hyaluronan Localized in Arthritic Synovium and the Presence of Proliferating Cells. A Role for Macrophage-derived Factors. *Arthritis Rheum.* 1992.

- <https://doi.org/10.1002/art.1780350405>.
- (63) Wells, A. F.; Larsson, E.; Tengblad, A.; Fellström, B.; Tufveson, G.; Klareskog, L.; Laurent, T. C. The Localization of Hyaluronan in Normal and Rejected Human Kidneys. *Transplantation* 1990. <https://doi.org/10.1097/00007890-199008000-00014>.
- (64) Schiller, S.; Mathews, M. B.; Cifonelli, J. A.; Dorfman, A. THE METABOLISM OF MUCOPOLYSACCHARIDES IN ANIMALS. *J. Biol. Chem.* 1956. [https://doi.org/10.1016/s0021-9258\(18\)65879-x](https://doi.org/10.1016/s0021-9258(18)65879-x).
- (65) Triggs-Raine, B. Biology of Hyaluronan: Insights from Genetic Disorders of Hyaluronan Metabolism. *World J. Biol. Chem.* 2015. <https://doi.org/10.4331/wjbc.v6.i3.110>.
- (66) Balazs, E. A.; Denlinger, J. L. Clinical Uses of Hyaluronan. *Ciba Foundation symposium.* 1989. <https://doi.org/10.1002/9780470513774.ch16>.
- (67) Price, R. D.; Myers, S.; Leigh, I. M.; Navsaria, H. A. The Role of Hyaluronic Acid in Wound Healing. *Am. J. Clin. Dermatol.* 2005. <https://doi.org/10.2165/00128071-200506060-00006>.
- (68) Gaffney, J.; Matou-Nasri, S.; Grau-Olivares, M.; Slevin, M. Therapeutic Applications of Hyaluronan. *Molecular BioSystems.* 2010. <https://doi.org/10.1039/b910552m>.
- (69) Sugahara, K.; Kitagawa, H. Recent Advances in the Study of the Biosynthesis and Functions of Sulfated Glycosaminoglycans. *Current Opinion in Structural Biology.* 2000. [https://doi.org/10.1016/S0959-440X\(00\)00125-1](https://doi.org/10.1016/S0959-440X(00)00125-1).
- (70) Caterson, B. Fell-Muir Lecture: Chondroitin Sulphate Glycosaminoglycans: Fun for Some and Confusion for Others. *International Journal of Experimental Pathology.* 2012. <https://doi.org/10.1111/j.1365-2613.2011.00807.x>.
- (71) Gallagher, J. Fell-Muir Lecture: Heparan Sulphate and the Art of Cell Regulation: A Polymer Chain Conducts the Protein Orchestra. *Int. J. Exp. Pathol.* 2015. <https://doi.org/10.1111/iep.12135>.
- (72) Lindahl, U.; Li, J. ping. Chapter 3 Interactions Between Heparan Sulfate and Proteins-Design and Functional Implications. *International Review of Cell and Molecular Biology.* 2009. [https://doi.org/10.1016/S1937-6448\(09\)76003-4](https://doi.org/10.1016/S1937-6448(09)76003-4).
- (73) Dhoot, G. K.; Gustafsson, M. K.; Ai, X.; Sun, W.; Standiford, D. M.; Emerson, J. Regulation of Wnt Signaling and Embryo Patterning by an Extracellular Sulfatase. *Science.* 2001. <https://doi.org/10.1126/science.293.5535.1663>.
- (74) Frese, M. A.; Milz, F.; Dick, M.; Lamanna, W. C.; Dierks, T. Characterization of the Human Sulfatase Sulfl and Its High Affinity Heparin/Heparan Sulfate Interaction Domain. *J. Biol. Chem.* 2009. <https://doi.org/10.1074/jbc.M109.035808>.
- (75) Capila, I.; Linhardt, R. J. Heparin - Protein Interactions. *Angewandte Chemie - International Edition.* 2002. [https://doi.org/10.1002/1521-3773\(20020201\)41:3<390::AID-ANIE390>3.0.CO;2-B](https://doi.org/10.1002/1521-3773(20020201)41:3<390::AID-ANIE390>3.0.CO;2-B).
- (76) Ori, A.; Wilkinson, M. C.; Fernig, D. G. The Heparanome and Regulation of Cell Function: Structures, Functions and Challenges. *Frontiers in Bioscience.* 2008. <https://doi.org/10.2741/3007>.
- (77) Mulloy, B.; Linhardt, R. J. Order out of Complexity - Protein Structures That Interact with Heparin. *Current Opinion in Structural Biology.* 2001. [https://doi.org/10.1016/S0959-440X\(00\)00257-8](https://doi.org/10.1016/S0959-440X(00)00257-8).
- (78) Rapraeger, A. C.; Krufka, A.; Olwin, B. B. Requirement of Heparan Sulfate for BFGF-Mediated Fibroblast Growth and Myoblast Differentiation. *Science.* 1991. <https://doi.org/10.1126/science.1646484>.
- (79) Yayon, A.; Klagsbrun, M.; Esko, J. D.; Leder, P.; Ornitz, D. M. Cell Surface, Heparin-like Molecules Are Required for Binding of Basic Fibroblast Growth Factor to Its High Affinity Receptor. *Cell* 1991. [https://doi.org/10.1016/0092-8674\(91\)90512-W](https://doi.org/10.1016/0092-8674(91)90512-W).
- (80) Ramani, V. C.; Yang, Y.; Ren, Y.; Nan, L.; Sanderson, R. D. Heparanase Plays a Dual Role in Driving Hepatocyte Growth Factor (HGF) Signaling by Enhancing HGF Expression and Activity. *J. Biol. Chem.* 2011. <https://doi.org/10.1074/jbc.M110.183277>.
- (81) Asai, T.; Watanabe, K.; Ichihara-Tanaka, K.; Kaneda, N.; Kojima, S.; Iguchi, A.; Inagaki, F.; Muramatsu, T. Identification of Heparin-Binding Sites in Midkine and Their Role in Neurite-Promotion. *Biochem. Biophys. Res. Commun.* 1997. <https://doi.org/10.1006/bbrc.1997.6905>.
- (82) Zou, P.; Zou, K.; Muramatsu, H.; Ichihara-Tanaka, K.; Habuchi, O.; Ohtake, S.; Ikematsu, S.; Sakuma, S.; Muramatsu, T. Glycosaminoglycan Structures Required for Strong Binding to Midkine, a Heparin-Binding Growth Factor. *Glycobiology* 2003. <https://doi.org/10.1093/glycob/cwg001>.
- (83) Proudfoot, A. E. I.; Handel, T. M.; Johnson, Z.; Lau, E. K.; LiWang, P.; Clark-Lewis, I.; Borlat, F.; Wells, T. N. C.; Kosco-Vilbois, M. H. Glycosaminoglycan Binding and Oligomerization Are Essential for the in Vivo Activity of Certain Chemokines. *Proc. Natl. Acad. Sci. U. S. A.* 2003. <https://doi.org/10.1073/pnas.0334864100>.
- (84) Lortat-Jacob, H.; Grosdidier, A.; Imberty, A. Structural Diversity of Heparan Sulfate Binding Domains in Chemokines. *Proc. Natl. Acad. Sci. U. S. A.* 2002. <https://doi.org/10.1073/pnas.032497699>.
- (85) Sadir, R.; Baleux, F.; Grosdidier, A.; Imberty, A.; Lortat-Jacob, H. Characterization of the Stromal Cell-Derived Factor-1 $\alpha$ -Heparin Complex. *J. Biol. Chem.* 2001. <https://doi.org/10.1074/jbc.M008110200>.
- (86) Lau, E. K.; Paavola, C. D.; Johnson, Z.; Gaudry, J. P.; Geretti, E.; Borlat, F.; Kungl, A. J.; Proudfoot, A. E.; Handel, T. M. Identification of the Glycosaminoglycan Binding Site of the CC Chemokine, MCP-1: Implications for Structure and Function in Vivo. *J. Biol. Chem.* 2004. <https://doi.org/10.1074/jbc.M311224200>.
- (87) Yu, Y.; Sweeney, M. D.; Saad, O. M.; Crown, S. E.; Handel, T. M.; Leary, J. A. Chemokine-Glycosaminoglycan Binding: Specificity for CCR2 Ligand Binding to Highly Sulfated Oligosaccharides Using FTICR Mass Spectrometry. *J. Biol. Chem.* 2005. <https://doi.org/10.1074/jbc.M505738200>.
- (88) Perez, S.; Makshakova, O.; Angulo, J.; Bedini, E.; Bisio, A.; de Paz, J. L.; Fadda, E.; Guerrini, M.; Hricovini, M.; Hricovini, M.; Lisacek, F.; Nieto, P. M.; Pagel, K.; Pairardi, G.; Richter, R.; Samsonov, S. A.; Vivès, R. A.; Nikitovic, D.; Ricard Blum, S. Glycosaminoglycans: What Remains To Be Deciphered? *JACS Au* 2022. <https://doi.org/10.1021/jacsau.2c00569>.
- (89) Mulloy, B.; Lever, R.; Page, C. P. Mast Cell Glycosaminoglycans. *Glycoconjugate Journal.* 2017. <https://doi.org/10.1007/s10719-016-9749-0>.
- (90) Ribatti, D. The Staining of Mast Cells: A Historical Overview. *International Archives of Allergy and Immunology.* 2018. <https://doi.org/10.1159/000487538>.
- (91) Herrera-Heredia, S. A.; Hsu, H. P.; Kao, C. Y.; Tsai, Y. H.; Yamaguchi, Y.; Roers, A.; Hsu, C. L.; Dzhalgalov, I. L. Heparin Is Required for the Formation of Granules in Connective Tissue Mast Cells. *Front. Immunol.* 2022. <https://doi.org/10.3389/fimmu.2022.1000405>.
- (92) Mulloy, B. The Non-Anticoagulant Promise of Heparin and Its Mimetics. *Current Opinion in Pharmacology.* 2019. <https://doi.org/10.1016/j.coph.2019.03.009>.
- (93) Beurskens, D. M. H.; Huckriede, J. P.; Schrijver, R.; Hemker, H. C.; Reutelingsperger, C. P.; Nicolaes, G. A. F. The Anticoagulant and Nonanticoagulant Properties of Heparin. *Thromb. Haemost.* 2020. <https://doi.org/10.1055/s-0040-1715460>.
- (94) Nahain, A. Al; Ignjatovic, V.; Monagle, P.; Tsanaktisidis, J.; Vamvounis, G.; Ferro, V. Anticoagulant Heparin Mimetics via RAFT Polymerization. *Biomacromolecules* 2020. <https://doi.org/10.1021/acs.biomac.9b01688>.
- (95) Altinkaya, E.; Aktas, A. Insulin and Heparin Therapies in Acute Pancreatitis Due to Hypertriglyceridemia. *Journal of the College of Physicians and Surgeons Pakistan.* 2021. <https://doi.org/10.29271/jcsp.2021.11.1337>.
- (96) Journey, A.; Alshehri, M.; Mahendra, A.; Anteet, M.; Yousef, M. A.; Khan, A. M. Therapeutic Approaches in Hypertriglyceridemia-Induced Acute Pancreatitis: A Literature Review of Available Therapies and Case Series. *J. Clin. Apher.* 2020.

- <https://doi.org/10.1002/jca.21763>.
- (97) Atallah, J.; Khachfe, H. H.; Berro, J.; Assi, H. I. The Use of Heparin and Heparin-like Molecules in Cancer Treatment: A Review. *Cancer Treatment and Research Communications*. 2020. <https://doi.org/10.1016/j.ctarc.2020.100192>.
- (98) Litov, L.; Petkov, P.; Rangelov, M.; Ilieva, N.; Lilkova, E.; Todorova, N.; Krachmarova, E.; Malinova, K.; Gospodinov, A.; Hristova, R.; Ivanov, I.; Nacheva, G. Molecular Mechanism of the Anti-Inflammatory Action of Heparin. *Int. J. Mol. Sci.* 2021. <https://doi.org/10.3390/ijms221910730>.
- (99) Fu, S.; Yu, S.; Wang, L.; Ma, X.; Li, X. Unfractionated Heparin Improves the Clinical Efficacy in Adult Sepsis Patients: A Systematic Review and Meta-Analysis. *BMC Anesthesiol.* 2022. <https://doi.org/10.1186/s12871-021-01545-w>.
- (100) Myint, K. M.; Yamamoto, Y.; Doi, T.; Kato, I.; Harashima, A.; Yonekura, H.; Watanabe, T.; Shinohara, H.; Takeuchi, M.; Tsuneyama, K.; Hashimoto, N.; Asano, M.; Takasawa, S.; Okamoto, H.; Yamamoto, H. RAGE Control of Diabetic Nephropathy in a Mouse Model: Effects of RAGE Gene Disruption and Administration of Low-Molecular Weight Heparin. *Diabetes* 2006. <https://doi.org/10.2337/db06-0221>.
- (101) Mycroft-West, C. J.; Devlin, A. J.; Cooper, L. C.; Guimond, S. E.; Procter, P.; Guerrini, M.; Miller, G. J.; Fernig, D. G.; Yates, E. A.; Lima, M. A.; Skidmore, M. A. Glycosaminoglycans from *Litopenaeus Vannamei* Inhibit the Alzheimer's Disease  $\beta$  Secretase, BACE1. *Mar. Drugs* 2021. <https://doi.org/10.3390/MD19040203>.
- (102) Stopschinski, B. E.; Thomas, T. L.; Nadji, S.; Darvish, E.; Fan, L.; Holmes, B. B.; Modi, A. R.; Finnell, J. G.; Kashmer, O. M.; Estill-Terpack, S.; Mirbaha, H.; Luu, H. S.; Diamond, M. I. A Synthetic Heparinoid Blocks Tau Aggregate Cell Uptake and Amplification. *J. Biol. Chem.* 2020. <https://doi.org/10.1074/jbc.RA119.010353>.
- (103) Clausen, T. M.; Sandoval, D. R.; Spliid, C. B.; Pihl, J.; Perrett, H. R.; Painter, C. D.; Narayanan, A.; Majowicz, S. A.; Kwong, E. M.; McVicar, R. N.; Thacker, B. E.; Glass, C. A.; Yang, Z.; Torres, J. L.; Golden, G. J.; Bartels, P. L.; Porell, R. N.; Garretson, A. F.; Laubach, L.; Feldman, J.; Yin, X.; Pu, Y.; Hauser, B. M.; Caradonna, T. M.; Kellman, B. P.; Martino, C.; Gordts, P. L. S. M.; Chanda, S. K.; Schmidt, A. G.; Godula, K.; Leibel, S. L.; Jose, J.; Corbett, K. D.; Ward, A. B.; Carlin, A. F.; Esko, J. D. SARS-CoV-2 Infection Depends on Cellular Heparan Sulfate and ACE2. *Cell* 2020. <https://doi.org/10.1016/j.cell.2020.09.033>.
- (104) van der Wal, L. I.; Kroft, L. J. M.; van Dam, L. F.; Cobbaert, C. M.; Eikenboom, J.; Huisman, M. V.; Helmerhorst, H. J. F.; Klok, F. A.; de Jonge, E. Early Effects of Unfractionated Heparin on Clinical and Radiological Signs and D-Dimer Levels in Patients with COVID-19 Associated Pulmonary Embolism: An Observational Cohort Study. *Thrombosis Research*. 2021. <https://doi.org/10.1016/j.thromres.2021.01.023>.
- (105) Hippensteel, J. A.; LaRiviere, W. B.; Colbert, J. F.; Langou t-Astri, C. J.; Schmidt, E. P. Heparin as a Therapy for COVID-19: Current Evidence and Future Possibilities. *American Journal of Physiology - Lung Cellular and Molecular Physiology*. 2020. <https://doi.org/10.1152/AJPLUNG.00199.2020>.
- (106) Kow, C. S.; Ramachandram, D. S.; Hasan, S. S. The Effect of Higher-Intensity Dosing of Anticoagulation on the Clinical Outcomes in Hospitalized Patients with COVID-19: A Meta-Analysis of Randomized Controlled Trials. *J. Infect. Chemother.* 2022. <https://doi.org/10.1016/j.jiac.2021.11.008>.
- (107) Lindahl, U.; Li, J. P. Heparin – An Old Drug with Multiple Potential Targets in Covid-19 Therapy. *Journal of Thrombosis and Haemostasis*. 2020. <https://doi.org/10.1111/jth.14898>.
- (108) Thachil, J. The Versatile Heparin in COVID-19. *Journal of Thrombosis and Haemostasis*. 2020. <https://doi.org/10.1111/jth.14821>.
- (109) Volpi, N. Chondroitin Sulfate Safety and Quality. *Molecules*. 2019. <https://doi.org/10.3390/molecules24081447>.
- (110) Sugahara, K.; Mikami, T.; Uyama, T.; Mizuguchi, S.; Nomura, K.; Kitagawa, H. Recent Advances in the Structural Biology of Chondroitin Sulfate and Dermatan Sulfate. *Current Opinion in Structural Biology*. 2003. <https://doi.org/10.1016/j.sbi.2003.09.011>.
- (111) Abourehab, M. A. S.; Baisakhiya, S.; Aggarwal, A.; Singh, A.; Abdelgawad, M. A.; Deepak, A.; Ansari, M. J.; Pramanik, S. Chondroitin Sulfate-Based Composites: A Tour d'horizon of Their Biomedical Applications. *Journal of Materials Chemistry B*. 2022. <https://doi.org/10.1039/d2tb01514e>.
- (112) Jackson, R. L.; Busch, S. J.; Cardin, A. D. Glycosaminoglycans: Molecular Properties, Protein Interactions, and Role in Physiological Processes. *Physiological Reviews*. 1991. <https://doi.org/10.1152/physrev.1991.71.2.481>.
- (113) Bruyere, O.; Reginster, J. Y. Glucosamine and Chondroitin Sulfate as Therapeutic Agents for Knee and Hip Osteoarthritis. *Drugs and Aging*. 2007. <https://doi.org/10.2165/00002512-200724070-00005>.
- (114) Zhu, X.; Sang, L.; Wu, D.; Rong, J.; Jiang, L. Effectiveness and Safety of Glucosamine and Chondroitin for the Treatment of Osteoarthritis: A Meta-Analysis of Randomized Controlled Trials. *Journal of Orthopaedic Surgery and Research*. 2018. <https://doi.org/10.1186/s13018-018-0871-5>.
- (115) Toida, T.; Sakai, S.; Akiyama, H.; Linhardt, R. J. Immunological Activity of Chondroitin Sulfate. *Advances in Pharmacology*. 2006. [https://doi.org/10.1016/S1054-3589\(05\)53019-9](https://doi.org/10.1016/S1054-3589(05)53019-9).
- (116) Vergés, J.; Montell, E.; Herrero, M.; Perna, C.; Cuevas, J.; Pérez, M.; Möller, I. Clinical and Histopathological Improvement of Psoriasis with Oral Chondroitin Sulfate: A Serendipitous Finding. *Dermatol. Online J.* 2005. <https://doi.org/10.5070/d32zh8x3vf>.
- (117) Du Souich, P.; García, A. G.; Vergés, J.; Montell, E. Immunomodulatory and Anti-Inflammatory Effects of Chondroitin Sulphate. *J. Cell. Mol. Med.* 2009. <https://doi.org/10.1111/j.1582-4934.2009.00826.x>.
- (118) Volpi, N. Anti-Inflammatory Activity of Chondroitin Sulphate: New Functions from an Old Natural Macromolecule. *Inflammopharmacology*. 2011. <https://doi.org/10.1007/s10787-011-0098-0>.
- (119) Jomphe, C.; Gabriac, M.; Hale, T. M.; Héroux, L.; Trudeau, L. É.; Deblois, D.; Montell, E.; Vergés, J.; Du Souich, P. Chondroitin Sulfate Inhibits the Nuclear Translocation of Nuclear Factor- $\kappa$ B in Interleukin-1 $\beta$ -Stimulated Chondrocytes. *Basic Clin. Pharmacol. Toxicol.* 2008. <https://doi.org/10.1111/j.1742-7843.2007.00158.x>.
- (120) Rani, A.; Patel, S.; Goyal, A. Chondroitin Sulfate (CS) Lyases: Structure, Function and Application in Therapeutics. *Curr. Protein Pept. Sci.* 2017. <https://doi.org/10.2174/1389203718666170102112805>.
- (121) Restaino, O. F.; Finamore, R.; Stellavato, A.; Diana, P.; Bedini, E.; Trifuoggi, M.; De Rosa, M.; Schiraldi, C. European Chondroitin Sulfate and Glucosamine Food Supplements: A Systematic Quality and Quantity Assessment Compared to Pharmaceuticals. *Carbohydr. Polym.* 2019. <https://doi.org/10.1016/j.carbpol.2019.114984>.
- (122) Vessella, G.; Vázquez, J. A.; Valcárcel, J.; Lagartera, L.; Monterrey, D. T.; Bastida, A.; García-junceda, E.; Bedini, E.; Fernández-mayoralas, A.; Revuelta, J. Deciphering Structural Determinants in Chondroitin Sulfate Binding to FGF-2: Paving the Way to Enhanced Predictability of Their Biological Functions. *Polymers (Basel)*. 2021. <https://doi.org/10.3390/polym13020313>.
- (123) Mizumoto, S.; Yamada, S. The Specific Role of Dermatan Sulfate as an Instructive Glycosaminoglycan in Tissue Development. *International Journal of Molecular Sciences*. 2022. <https://doi.org/10.3390/ijms23137485>.
- (124) Liaw, P. C. Y.; Becker, D. L.; Stafford, A. R.; Fredenburgh, J. C.; Weitz, J. I. Molecular Basis for the Susceptibility of Fibrin-Bound Thrombin to Inactivation by Heparin Cofactor II in the Presence of Dermatan Sulfate but Not Heparin. *J. Biol. Chem.* 2001. <https://doi.org/10.1074/jbc.M010584200>.
- (125) Shirk, R. A.; Parthasarathy, N.; Antonio, J. D. S.; Church, F. C.; Wagner, W. D. Altered Dermatan Sulfate Structure and Reduced Heparin Cofactor II- Stimulating Activity of Biglycan and Decorin from Human Atherosclerotic Plaque. *J. Biol. Chem.* 2000. <https://doi.org/10.1074/jbc.M001659200>.

- (126) Nelimarkka, L.; Salminen, H.; Kuopio, T.; Nikkari, S.; Ekfors, T.; Laine, J.; Pelliniemi, L.; Järveläinen, H. Decorin Is Produced by Capillary Endothelial Cells in Inflammation-Associated Angiogenesis. *Am. J. Pathol.* 2001. [https://doi.org/10.1016/S0002-9440\(10\)63975-2](https://doi.org/10.1016/S0002-9440(10)63975-2).
- (127) Denholm, E. M.; Lin, Y. Q.; Silver, P. J. Anti-Tumor Activities of Chondroitinase AC and Chondroitinase B: Inhibition of Angiogenesis, Proliferation and Invasion. *Eur. J. Pharmacol.* 2001. [https://doi.org/10.1016/S0014-2999\(01\)00884-6](https://doi.org/10.1016/S0014-2999(01)00884-6).
- (128) Wei, J.; Hu, M.; Huang, K.; Lin, S.; Du, H. Roles of Proteoglycans and Glycosaminoglycans in Cancer Development and Progression. *International Journal of Molecular Sciences.* 2020. <https://doi.org/10.3390/ijms211175983>.
- (129) Thelin, M. A.; Svensson, K. J.; Shi, X.; Bagher, M.; Axelsson, J.; Isinger-Ekstrand, A.; Van Kuppevelt, T. H.; Johansson, J.; Nilbert, M.; Zaia, J.; Belting, M.; Maccarana, M.; Malmström, A. Dermatan Sulfate Is Involved in the Tumorigenic Properties of Esophagus Squamous Cell Carcinoma. *Cancer Res.* 2012. <https://doi.org/10.1158/0008-5472.CAN-11-1351>.
- (130) Liao, W. C.; Liao, C. K.; Tsai, Y. H.; Tseng, T. J.; Chuang, L. C.; Lan, C. T.; Chang, H. M.; Liu, C. H. DSE Promotes Aggressive Glioma Cell Phenotypes by Enhancing HB-EGF/ERBB Signaling. *PLoS One* 2018. <https://doi.org/10.1371/journal.pone.0198364>.
- (131) Koźma, E. M.; Wisowski, G.; Latocha, M.; Kusz, D.; Olczyk, K. Complex Influence of Dermatan Sulphate on Breast Cancer Cells. *Exp. Biol. Med.* 2014. <https://doi.org/10.1177/1535370214538590>.
- (132) Seidler, D. G.; Faiyaz-Ul-Haque, M.; Hansen, U.; Yip, G. W.; Zaidi, S. H. E.; Teebi, A. S.; Kiesel, L.; Götte, M. Defective Glycosylation of Decorin and Biglycan, Altered Collagen Structure, and Abnormal Phenotype of the Skin Fibroblasts of an Ehlers-Danlos Syndrome Patient Carrying the Novel Arg270Cys Substitution in Galactosyltransferase I (B4GalT-7). *J. Mol. Med.* 2006. <https://doi.org/10.1007/s00109-006-0046-4>.
- (133) Syx, D.; Van Damme, T.; Symoens, S.; Maiburg, M. C.; van de Laar, I.; Morton, J.; Suri, M.; Del campo, M.; Hausser, I.; Hermans-Lê, T.; De Paepe, A.; Malfait, F. Genetic Heterogeneity and Clinical Variability in Musculocontractural Ehlers-Danlos Syndrome Caused by Impaired Dermatan Sulfate Biosynthesis. *Hum. Mutat.* 2015. <https://doi.org/10.1002/humu.22774>.
- (134) Wisowski, G.; Pudełko, A.; Olczyk, K.; Paul-Samojedny, M.; Koźma, E. M. Dermatan Sulfate Affects Breast Cancer Cell Function via the Induction of Necroptosis. *Cells* 2022. <https://doi.org/10.3390/cells11010173>.
- (135) Hayes, A. J.; Melrose, J. Neural Tissue Homeostasis and Repair Is Regulated via CS and DS Proteoglycan Motifs. *Frontiers in Cell and Developmental Biology.* 2021. <https://doi.org/10.3389/fcell.2021.696640>.
- (136) Mizumoto, S.; Kwok, J. C. F.; Whitelock, J. M.; Li, F.; Perris, R. Editorial: Roles of Chondroitin Sulfate and Dermatan Sulfate as Regulators for Cell and Tissue Development. *Frontiers in Cell and Developmental Biology.* 2022. <https://doi.org/10.3389/fcell.2022.941178>.
- (137) Nadafi, R.; Koning, J. J.; Veninga, H.; Stachtea, X. N.; Konijn, T.; Zwiers, A.; Malmström, A.; den Haan, J. M. M.; Mebius, R. E.; Maccarana, M.; Reijmers, R. M. Dendritic Cell Migration to Skin-Draining Lymph Nodes Is Controlled by Dermatan Sulfate and Determines Adaptive Immunity Magnitude. *Front. Immunol.* 2018. <https://doi.org/10.3389/fimmu.2018.00206>.
- (138) Plichta, J. K.; Radek, K. A. Sugar-Coating Wound Repair. *J. Burn Care Res.* 2012. <https://doi.org/10.1097/bcr.0b013e318240540a>.
- (139) Pomin, V. H.; Piquet, A. A.; Pereira, M. S.; Mourão, P. A. S. Residual Keratan Sulfate in Chondroitin Sulfate Formulations for Oral Administration. *Carbohydr. Polym.* 2012. <https://doi.org/10.1016/j.carbpol.2012.06.009>.
- (140) Choi, H. U.; Meyer, K. The Structure of Keratan Sulphates from Various Sources. *Biochem. J.* 1975. <https://doi.org/10.1042/bj1510543>.
- (141) Catterson, B.; Melrose, J. Keratan Sulfate, a Complex Glycosaminoglycan with Unique Functional Capability. *Glycobiology.* 2018. <https://doi.org/10.1093/glycob/cwy003>.
- (142) Krusius, T.; Finne, J.; Margolis, R. K.; Margolis, R. U. Identification of an O-Glycosidic Mannose-Linked Sialylated Tetrasaccharide and Keratan Sulfate Oligosaccharides in the Chondroitin Sulfate Proteoglycan of Brain. *J. Biol. Chem.* 1986. [https://doi.org/10.1016/s0021-9258\(19\)83901-7](https://doi.org/10.1016/s0021-9258(19)83901-7).
- (143) Pomin, V. H. Keratan Sulfate: An up-to-Date Review. *International Journal of Biological Macromolecules.* 2015. <https://doi.org/10.1016/j.ijbiomac.2014.08.029>.
- (144) Weyers, A.; Yang, B.; Solakyildirim, K.; Yee, V.; Li, L.; Zhang, F.; Linhardt, R. J. Isolation of Bovine Corneal Keratan Sulfate and Its Growth Factor and Morphogen Binding. *FEBS J.* 2013. <https://doi.org/10.1111/febs.12165>.
- (145) Conrad, A. H.; Zhang, Y.; Tasheva, E. S.; Conrad, G. W. Proteomic Analysis of Potential Keratan Sulfate, Chondroitin Sulfate A, and Hyaluronic Acid Molecular Interactions. *Investig. Ophthalmol. Vis. Sci.* 2010. <https://doi.org/10.1167/iovs.09-4914>.
- (146) Kiani, C.; Chen, L.; Wu, Y. J.; Yee, A. J.; Yang, B. B. Structure and Function of Aggrecan. *Cell Research.* 2002. <https://doi.org/10.1038/sj.cr.7290106>.
- (147) Hamaya, E.; Fujisawa, T.; Tamura, M. Osteoadherin Serves Roles in the Regulation of Apoptosis and Growth in MC3T3-E1 Osteoblast Cells. *Int. J. Mol. Med.* 2019. <https://doi.org/10.3892/ijmm.2019.4376>.
- (148) Sommarin, Y.; Wendel, M.; Shen, Z.; Hellman, U.; Heinegård, D. Osteoadherin, a Cell-Binding Keratan Sulfate Proteoglycan in Bone, Belongs to the Family of Leucine-Rich Repeat Proteins of the Extracellular Matrix. *J. Biol. Chem.* 1998. <https://doi.org/10.1074/jbc.273.27.16723>.
- (149) Rehn, A. P.; Cerny, R.; Sugars, R. V.; Kaukua, N.; Wendel, M. Osteoadherin Is Upregulated by Mature Osteoblasts and Enhances Their in Vitro Differentiation and Mineralization. *Calcif. Tissue Int.* 2008. <https://doi.org/10.1007/s00223-008-9138-1>.
- (150) Nikdin, H.; Olsson, M. L.; Hulthenby, K.; Sugars, R. V. Osteoadherin Accumulates in the Predentin towards the Mineralization Front in the Developing Tooth. *PLoS One* 2012. <https://doi.org/10.1371/journal.pone.0031525>.
- (151) Liu, C. Y.; Birk, D. E.; Hassell, J. R.; Kane, B.; Kao, W. W. Y. Keratan-Deficient Mice Display Alterations in Corneal Structure. *J. Biol. Chem.* 2003. <https://doi.org/10.1074/jbc.M301169200>.
- (152) Kampmann, A.; Fernández, B.; Deindl, E.; Kubin, T.; Pipp, F.; Eitenmüller, I.; Hofer, I. E.; Schaper, W.; Zimmermann, R. The Proteoglycan Osteoglycin/Mimecan Is Correlated with Arteriogenesis. *Mol. Cell. Biochem.* 2009. <https://doi.org/10.1007/s11010-008-9935-x>.
- (153) Peal, D. S.; Burns, C. G.; Macrae, C. A.; Milan, D. Chondroitin Sulfate Expression Is Required for Cardiac Atrioventricular Canal Formation. *Dev. Dyn.* 2009. <https://doi.org/10.1002/dvdy.22154>.
- (154) Jana, M.; Bandyopadhyay, S. Conformational Flexibility of a Protein-Carbohydrate Complex and the Structure and Ordering of Surrounding Water. *Phys. Chem. Chem. Phys.* 2012. <https://doi.org/10.1039/c2cp24104h>.
- (155) Nurisso, A.; Blanchard, B.; Audfray, A.; Rydner, L.; Oscarson, S.; Varrot, A.; Imberty, A. Role of Water Molecules in Structure and Energetics of Pseudomonas Aeruginosa Lectin I Interacting with Disaccharides. *J. Biol. Chem.* 2010. <https://doi.org/10.1074/jbc.M110.108340>.
- (156) Sepuru, K. M.; Nagarajan, B.; Desai, U. R.; Rajarathnam, K. Molecular Basis of Chemokine CXCL5-Glycosaminoglycan Interactions. *J. Biol. Chem.* 2016. <https://doi.org/10.1074/jbc.M116.745265>.
- (157) Shanthamurthy, C. D.; Gimeno, A.; Leviatan Ben-Arye, S.; Kumar, N. V.; Jain, P.; Padler-Karavani, V.; Jimenez-Barbero, J.; Kikkeri, R. Sulfation Code and Conformational Plasticity of 1-Ilduronic Acid Homo-Oligosaccharides Mimic the Biological Functions of Heparan Sulfate. *ACS Chem. Biol.* 2021. <https://doi.org/10.1021/acscchembio.1c00582>.
- (158) Teyra, J.; Samsonov, S. A.; Schreiber, S.; Pisabarro, M. T. SCOWLP Update: 3D Classification of Protein-Protein, -Peptide, -

- Saccharide and -Nucleic Acid Interactions, and Structure-Based Binding Inferences across Folds. *BMC Bioinformatics* 2011. <https://doi.org/10.1186/1471-2105-12-398>.
- (159) Samsonov, S. A.; Teyra, J.; Pisabarro, M. T. Docking Glycosaminoglycans to Proteins: Analysis of Solvent Inclusion. *J. Comput. Aided. Mol. Des.* 2011. <https://doi.org/10.1007/s10822-011-9433-1>.
- (160) Liu, Q.; Brady, J. W. Anisotropic Solvent Structuring in Aqueous Sugar Solutions. *J. Am. Chem. Soc.* 1996. <https://doi.org/10.1021/ja962108d>.
- (161) Pagnotta, S. E.; McLain, S. E.; Soper, A. K.; Bruni, F.; Ricci, M. A. Water and Trehalose: How Much Do They Interact with Each Other? *J. Phys. Chem. B* 2010. <https://doi.org/10.1021/jp911940h>.
- (162) Pol-Fachin, L.; Verli, H. Depiction of the Forces Participating in the 2-O-Sulfo- $\alpha$ -L-Iduronic Acid Conformational Preference in Heparin Sequences in Aqueous Solutions. *Carbohydr. Res.* 2008. <https://doi.org/10.1016/j.carres.2008.04.016>.
- (163) Ruiz Hernandez, S. E.; Streeter, I.; De Leeuw, N. H. The Effect of Water on the Binding of Glycosaminoglycan Saccharides to Hydroxyapatite Surfaces: A Molecular Dynamics Study. *Phys. Chem. Chem. Phys.* 2015. <https://doi.org/10.1039/c5cp02630j>.
- (164) Wang, X.; Xu, H.; Huang, Y.; Gu, S.; Jiang, J. X. Coupling Effect of Water and Proteoglycans on the in Situ Toughness of Bone. *J. Bone Miner. Res.* 2016. <https://doi.org/10.1002/jbmr.2774>.
- (165) Sarkar, A.; Desai, U. R. A Simple Method for Discovering Druggable, Specific Glycosaminoglycan-Protein Systems. Elucidation of Key Principles from Heparin/Heparan Sulfate-Binding Proteins. *PLoS One* 2015. <https://doi.org/10.1371/journal.pone.0141127>.
- (166) Sarkar, A.; Yu, W.; Desai, U. R.; Mackerell, A. D.; Mosier, P. D. Estimating Glycosaminoglycan-Protein Interaction Affinity: Water Dominates the Specific Antithrombin-Heparin Interaction. *Glycobiology* 2016. <https://doi.org/10.1093/glycob/cww073>.
- (167) Beldowski, P.; Mazurkiewicz, A.; Topoliński, T.; Małek, T. Hydrogen and Water Bonding between Glycosaminoglycans and Phospholipids in the Synovial Fluid: Molecular Dynamics Study. *Materials (Basel)*. 2019. <https://doi.org/10.3390/ma12132060>.
- (168) van Dam, E. P.; Giubertoni, G.; Burla, F.; Koenderink, G. H.; Bakker, H. J. Hyaluronan Biopolymers Release Water upon PH-Induced Gelation. *Phys. Chem. Chem. Phys.* 2020. <https://doi.org/10.1039/d0cp00215a>.
- (169) Samsonov, S. A.; Teyra, J.; Pisabarro, M. T. Docking Glycosaminoglycans to Proteins: Analysis of Solvent Inclusion. *J. Comput. Aided. Mol. Des.* 2011. <https://doi.org/10.1007/s10822-011-9433-1>.
- (170) Gandhi, N. S.; Mancera, R. L. Free Energy Calculations of Glycosaminoglycan-Protein Interactions. *Glycobiology* 2009. <https://doi.org/10.1093/glycob/cwp101>.
- (171) Künze, G.; Huster, D.; Samsonov, S. A. Investigation of the Structure of Regulatory Proteins Interacting with Glycosaminoglycans by Combining NMR Spectroscopy and Molecular Modeling - The Beginning of a Wonderful Friendship. *Biological Chemistry*. 2021. <https://doi.org/10.1515/hsz-2021-0119>.
- (172) Pomin, V. H.; Wang, X. Glycosaminoglycan-Protein Interactions by Nuclear Magnetic Resonance (NMR) Spectroscopy. *Molecules*. 2018. <https://doi.org/10.3390/molecules23092314>.
- (173) Hoult, D. I.; Bhakar, B. NMR Signal Reception: Virtual Photons and Coherent Spontaneous Emission. *Concepts Magn. Reson.* 1997. [https://doi.org/10.1002/\(sici\)1099-0534\(1997\)9:5<277::aid-cmr1>3.0.co;2-w](https://doi.org/10.1002/(sici)1099-0534(1997)9:5<277::aid-cmr1>3.0.co;2-w).
- (174) Taylor, D. G.; Inamdar, R.; Bushell, M. C. NMR Imaging in Theory and in Practice. *Physics in Medicine and Biology*. 1988. <https://doi.org/10.1088/0031-9155/33/6/001>.
- (175) Pomin, V. H. Biological Findings from the Recent NMR-Based Studies of Glycosaminoglycan-Protein Interactions. *Glycobiology*. 2014. <https://doi.org/10.1093/glycob/cwu065>.
- (176) Pichert, A.; Samsonov, S. A.; Theisgen, S.; Thomas, L.; Baumann, L.; Schiller, J.; Beck-Sickinger, A. G.; Huster, D.; Pisabarro, M. T. Characterization of the Interaction of Interleukin-8 with Hyaluronan, Chondroitin Sulfate, Dermatan Sulfate and Their Sulfated Derivatives by Spectroscopy and Molecular Modeling. *Glycobiology* 2012. <https://doi.org/10.1093/glycob/cwr120>.
- (177) Mulloy, B.; Forster, M. J.; Jones, C.; Davies, D. B. N.m.r. And Molecular-Modelling Studies of the Solution Conformation of Heparin. *Biochem. J.* 1993. <https://doi.org/10.1042/bj2930849>.
- (178) 1HPN 10.2210/pdb1HPN/pdb <https://www.rcsb.org/structure/1HPN>.
- (179) Sattelle, B. M.; Shakeri, J.; Roberts, I. S.; Almond, A. A 3D-Structural Model of Unsulfated Chondroitin from High-Field NMR: 4-Sulfation Has Little Effect on Backbone Conformation. *Carbohydr. Res.* 2010. <https://doi.org/10.1016/j.carres.2009.11.013>.
- (180) 2KQO <https://www.rcsb.org/structure/2KQO>. <https://doi.org/10.2210/pdb2KQO/pdb>.
- (181) Almond, A.; DeAngelis, P. L.; Blundell, C. D. Hyaluronan: The Local Solution Conformation Determined by NMR and Computer Modeling Is Close to a Contracted Left-Handed 4-Fold Helix. *J. Mol. Biol.* 2006. <https://doi.org/10.1016/j.jmb.2006.02.077>.
- (182) 2BVK <https://www.rcsb.org/structure/2BVK>. <https://doi.org/10.2210/pdb2BVK/pdb>.
- (183) Garcia-Mayoral, M. F.; Canales, A.; Diaz, D.; López-Prados, J.; Moussaoui, M.; De Paz, J. L.; Angulo, J.; Nieto, P. M.; Jiménez-Barbero, J.; Boix, E.; Bruix, M. Insights into the Glycosaminoglycan-Mediated Cytotoxic Mechanism of Eosinophil Cationic Protein Revealed by NMR. *ACS Chem. Biol.* 2013. <https://doi.org/10.1021/cb300386v>.
- (184) Sanderson, P. N.; Huckerby, T. N.; Nieduszynski, I. A. Conformational Equilibria of  $\alpha$ -L-Iduronate Residues in Disaccharides Derived from Heparin. *Biochem. J.* 1987. <https://doi.org/10.1042/bj2430175>.
- (185) Hricovini, M.; Guerrini, M.; Bisio, A.; Torri, G.; Petitou, M.; Casu, B. Conformation of Heparin Pentasaccharide Bound to Antithrombin III. *Biochem. J.* 2001. <https://doi.org/10.1042/0264-6021:3590265>.
- (186) Kunze, G.; Köhling, S.; Vogel, A.; Rademann, J.; Huster, D. Identification of the Glycosaminoglycan Binding Site of Interleukin-10 by NMR Spectroscopy. *J. Biol. Chem.* 2016. <https://doi.org/10.1074/jbc.M115.681759>.
- (187) Köhling, S.; Künze, G.; Lemmnitzer, K.; Bermudez, M.; Wolber, G.; Schiller, J.; Huster, D.; Rademann, J. Chemoenzymatic Synthesis of Nonasulfated Tetrahyaluronan with a Paramagnetic Tag for Studying Its Complex with Interleukin-10. *Chem. - A Eur. J.* 2016. <https://doi.org/10.1002/chem.201504459>.
- (188) Park, Y.; Jowitz, T. A.; Day, A. J.; Prestegard, J. H. Nuclear Magnetic Resonance Insight into the Multiple Glycosaminoglycan Binding Modes of the Link Module from Human TSG-6. *Biochemistry* 2016. <https://doi.org/10.1021/acs.biochem.5b01148>.
- (189) Moure, M. J.; Eletsky, A.; Gao, Q.; Morris, L. C.; Yang, J. Y.; Chapla, D.; Zhao, Y.; Zong, C.; Amster, I. J.; Moremen, K. W.; Boons, G. J.; Prestegard, J. H. Paramagnetic Tag for Glycosylation Sites in Glycoproteins: Structural Constraints on Heparan Sulfate Binding to Robo1. *ACS Chem. Biol.* 2018. <https://doi.org/10.1021/acscchembio.8b00511>.
- (190) Scott, J. E.; Cummings, C.; Greiling, H.; Stuhlsatz, H. W.; Gregory, J. D.; Damle, S. P. Examination of Corneal Proteoglycans and Glycosaminoglycans by Rotary Shadowing and Electron Microscopy. *Int. J. Biol. Macromol.* 1990. [https://doi.org/10.1016/0141-8130\(90\)90029-A](https://doi.org/10.1016/0141-8130(90)90029-A).
- (191) Maloney, F. P.; Kuklewicz, J.; Corey, R. A.; Bi, Y.; Ho, R.; Mateusiak, L.; Pardon, E.; Steyaert, J.; Stansfeld, P. J.; Zimmer, J. Structure, Substrate Recognition and Initiation of Hyaluronan Synthase. *Nature* 2022. <https://doi.org/10.1038/s41586-022-04534-2>.
- (192) Koutsakis, C.; Franchi, M.; Tavianatou, A. G.; Masola, V.; Karamanos, N. K. Studying the Effects of Glycosaminoglycans in Cell Morphological Aspect with Scanning Electron Microscopy. *Methods Mol. Biol.* 2023. [https://doi.org/10.1007/978-1-0716-2946-8\\_8](https://doi.org/10.1007/978-1-0716-2946-8_8).
- (193) Raspanti, M.; Protasoni, M.; Zecca, P. A.; Reguzzoni, M. Slippery When Wet: The Free Surface of the Articular Cartilage. *Microsc. Res. Tech.* 2021. <https://doi.org/10.1002/jemt.23684>.



- (194) Watanabe, T.; Kametani, K.; Koyama, Y. I.; Suzuki, D.; Imamura, Y.; Takehana, K.; Hiramatsu, K. Ring-Mesh Model of Proteoglycan Glycosaminoglycan Chains in Tendon Based on Three-Dimensional Reconstruction by Focused Ion Beam Scanning Electron Microscopy. *J. Biol. Chem.* 2016. <https://doi.org/10.1074/jbc.M116.733857>.
- (195) PDB Data Distribution by Experimental Method and Molecular Type (26.05.2023) <https://www.rcsb.org/stats/summary>.
- (196) Huang, W.; Boju, L.; Tkalec, L.; Su, H.; Yang, H. O.; Gunay, N. S.; Linhardt, R. J.; Yeong, S. K.; Matte, A.; Cygler, M. Active Site of Chondroitin AC Lyase Revealed by the Structure of Enzyme - Oligosaccharide Complexes and Mutagenesis. *Biochemistry* 2001. <https://doi.org/10.1021/bi0024254>.
- (197) Tesmer, J. J. G.; Dessauer, C. W.; Sunahara, R. K.; Murray, L. D.; Johnson, R. A.; Gilman, A. G.; Sprang, S. R. Molecular Basis for P-Site Inhibition of Adenylyl Cyclase. *Biochemistry* 2000. <https://doi.org/10.1021/bi0015562>.
- (198) Shao, C.; Zhang, F.; Kemp, M. M.; Linhardt, R. J.; Waisman, D. M.; Head, J. F.; Seaton, B. A. Crystallographic Analysis of Calcium-Dependent Heparin Binding to Annexin A2. *J. Biol. Chem.* 2006. <https://doi.org/10.1074/jbc.M604502200>.
- (199) Capila, I.; Hernáiz, M. J.; Mo, Y. D.; Mealy, T. R.; Campos, B.; Dedman, J. R.; Linhardt, R. J.; Seaton, B. A. Annexin V-Heparin Oligosaccharide Complex Suggests Heparan Sulfate-Mediated Assembly on Cell Surfaces. *Structure* 2001. [https://doi.org/10.1016/S0969-2126\(00\)00549-9](https://doi.org/10.1016/S0969-2126(00)00549-9).
- (200) Moon, A. F.; Edavettal, S. C.; Krahn, J. M.; Munoz, E. M.; Negishi, M.; Linhardt, R. J.; Liu, J.; Pedersen, L. C. Structural Analysis of the Sulfotransferase (3-O-Sulfotransferase Isoform 3) Involved in the Biosynthesis of an Entry Receptor for Herpes Simplex Virus 1. *J. Biol. Chem.* 2004. <https://doi.org/10.1074/jbc.M405013200>.
- (201) Shaya, D.; Tocilj, A.; Li, Y.; Myette, J.; Venkataraman, G.; Sasisekharan, R.; Cygler, M. Crystal Structure of Heparinase II from *Pedobacter heparinus* and Its Complex with a Disaccharide Product. *J. Biol. Chem.* 2006. <https://doi.org/10.1074/jbc.M512055200>.
- (202) Wu, L.; Viola, C. M.; Brzozowski, A. M.; Davies, G. J. Structural Characterization of Human Heparanase Reveals Insights into Substrate Recognition. *Nat. Struct. Mol. Biol.* 2015. <https://doi.org/10.1038/nsmb.3136>.
- (203) MatrixDB <http://matrixdb.univ-lyon1.fr>.
- (204) Chautard, E.; Ballut, L.; Thierry-Mieg, N.; Ricard-Blum, S. MatrixDB, a Database Focused on Extracellular Protein-Protein and Protein-Carbohydrate Interactions. *Bioinformatics* 2009. <https://doi.org/10.1093/bioinformatics/btp025>.
- (205) GAG-DB <https://www.gagdb.glycopedia.eu>.
- (206) Pérez, S.; Bonnardel, F.; Lisacek, F.; Imberty, A.; Blum, S. R.; Makshakova, O. GAG-DB, the New Interface of the Three-Dimensional Landscape of Glycosaminoglycans. *Biomolecules* 2020. <https://doi.org/10.3390/biom10121660>.
- (207) Almond, A.; Sheehan, J. K. Glycosaminoglycan Conformation: Do Aqueous Molecular Dynamics Simulations Agree with x-Ray Fiber Diffraction? *Glycobiology* 2000. <https://doi.org/10.1093/glycob/10.3.329>.
- (208) Bakhtiar, R. Surface Plasmon Resonance Spectroscopy: A Versatile Technique in a Biochemist's Toolbox. *J. Chem. Educ.* 2013. <https://doi.org/10.1021/ed200549g>.
- (209) Zhang, F.; Datta, P.; Dordick, J. S.; Linhardt, R. J. Evaluating Heparin Products for Heparin-Induced Thrombocytopenia Using Surface Plasmon Resonance. *J. Pharm. Sci.* 2020. <https://doi.org/10.1016/j.xphs.2019.10.040>.
- (210) Schasfoort, R. B. M.; Vallet, S. D.; Deddens, L.; Vonarburg, A.; Salza, R.; Faye, C.; Aranyos, A.; Thierry-Mieg, N.; Ricard-Blum, S. *Handbook of Surface Plasmon Resonance 2nd Edition*; 2017.
- (211) Munakata, H.; Takagaki, K.; Majima, M.; Endo, M. Interaction between Collagens and Glycosaminoglycans Investigated Using a Surface Plasmon Resonance Biosensor. *Glycobiology* 1999. <https://doi.org/10.1093/glycob/9.10.1023>.
- (212) Théoleyre, S.; Kwan Tat, S.; Vusio, P.; Blanchard, F.; Gallagher, J.; Ricard-Blum, S.; Fortun, Y.; Padrines, M.; Rédini, F.; Heymann, D. Characterization of Osteoprotegerin Binding to Glycosaminoglycans by Surface Plasmon Resonance: Role in the Interactions with Receptor Activator of Nuclear Factor KB Ligand (RANKL) and RANK. *Biochem. Biophys. Res. Commun.* 2006. <https://doi.org/10.1016/j.bbrc.2006.06.120>.
- (213) Rusnati, M.; Bugatti, A. Surface Plasmon Resonance Analysis of Heparin-Binding Angiogenic Growth Factors. In *Methods in Molecular Biology*; 2016. [https://doi.org/10.1007/978-1-4939-3999-2\\_7](https://doi.org/10.1007/978-1-4939-3999-2_7).
- (214) Daniel R. Sandoval, Alejandro Gomez Toledo, Chelsea D. Painter, Ember M. Tota, XM. Osman Sheikh, Alan M. V. West, Martin M. Frank, Lance Wells, Ding Xu, Roy Bicknell, Kevin D. Corbett, Xj. D. E. Proteomics-Based Screening of the Endothelial Heparansulfate Interactome Reveals That C-Type Lectin 14a (CLEC14A) Is a Heparin-Binding Protein. *J Biol Chem.* 2020. <https://doi.org/10.1074/jbc.RA119.011639>.
- (215) Yan, L.; Song, Y.; Xia, K.; He, P.; Zhang, F.; Chen, S.; Pouliot, R.; Weiss, D. J.; Tandon, R.; Bates, J. T.; Ederer, D. R.; Mitra, D.; Sharma, P.; Davis, A.; Linhardt, R. J. Heparan Sulfates from Bat and Human Lung and Their Binding to the Spike Protein of SARS-CoV-2 Virus. *Carbohydr. Polym.* 2021. <https://doi.org/10.1016/j.carbpol.2021.117797>.
- (216) Przybylski, C.; Gonnet, F.; Saesen, E.; Lortat-Jacob, H.; Daniel, R. Surface Plasmon Resonance Imaging Coupled to On-Chip Mass Spectrometry: A New Tool to Probe Protein-GAG Interactions. *Anal. Bioanal. Chem.* 2020. <https://doi.org/10.1007/s00216-019-02267-2>.
- (217) Yu, Y.; Zhang, F.; Renois-Predelus, G.; Amster, J. I.; Linhardt, R. J. L. Filter-Entrapment Enrichment Pull-down Assay for Glycosaminoglycan Structural Characterization and Protein Interaction. *Carbohydr. Polym.* 2020. <https://doi.org/10.1016/j.carbpol.2020.116623>.
- (218) Shi, D.; He, P.; Song, Y.; Cheng, S.; Linhardt, R. J.; Dordick, J. S.; Chi, L.; Zhang, F. Kinetic and Structural Aspects of Glycosaminoglycan-Monkeypox Virus Protein A29 Interactions Using Surface Plasmon Resonance. *Molecules* 2022. <https://doi.org/10.3390/molecules27185898>.
- (219) Rich, R. L.; Myszka, D. G. Higher-Throughput, Label-Free, Real-Time Molecular Interaction Analysis. *Analytical Biochemistry.* 2007. <https://doi.org/10.1016/j.ab.2006.10.040>.
- (220) Kamat, V.; Rafique, A. Designing Binding Kinetic Assay on the Bio-Layer Interferometry (BLI) Biosensor to Characterize Antibody-Antigen Interactions. *Anal. Biochem.* 2017. <https://doi.org/10.1016/j.ab.2017.08.002>.
- (221) Laguri, C.; Sadir, R.; Gout, E.; Vivès, R. R.; Lortat-Jacob, H. Preparation and Characterization of Heparan Sulfate-Derived Oligosaccharides to Investigate Protein-GAG Interaction and HS Biosynthesis Enzyme Activity. In *Methods in Molecular Biology*; 2022. [https://doi.org/10.1007/978-1-0716-1398-6\\_11](https://doi.org/10.1007/978-1-0716-1398-6_11).
- (222) Groner, M.; Ng, T.; Wang, W.; Udit, A. K. Bio-Layer Interferometry of a Multivalent Sulfated Virus Nanoparticle with Heparin-like Anticoagulant Activity. *Anal. Bioanal. Chem.* 2015. <https://doi.org/10.1007/s00216-015-8735-x>.
- (223) Kowarschik, S.; Schöllkopf, J.; Müller, T.; Tian, S.; Knerr, J.; Bakker, H.; Rein, S.; Dong, M.; Weber, S.; Grosse, R.; Schmidt, G. Yersinia Pseudotuberculosis Cytotoxic Factor Interacts with Glycosaminoglycans. *FASEB J.* 2021. <https://doi.org/10.1096/fj.202001630R>.
- (224) Sharma, N.; Haggstrom, L.; Sohrabipour, S.; Dwivedi, D. J.; Liaw, P. C. Investigations of the Effectiveness of Heparin Variants as Inhibitors of Histones. *J. Thromb. Haemost.* 2022. <https://doi.org/10.1111/jth.15706>.
- (225) Vallet, S. D.; Berthollier, C.; Salza, R.; Muller, L.; Ricard-Blum, S. The Interactome of Cancer-Related Lysyl Oxidase and Lysyl Oxidase-like Proteins. *Cancers (Basel).* 2021. <https://doi.org/10.3390/cancers13010071>.

- (226) Diaz-Salmeron, R.; Michel, J. P.; Hadji, H.; Gout, E.; Vivès, R. R.; Ponchel, G.; Bouchemal, K. Role of the Interactions of Soft Hyaluronan Nanomaterials with CD44 and Supported Bilayer Membranes in the Cellular Uptake. *Colloids Surfaces B Biointerfaces* 2021. <https://doi.org/10.1016/j.colsurfb.2021.111916>.
- (227) Coburger, I.; Dahms, S. O.; Roeser, D.; Gührs, K. H.; Hortschansky, P.; Than, M. E. Analysis of the Overall Structure of the Multi-Domain Amyloid Precursor Protein (APP). *PLoS One* 2013. <https://doi.org/10.1371/journal.pone.0081926>.
- (228) Liang, A.; Desai, U. Advances in Studying Glycosaminoglycan-Protein Interactions Using Capillary Electrophoresis. In *Methods in Molecular Biology*; 2022. [https://doi.org/10.1007/978-1-0716-1398-6\\_30](https://doi.org/10.1007/978-1-0716-1398-6_30).
- (229) Heegaard, N. H. H.; Mortensen, H. D.; Roepstorff, P. Demonstration of a Heparin-Binding Site in Serum Amyloid P Component Using Affinity Capillary Electrophoresis as an Adjunct Technique. *J. Chromatogr. A* 1995. [https://doi.org/10.1016/0021-9673\(95\)00644-3](https://doi.org/10.1016/0021-9673(95)00644-3).
- (230) Heegaard, N. H. H.; He, X.; Blomberg, L. G. Binding of Ca<sup>2+</sup>, Mg<sup>2+</sup>, and Heparin by Human Serum Amyloid P Component in Affinity Capillary Electrophoresis. *Electrophoresis* 2006. <https://doi.org/10.1002/elps.200600005>.
- (231) Hamazaki, H. Ca<sup>2+</sup>-Mediated Association of Human Serum Amyloid P Component with Heparan Sulfate and Dermatan Sulfate. *J. Biol. Chem.* 1987. [https://doi.org/10.1016/s0021-9258\(19\)75657-9](https://doi.org/10.1016/s0021-9258(19)75657-9).
- (232) Li, X. A.; Hatanaka, K.; Guo, L.; Kitamura, Y.; Yamamoto, A. Binding of Serum Amyloid P Component to Heparin in Human Serum. *BBA - Gen. Subj.* 1994. [https://doi.org/10.1016/0304-4165\(94\)90034-5](https://doi.org/10.1016/0304-4165(94)90034-5).
- (233) Heegaard, N. H. H.; Roepstorff, P.; Melberg, S. G.; Nissen, M. H. Cleaved B2-Microglobulin Partially Attains a Conformation That Has Amyloidogenic Features. *J. Biol. Chem.* 2002. <https://doi.org/10.1074/jbc.M108837200>.
- (234) Heegaard, N. H. H.; Robey, F. A. Use of Capillary Zone Electrophoresis To Evaluate the Binding of Anionic Carbohydrates to Synthetic Peptides Derived from Human Serum Amyloid P Component. *Anal. Chem.* 1992. <https://doi.org/10.1021/ac00045a004>.
- (235) Hernaiz, M. J.; LeBrun, L. A.; Wu, Y.; Sen, J. W.; Linhardt, R. J.; Heegaard, N. H. H. Characterization of Heparin Binding by a Peptide from Amyloid P Component Using Capillary Electrophoresis, Surface Plasmon Resonance and Isothermal Titration Calorimetry. *Eur. J. Biochem.* 2002. <https://doi.org/10.1046/j.1432-1033.2002.02964.x>.
- (236) Hattori, T.; Kimura, K.; Seyrek, E.; Dubin, P. L. Binding of Bovine Serum Albumin to Heparin Determined by Turbidimetric Titration and Frontal Analysis Continuous Capillary Electrophoresis. *Anal. Biochem.* 2001. <https://doi.org/10.1006/abio.2001.5129>.
- (237) Lipponen, K.; Liu, Y.; Stege, P. W.; Öörni, K.; Kovanen, P. T.; Riekkola, M. L. Capillary Electrochromatography and Quartz Crystal Microbalance, Valuable Techniques in the Study of Heparin-Lipoprotein Interactions. *Anal. Biochem.* 2012. <https://doi.org/10.1016/j.ab.2012.02.017>.
- (238) Wu, X.; Linhardt, R. J. Capillary Affinity Chromatography and Affinity Capillary Electrophoresis of Heparin Binding Proteins. In *Electrophoresis*; 1998. <https://doi.org/10.1002/elps.1150191514>.
- (239) Seyrek, E.; Dubin, P. L.; Henriksen, J. Nonspecific Electrostatic Binding Characteristics of the Heparin-Antithrombin Interaction. *Biopolymers* 2007. <https://doi.org/10.1002/bip.20731>.
- (240) Mozafari, M.; Balasupramaniam, S.; Preu, L.; El Deeb, S.; Reiter, C. G.; Wätzig, H. Using Affinity Capillary Electrophoresis and Computational Models for Binding Studies of Heparinoids with P-Selectin and Other Proteins. *Electrophoresis* 2017. <https://doi.org/10.1002/elps.201600480>.
- (241) Zhang, M.; Li, G.; Zhang, Y.; Kang, J. Quantitative Analysis of Antithrombin III Binding Site in Low Molecular Weight Heparins by Exhaustive Heparinases Digestion and Capillary Electrophoresis. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 2017. <https://doi.org/10.1016/j.jchromb.2017.08.047>.
- (242) Miles, A. J.; Janes, R. W.; Wallace, B. A. Tools and Methods for Circular Dichroism Spectroscopy of Proteins: A Tutorial Review. *Chemical Society Reviews*. 2021. <https://doi.org/10.1039/d0cs00558d>.
- (243) Matsuo, K.; Gekko, K. Synchrotron-Radiation Vacuum-Ultraviolet Circular-Dichroism Spectroscopy for Characterizing the Structure of Saccharides. In *Advances in Experimental Medicine and Biology*; 2018. [https://doi.org/10.1007/978-981-13-2158-0\\_6](https://doi.org/10.1007/978-981-13-2158-0_6).
- (244) CHAKRABARTI, B. Carboxyl and Amide Transitions in the Circular Dichroism of Glycosaminoglycans; 1981. <https://doi.org/10.1021/bk-1981-0150.ch019>.
- (245) Matsuo, K.; Namatame, H.; Taniguchi, M.; Gekko, K. Vacuum-Ultraviolet Circular Dichroism Analysis of Glycosaminoglycans by Synchrotron-Radiation Spectroscopy. *Biosci. Biotechnol. Biochem.* 2009. <https://doi.org/10.1271/bbb.80605>.
- (246) Stanley, F. E.; Warner, A. M.; Gutierrez, S. M.; Stalcup, A. M. Heparin-Induced Circular Dichroism of Chloroquine. *Biochem. Biophys. Res. Commun.* 2009. <https://doi.org/10.1016/j.bbrc.2009.07.098>.
- (247) Zsila, F.; Gedeon, G. Binding of Anti-Prion Agents to Glycosaminoglycans: Evidence from Electronic Absorption and Circular Dichroism Spectroscopy. *Biochem. Biophys. Res. Commun.* 2006. <https://doi.org/10.1016/j.bbrc.2006.06.033>.
- (248) Stone, A. L.; Beeler, D.; Oosta, G.; Rosenberg, R. D. Circular Dichroism Spectroscopy of Heparin-Antithrombin Interactions. *Proc. Natl. Acad. Sci. U. S. A.* 1982. <https://doi.org/10.1073/pnas.79.23.7190>.
- (249) Satish, L.; Santra, S.; Tsurkan, M. V.; Werner, C.; Jana, M.; Sahoo, H. Conformational Changes of GDNF-Derived Peptide Induced by Heparin, Heparan Sulfate, and Sulfated Hyaluronic Acid – Analysis by Circular Dichroism Spectroscopy and Molecular Dynamics Simulation. *Int. J. Biol. Macromol.* 2021. <https://doi.org/10.1016/j.ijbiomac.2021.05.194>.
- (250) Zsila, F. Glycosaminoglycans Are Potential Pharmacological Targets for Classic DNA Minor Groove Binder Drugs Berenil and Pentamidine. *Phys. Chem. Chem. Phys.* 2015. <https://doi.org/10.1039/c5cp03153b>.
- (251) Zsila, F.; Samsonov, S. A.; Maszota-Zieleniak, M. Mind Your Dye: The Amyloid Sensor Thioflavin t Interacts with Sulfated Glycosaminoglycans Used to Induce Cross- $\beta$ -Sheet Motifs. *J. Phys. Chem. B* 2020. <https://doi.org/10.1021/acs.jpcc.0c08273>.
- (252) Maszota-Zieleniak, M.; Zsila, F.; Samsonov, S. A. Computational Insights into Heparin-Small Molecule Interactions: Evaluation of the Balance between Stacking and Non-Stacking Binding Modes. *Carbohydr. Res.* 2021. <https://doi.org/10.1016/j.carres.2021.108390>.
- (253) Stanley, F. E.; Warner, A. M.; McWilliams, K. M.; Stalcup, A. M. Heparin-Induced Circular Dichroism of an Achiral, Bicyclic Species. *Chirality* 2011. <https://doi.org/10.1002/chir.20873>.
- (254) Uciechowska-Kaczmarzyk, U.; Babik, S.; Zsila, F.; Bojarski, K. K.; Beke-Somfai, T.; Samsonov, S. A. Molecular Dynamics-Based Model of VEGF-A and Its Heparin Interactions. *J. Mol. Graph. Model.* 2018. <https://doi.org/10.1016/j.jmgl.2018.04.015>.
- (255) Nordsieck, K.; Baumann, L.; Hintze, V.; Pisabarro, M. T.; Schnabelrauch, M.; Beck-Sickinger, A. G.; Samsonov, S. A. The Effect of Interleukin-8 Truncations on Its Interactions with Glycosaminoglycans. *Biopolymers* 2018. <https://doi.org/10.1002/bip.23103>.
- (256) Zsila, F.; Juhász, T.; Kohut, G.; Beke-Somfai, T. Heparin and Heparan Sulfate Binding of the Antiparasitic Drug Imidocarb: Circular Dichroism Spectroscopy, Isothermal Titration Calorimetry, and Computational Studies. *J. Phys. Chem. B* 2018. <https://doi.org/10.1021/acs.jpcc.7b08876>.
- (257) Rudd, T.; Yates, E.; Hricovini, M. Spectroscopic and Theoretical Approaches for the Determination of Heparin Saccharide Structure and the Study of Protein-Glycosaminoglycan Complexes in Solution. *Curr. Med. Chem.* 2009. <https://doi.org/10.2174/092986709789878193>.
- (258) Johnson, C. M. Isothermal Titration Calorimetry. In *Methods in Molecular Biology*; 2021. [https://doi.org/10.1007/978-1-0716-1197-5\\_5](https://doi.org/10.1007/978-1-0716-1197-5_5).

- (259) Dutta, A. K.; Rösgen, J.; Rajarathnam, K. Using Isothermal Titration Calorimetry to Determine Thermodynamic Parameters of Protein–Glycosaminoglycan Interactions. *Methods Mol. Biol.* 2015. [https://doi.org/10.1007/978-1-4939-1714-3\\_25](https://doi.org/10.1007/978-1-4939-1714-3_25).
- (260) Pilipenko, I.; Korzhikov-Vlakh, V.; Valtari, A.; Anufrikov, Y.; Kalinin, S.; Ruponen, M.; Krasavin, M.; Urtti, A.; Tennikova, T. Mucoadhesive Properties of Nanogels Based on Stimuli-Sensitive Glycosaminoglycan–Graft-PNIPAAm Copolymers. *Int. J. Biol. Macromol.* 2021. <https://doi.org/10.1016/j.ijbiomac.2021.07.070>.
- (261) Malicka, W.; Haag, R.; Ballauff, M. Interaction of Heparin with Proteins: Hydration Effects. *J. Phys. Chem. B* 2022. <https://doi.org/10.1021/acs.jpcc.2c04928>.
- (262) Muñoz, E.; Sabin, J. The Use of ITC and the Software AFFINImeter for the Quantification of the Anticoagulant Pentasaccharide in Low Molecular Weight Heparin. In *Methods in Molecular Biology*; 2019. [https://doi.org/10.1007/978-1-4939-9179-2\\_15](https://doi.org/10.1007/978-1-4939-9179-2_15).
- (263) Walkowiak, J. J.; Ballauff, M.; Zimmermann, R.; Freudenberg, U.; Werner, C. Thermodynamic Analysis of the Interaction of Heparin with Lysozyme. *Biomacromolecules* 2020. <https://doi.org/10.1021/acs.biomac.0c00780>.
- (264) Lauth, L. M.; Voigt, B.; Bhatia, T.; Machner, L.; Balbach, J.; Ott, M. Heparin Promotes Rapid Fibrillation of the Basic Parathyroid Hormone at Physiological pH. *FEBS Lett.* 2022. <https://doi.org/10.1002/1873-3468.14455>.
- (265) Havlíková, M.; Jugl, A.; Krouská, J.; Szabová, J.; Mravcová, L.; Venerová, T.; Chang, C. H.; Pekař, M.; Mravec, F. Interactions between Cationic Ion Pair Amphiphile Vesicles and Hyaluronan - A Physicochemical Study. *Langmuir* 2021. <https://doi.org/10.1021/acs.langmuir.1c00993>.
- (266) Corredor, M.; Carbajo, D.; Domingo, C.; Pérez, Y.; Bujons, J.; Messegue, A.; Alfonso, I. Dynamic Covalent Identification of an Efficient Heparin Ligand. *Angew. Chemie - Int. Ed.* 2018. <https://doi.org/10.1002/anie.201806770>.
- (267) Hamza, A.; Samad, A.; Parray, Z. A.; Ara, S.; Ahmed, A.; Almajhdi, F. N.; Hussain, T.; Islam, A.; Parveen, S. Mutation in the CX3C Motif of G Protein Disrupts Its Interaction with Heparan Sulfate: A Calorimetric, Spectroscopic, and Molecular Docking Study. *Int. J. Mol. Sci.* 2022. <https://doi.org/10.3390/ijms23041950>.
- (268) Adrian, E.; Trellová, D.; Filová, E.; Kumorek, M.; Lobaz, V.; Poreba, R.; Janoušková, O.; Pop-georgievski, O.; Lacík, I.; Kubies, D. Complexation of Cxcl12, Fgf-2 and Vegf with Heparin Modulates the Protein Release from Alginate Microbeads. *Int. J. Mol. Sci.* 2021. <https://doi.org/10.3390/ijms222111666>.
- (269) Scrocco, E.; Tomasi, J. The Electrostatic Molecular Potential as a Tool for the Interpretation of Molecular Properties. In *New Concepts II*; 2007. [https://doi.org/10.1007/3-540-06399-4\\_6](https://doi.org/10.1007/3-540-06399-4_6).
- (270) Petrongolo, C.; Preston, H. J. T.; Kaufman, J. J. Ab Initio LCAO-MO-SCF Calculation of the Electrostatic Molecular Potential of Chlorpromazine and Promazine. *Int. J. Quantum Chem.* 1978. <https://doi.org/10.1002/qua.560130402>.
- (271) Bojarski, K. K.; Sieradzian, A. K.; Samsonov, S. A. Molecular Dynamics Insights into Protein–Glycosaminoglycan Systems from Microsecond-Scale Simulations. *Biopolymers* 2019, 110 (7), e23252. <https://doi.org/10.1002/bip.23252>.
- (272) Pagielska, M.; Samsonov, S. A. Molecular Dynamics-Based Comparative Analysis of Chondroitin and Dermatan Sulfates. *Biomolecules* 2023. <https://doi.org/10.3390/biom13020247>.
- (273) Samsonov, S. A. Computational Analysis of Solvent Inclusion in Docking Studies of Protein–Glycosaminoglycan Systems. In *Methods in Molecular Biology*; 2018. [https://doi.org/10.1007/978-1-4939-7756-7\\_22](https://doi.org/10.1007/978-1-4939-7756-7_22).
- (274) Hileman, R. E.; Fromm, J. R.; Weiler, J. M.; Linhardt, R. J. Glycosaminoglycan-Protein Interactions: Definition of Consensus Sites in Glycosaminoglycan Binding Proteins. *BioEssays*. 1998. [https://doi.org/10.1002/\(SICI\)1521-1878\(199802\)20:2<156::AID-BIES8>3.0.CO;2-R](https://doi.org/10.1002/(SICI)1521-1878(199802)20:2<156::AID-BIES8>3.0.CO;2-R).
- (275) Bitomsky, W.; Wade, R. C. Docking of Glycosaminoglycans to Heparin-Binding Proteins: Validation for AFGF, BFGF, and Antithrombin and Application to IL-8. *J. Am. Chem. Soc.* 1999. <https://doi.org/10.1021/ja983319g>.
- (276) Kuhn, L. A.; Griffin, J. H.; Fisher, C. L.; Greengard, J. S.; Bouma, B. N.; Espana, F.; Tainer, J. A. Elucidating the Structural Chemistry of Glycosaminoglycan Recognition by Protein C Inhibitor. *Proc. Natl. Acad. Sci. U. S. A.* 1990. <https://doi.org/10.1073/pnas.87.21.8506>.
- (277) Samsonov, S. A.; Bichmann, L.; Pisabarro, M. T. Coarse-Grained Model of Glycosaminoglycans. *J. Chem. Inf. Model.* 2015. <https://doi.org/10.1021/ci500669w>.
- (278) Gehrcke, J. P.; Pisabarro, M. T. Identification and Characterization of a Glycosaminoglycan Binding Site on Interleukin-10 via Molecular Simulation Methods. *J. Mol. Graph. Model.* 2015. <https://doi.org/10.1016/j.jmgm.2015.09.003>.
- (279) Honig, B.; Nicholls, A. Classical Electrostatics in Biology and Chemistry. *Science (80-. )*. 1995. <https://doi.org/10.1126/science.7761829>.
- (280) Warwicker, J.; Watson, H. C. Calculation of the Electric Potential in the Active Site Cleft Due to  $\alpha$ -Helix Dipoles. *J. Mol. Biol.* 1982. [https://doi.org/10.1016/0022-2836\(82\)90505-8](https://doi.org/10.1016/0022-2836(82)90505-8).
- (281) Samsonov, S. A.; Pisabarro, M. T. Computational Analysis of Interactions in Structurally Available Protein-Glycosaminoglycan Complexes. *Glycobiology* 2016. <https://doi.org/10.1093/glycob/cww055>.
- (282) Lopez-Vallejo, F.; Caulfield, T.; Martinez-Mayorga, K.; A. Giulianotti, M.; Nefzi, A.; A. Houghten, R.; L. Medina-Franco, J. Integrating Virtual Screening and Combinatorial Chemistry for Accelerated Drug Discovery. *Comb. Chem. High Throughput Screen.* 2011. <https://doi.org/10.2174/138620711795767866>.
- (283) Lengauer, T.; Rarey, M. Computational Methods for Biomolecular Docking. *Curr. Opin. Struct. Biol.* 1996. [https://doi.org/10.1016/S0959-440X\(96\)80061-3](https://doi.org/10.1016/S0959-440X(96)80061-3).
- (284) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and Scoring in Virtual Screening for Drug Discovery: Methods and Applications. *Nature Reviews Drug Discovery.* 2004. <https://doi.org/10.1038/nrd1549>.
- (285) Ferreira, L. G.; Dos Santos, R. N.; Oliva, G.; Andricopulo, A. D. Molecular Docking and Structure-Based Drug Design Strategies. *Molecules.* 2015. <https://doi.org/10.3390/molecules200713384>.
- (286) Levitt, D. G.; Banaszak, L. J. POCKET: A Computer Graphics Method for Identifying and Displaying Protein Cavities and Their Surrounding Amino Acids. *J. Mol. Graph.* 1992. [https://doi.org/10.1016/0263-7855\(92\)80074-N](https://doi.org/10.1016/0263-7855(92)80074-N).
- (287) Goodford, P. J. A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules. *J. Med. Chem.* 1985. <https://doi.org/10.1021/jm00145a002>.
- (288) Laskowski, R. A. SURFNET: A Program for Visualizing Molecular Surfaces, Cavities, and Intermolecular Interactions. *J. Mol. Graph.* 1995. [https://doi.org/10.1016/0263-7855\(95\)00073-9](https://doi.org/10.1016/0263-7855(95)00073-9).
- (289) Brady, G. P.; Stouten, P. F. W. Fast Prediction and Visualization of Protein Binding Pockets with PASS. *J. Comput. Aided. Mol. Des.* 2000. <https://doi.org/10.1023/A:1008124202956>.
- (290) Oliveira, J. S.; Pereira, J. H.; Canduri, F.; Rodrigues, N. C.; de Souza, O. N.; de Azevedo, W. F.; Basso, L. A.; Santos, D. S. Crystallographic and Pre-Steady-State Kinetics Studies on Binding of NADH to Wild-Type and Isoniazid-Resistant Enoyl-ACP(CoA) Reductase Enzymes from Mycobacterium Tuberculosis. *J. Mol. Biol.* 2006. <https://doi.org/10.1016/j.jmb.2006.03.055>.
- (291) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Protein-Ligand Docking: Current Status and Future Challenges. *Proteins: Structure, Function and Genetics.* 2006. <https://doi.org/10.1002/prot.21082>.
- (292) Dias, R.; de Azevedo Jr., W. Molecular Docking Algorithms. *Curr. Drug Targets* 2008. <https://doi.org/10.2174/138945008786949432>.

- (293) Kramer, B.; Rarey, M.; Lengauer, T. Evaluation of the FlexX Incremental Construction Algorithm for Protein-Ligand Docking. *Proteins Struct. Funct. Genet.* 1999. [https://doi.org/10.1002/\(SICI\)1097-0134\(19991101\)37:2<228::AID-PROT8>3.0.CO;2-8](https://doi.org/10.1002/(SICI)1097-0134(19991101)37:2<228::AID-PROT8>3.0.CO;2-8).
- (294) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A Fast Flexible Docking Method Using an Incremental Construction Algorithm. *J. Mol. Biol.* 1996. <https://doi.org/10.1006/jmbi.1996.0477>.
- (295) Schellhammer, I.; Rarey, M. FlexX-Scan: Fast, Structure-Based Virtual Screening. *Proteins Struct. Funct. Bioinforma.* 2004. <https://doi.org/10.1002/prot.20217>.
- (296) Moré, J. J.; Wu, Z. Distance Geometry Optimization for Protein Structures. *J. Glob. Optim.* 1999. <https://doi.org/10.1023/A:1008380219900>.
- (297) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A Geometric Approach to Macromolecule-Ligand Interactions. *J. Mol. Biol.* 1982. [https://doi.org/10.1016/0022-2836\(82\)90153-X](https://doi.org/10.1016/0022-2836(82)90153-X).
- (298) Chen, R.; Li, L.; Weng, Z. ZDOCK: An Initial-Stage Protein-Docking Algorithm. *Proteins Struct. Funct. Genet.* 2003. <https://doi.org/10.1002/prot.10389>.
- (299) Yadava, U. Search Algorithms and Scoring Methods in Protein-Ligand Docking. *Endocrinol. Int. J.* 2018. <https://doi.org/10.15406/emij.2018.06.00212>.
- (300) Pérez, S.; Tvaroška, I. Carbohydrate-Protein Interactions: Molecular Modeling Insights. In *Advances in Carbohydrate Chemistry and Biochemistry*; 2014. <https://doi.org/10.1016/B978-0-12-800128-8.00001-7>.
- (301) Read, R. J.; Hart, T. N.; Cummings, M. D.; Ness, S. R. Monte Carlo Algorithms for Docking to Proteins. *Supramol. Chem.* 1995. <https://doi.org/10.1080/10610279508032529>.
- (302) Liu, M.; Wang, S. MCDOCK: A Monte Carlo Simulation Approach to the Molecular Docking Problem. *J. Comput. Aided. Mol. Des.* 1999. <https://doi.org/10.1023/A:1008005918983>.
- (303) Gardiner, E. J.; Willett, P.; Artymiuk, P. J. Protein Docking Using a Genetic Algorithm. *Proteins Struct. Funct. Genet.* 2001. <https://doi.org/10.1002/prot.1070>.
- (304) Steinmann, C.; Jensen, J. H. Using a Genetic Algorithm to Find Molecules with Good Docking Scores. *PeerJ Phys. Chem.* 2021. <https://doi.org/10.7717/peerj-pchem.18>.
- (305) Baxter, C. A.; Murray, C. W.; Clark, D. E.; Westhead, D. R.; Eldridge, M. D. Flexible Docking Using Tabu Search and an Empirical Estimate of Binding Affinity. *Proteins Struct. Funct. Genet.* 1998. [https://doi.org/10.1002/\(SICI\)1097-0134\(19981115\)33:3<367::AID-PROT6>3.0.CO;2-W](https://doi.org/10.1002/(SICI)1097-0134(19981115)33:3<367::AID-PROT6>3.0.CO;2-W).
- (306) Glover, F. Future Paths for Integer Programming and Links to Artificial Intelligence. *Comput. Oper. Res.* 1986. [https://doi.org/10.1016/0305-0548\(86\)90048-1](https://doi.org/10.1016/0305-0548(86)90048-1).
- (307) Aqvist, J.; Luzhkov, V. B.; Brandsdal, B. O. Ligand Binding Affinities from MD Simulations. *Acc. Chem. Res.* 2002. <https://doi.org/10.1021/ar010014p>.
- (308) Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA Methods to Estimate Ligand-Binding Affinities. *Expert Opinion on Drug Discovery.* 2015. <https://doi.org/10.1517/17460441.2015.1032936>.
- (309) Guedes, I. A.; Pereira, F. S. S.; Dardenne, L. E. Empirical Scoring Functions for Structure-Based Virtual Screening: Applications, Critical Aspects, and Challenges. *Frontiers in Pharmacology.* 2018. <https://doi.org/10.3389/fphar.2018.01089>.
- (310) Korb, O.; Stützel, T.; Exner, T. E. Empirical Scoring Functions for Advanced Protein-Ligand Docking with PLANTS. *J. Chem. Inf. Model.* 2009. <https://doi.org/10.1021/ci800298z>.
- (311) Yang, C.; Zhang, Y. Lin\_F9: A Linear Empirical Scoring Function for Protein-Ligand Docking. *J. Chem. Inf. Model.* 2021. <https://doi.org/10.1021/acs.jcim.1c00737>.
- (312) Barradas-Bautista, D.; Moal, I. H.; Fernández-Recio, J. A Systematic Analysis of Scoring Functions in Rigid-Body Protein Docking: The Delicate Balance between the Predictive Rate Improvement and the Risk of Overtraining. *Proteins Struct. Funct. Bioinforma.* 2017. <https://doi.org/10.1002/prot.25289>.
- (313) Muegge, I.; Martin, Y. C. A General and Fast Scoring Function for Protein-Ligand Interactions: A Simplified Potential Approach. *J. Med. Chem.* 1999. <https://doi.org/10.1021/jm980536j>.
- (314) Liu, J.; Wang, R. Classification of Current Scoring Functions. *J. Chem. Inf. Model.* 2015. <https://doi.org/10.1021/ci500731a>.
- (315) Ballester, P. J.; Mitchell, J. B. O. A Machine Learning Approach to Predicting Protein-Ligand Binding Affinity with Applications to Molecular Docking. *Bioinformatics* 2010. <https://doi.org/10.1093/bioinformatics/btq112>.
- (316) Guedes, I. A.; Barreto, A. M. S.; Marinho, D.; Krempser, E.; Kuenemann, M. A.; Sperandio, O.; Dardenne, L. E.; Miteva, M. A. New Machine Learning and Physics-Based Scoring Functions for Drug Discovery. *Sci. Rep.* 2021. <https://doi.org/10.1038/s41598-021-82410-1>.
- (317) Ashtawy, H. M.; Mahapatra, N. R. A Comparative Assessment of Predictive Accuracies of Conventional and Machine Learning Scoring Functions for Protein-Ligand Binding Affinity Prediction. *IEEE/ACM Trans. Comput. Biol. Bioinforma.* 2015. <https://doi.org/10.1109/TCBB.2014.2351824>.
- (318) Kozakov, D.; Clodfelter, K. H.; Vajda, S.; Camacho, C. J. Optimal Clustering for Detecting Near-Native Conformations in Protein Docking. *Biophys. J.* 2005. <https://doi.org/10.1529/biophysj.104.058768>.
- (319) Cardin, A. D.; Weintraub, H. J. R. Molecular Modeling of Protein-Glycosaminoglycan Interactions. *Arteriosclerosis* 1989. <https://doi.org/10.1161/01.atv.9.1.21>.
- (320) Uciechowska-Kaczmarzyk, U.; de Beauchene, I.; Samsonov, S. A. Docking Software Performance in Protein-Glycosaminoglycan Systems. *J. Mol. Graph. Model.* 2019, 90, 42–50. <https://doi.org/10.1016/j.jmkgm.2019.04.001>.
- (321) Griffith, A. R.; Rogers, C. J.; Miller, G. M.; Abrol, R.; Hsieh-Wilson, L. C.; Goddard, W. A. Predicting Glycosaminoglycan Surface Protein Interactions and Implications for Studying Axonal Growth. *Proc. Natl. Acad. Sci. U. S. A.* 2017. <https://doi.org/10.1073/pnas.1715093115>.
- (322) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function. *J. Comput. Chem.* 1998. [https://doi.org/10.1002/\(SICI\)1096-987X\(19981115\)19:14<1639::AID-JCC10>3.0.CO;2-B](https://doi.org/10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B).
- (323) Frank, M. Computational Docking as a Tool for the Rational Design of Carbohydrate-Based Drugs. *Top. Med. Chem.* 2014. [https://doi.org/10.1007/7355\\_2014\\_42](https://doi.org/10.1007/7355_2014_42).
- (324) Nivedha, A. K.; Thieker, D. F.; Makeneni, S.; Hu, H.; Woods, R. J. Vina-Carb: Improving Glycosidic Angles during Carbohydrate Docking. *J. Chem. Theory Comput.* 2016, 12 (2), 892–901. <https://doi.org/10.1021/acs.jctc.5b00834>.
- (325) Boittier, E. D.; Burns, J. M.; Gandhi, N. S.; Ferro, V. GlycoTorch Vina: Docking Designed and Tested for Glycosaminoglycans. *J. Chem. Inf. Model.* 2020. <https://doi.org/10.1021/acs.jcim.0c00373>.
- (326) Eberhardt, J.; Santos-Martins, D.; Tillack, A. F.; Forli, S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J. Chem. Inf. Model.* 2021. <https://doi.org/10.1021/acs.jcim.1c00203>.
- (327) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. I. Method and Assessment of Docking Accuracy. *J. Med. Chem.* 2004. <https://doi.org/10.1021/jm0306430>.

- (328) Desta, I. T.; Porter, K. A.; Xia, B.; Kozakov, D.; Vajda, S. Performance and Its Limits in Rigid Body Protein-Protein Docking. *Structure* 2020. <https://doi.org/10.1016/J.STR.2020.06.006>.
- (329) Mottarella, S. E.; Beglov, D.; Beglova, N.; Nugent, M. A.; Kozakov, D.; Vajda, S. Docking Server for the Identification of Heparin Binding Sites on Proteins. *J. Chem. Inf. Model.* 2014, 54 (7), 2068–2078. <https://doi.org/10.1021/CI500115J>.
- (330) Van Zundert, G. C. P.; Rodrigues, J. P. G. L. M.; Trellet, M.; Schmitz, C.; Kastiris, P. L.; Karaca, E.; Melquiond, A. S. J.; Van Dijk, M.; De Vries, S. J.; Bonvin, A. M. J. J. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* 2016. <https://doi.org/10.1016/J.JMB.2015.09.014>.
- (331) Aurélien Grosdidier, Vincent Zoete, O. M. SwissDock, a Protein-Small Molecule Docking Web Service Based on EADock DSS. *Nucleic Acids Res.* 2011. <https://doi.org/10.1093/NAR/GKR366>.
- (332) Vuorio, J.; Vattulainen, I.; Martinez-Seara, H. Atomistic Fingerprint of Hyaluronan-CD44 Binding. *PLOS Comput. Biol.* 2017. <https://doi.org/10.1371/journal.pcbi.1005663>.
- (333) Samsonov, S. A.; Gehrcke, J. P.; Pisabarro, M. T. Flexibility and Explicit Solvent in Molecular-Dynamics-Based Docking of Protein-Glycosaminoglycan Systems. *J. Chem. Inf. Model.* 2014. <https://doi.org/10.1021/ci4006047>.
- (334) Jarzynski, C. Nonequilibrium Equality for Free Energy Differences. *Phys. Rev. Lett.* 1997. <https://doi.org/10.1103/PhysRevLett.78.2690>.
- (335) Samsonov, S. A.; Zacharias, M.; Chauvot de Beauchene, I. Modeling Large Protein-Glycosaminoglycan Complexes Using a Fragment-Based Approach. *J. Comput. Chem.* 2019. <https://doi.org/10.1002/jcc.25797>.
- (336) Siebenmorgen, T.; Engelhard, M.; Zacharias, M. Prediction of Protein-Protein Complexes Using Replica Exchange with Repulsive Scaling. *J. Comput. Chem.* 2020. <https://doi.org/10.1002/jcc.26187>.
- (337) Siebenmorgen, T.; Zacharias, M. Efficient Refinement and Free Energy Scoring of Predicted Protein-Protein Complexes Using Replica Exchange with Repulsive Scaling. *J. Chem. Inf. Model.* 2020. <https://doi.org/10.1021/acs.jcim.0c00853>.
- (338) Siebenmorgen, T.; Zacharias, M. Evaluation of Predicted Protein-Protein Complexes by Binding Free Energy Simulations. *J. Chem. Theory Comput.* 2019. <https://doi.org/10.1021/acs.jctc.8b01022>.
- (339) van Gunsteren, W. F.; Weiner, P. K.; Wilkinson, A. J. *Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications. Vol. 3*; Springer Science & Business Media, 2013. <https://doi.org/10.1007/978-94-017-1120-3>.
- (340) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications. Vol. 1*; Elsevier, 2001.
- (341) Karplus, M.; Petsko, G. A. Molecular Dynamics Simulations in Biology. *Nature.* 1990. <https://doi.org/10.1038/347631a0>.
- (342) Yang, L.; Tan, C. H.; Hsieh, M. J.; Wang, J.; Duan, Y.; Cieplak, P.; Caldwell, J.; Kollman, P. A.; Luo, R. New-Generation Amber United-Atom Force Field. *J. Phys. Chem. B* 2006. <https://doi.org/10.1021/jp060163v>.
- (343) Scott, W. R. P.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krüger, P.; Van Gunsteren, W. F. The GROMOS Biomolecular Simulation Program Package. *J. Phys. Chem. A* 1999. <https://doi.org/10.1021/jp984217f>.
- (344) Jorgensen, W. L.; Tirado-Rives, J. The OPLS Potential Functions for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J. Am. Chem. Soc.* 1988. <https://doi.org/10.1021/ja00214a001>.
- (345) Halgren, T. A. Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94. *J. Comput. Chem.* 1996. [https://doi.org/10.1002/\(SICI\)1096-987X\(199604\)17:5/6<490::AID-JCC1>3.0.CO;2-P](https://doi.org/10.1002/(SICI)1096-987X(199604)17:5/6<490::AID-JCC1>3.0.CO;2-P).
- (346) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How Fast-Folding Proteins Fold. *Science (80-. )*. 2011. <https://doi.org/10.1126/science.1208351>.
- (347) Feig, M.; Yu, I.; Wang, P. H.; Nawrocki, G.; Sugita, Y. Crowding in Cellular Environments at an Atomistic Level from Computer Simulations. *J. Phys. Chem. B* 2017. <https://doi.org/10.1021/acs.jpcc.7b03570>.
- (348) Takahashi, K. Z.; Nishimura, R.; Yamato, N.; Yasuoka, K.; Masubuchi, Y. Onset of Static and Dynamic Universality among Molecular Models of Polymers. *Sci. Rep.* 2017. <https://doi.org/10.1038/s41598-017-08501-0>.
- (349) Jarin, Z.; Newhouse, J.; Voth, G. A. Coarse-Grained Force Fields from the Perspective of Statistical Mechanics: Better Understanding of the Origins of a MARTINI Hangover. *J. Chem. Theory Comput.* 2021. <https://doi.org/10.1021/acs.jctc.0c00638>.
- (350) Hilpert, C.; Beranger, L.; Souza, P. C. T.; Vainikka, P. A.; Nieto, V.; Marrink, S. J.; Monticelli, L.; Launay, G. Facilitating CG Simulations with MAD: The MARTINI Database Server. *J. Chem. Inf. Model.* 2023. <https://doi.org/10.1021/acs.jcim.2c01375>.
- (351) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; De Vries, A. H. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* 2007. <https://doi.org/10.1021/jp071097f>.
- (352) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S. J. The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J. Chem. Theory Comput.* 2008. <https://doi.org/10.1021/ct700324x>.
- (353) Souza, P. C. T.; Alessandri, R.; Barnoud, J.; Thallmair, S.; Faustino, I.; Grünewald, F.; Patmanidis, I.; Abdizadeh, H.; Bruininks, B. M. H.; Wassenaar, T. A.; Kroon, P. C.; Melcr, J.; Nieto, V.; Corradi, V.; Khan, H. M.; Domański, J.; Javanainen, M.; Martinez-Seara, H.; Reuter, N.; Best, R. B.; Vattulainen, I.; Monticelli, L.; Periole, X.; Tieleman, D. P.; de Vries, A. H.; Marrink, S. J. Martini 3: A General Purpose Force Field for Coarse-Grained Molecular Dynamics. *Nat. Methods* 2021. <https://doi.org/10.1038/s41592-021-01098-3>.
- (354) Qi, Y.; Ingólfsson, H. I.; Cheng, X.; Lee, J.; Marrink, S. J.; Im, W. CHARMM-GUI Martini Maker for Coarse-Grained Simulations with the Martini Force Field. *J. Chem. Theory Comput.* 2015. <https://doi.org/10.1021/acs.jctc.5b00513>.
- (355) Maisuradze, G. G.; Senet, P.; Czaplewski, C.; Liwo, A.; Scheraga, H. A. Investigation of Protein Folding by Coarse-Grained Molecular Dynamics with the UNRES Force Field. *J. Phys. Chem. A* 2010. <https://doi.org/10.1021/jp9117776>.
- (356) Liwo, A.; Czaplewski, C.; Sieradzan, A. K.; Lubecka, E. A.; Lipska, A. G.; Golon, Ł.; Karczyńska, A.; Krupa, P.; Mozolewska, M. A.; Makowski, M.; Ganzynkiewicz, R.; Gieldoń, A.; Maciejczyk, M. Scale-Consistent Approach to the Derivation of Coarse-Grained Force Fields for Simulating Structure, Dynamics, and Thermodynamics of Biopolymers. In *Progress in Molecular Biology and Translational Science*; 2020. <https://doi.org/10.1016/bs.pmbts.2019.12.004>.
- (357) Kar, P.; Gopal, S. M.; Cheng, Y. M.; Panahi, A.; Feig, M. Transferring the PRIMO Coarse-Grained Force Field to the Membrane Environment: Simulations of Membrane Proteins and Helix-Helix Association. *J. Chem. Theory Comput.* 2014. <https://doi.org/10.1021/ct500443v>.
- (358) Basdevant, N.; Borgis, D.; Ha-Duong, T. Modeling Protein-Protein Recognition in Solution Using the Coarse-Grained Force Field SCORPION. *J. Chem. Theory Comput.* 2013. <https://doi.org/10.1021/ct300943w>.
- (359) Davtyan, A.; Schafer, N. P.; Zheng, W.; Clementi, C.; Wolynes, P. G.; Papoian, G. A. AWSEM-MD: Protein Structure Prediction Using Coarse-Grained Physical Potentials and Bioinformatically Based Local Structure Biasing. *J. Phys. Chem. B* 2012. <https://doi.org/10.1021/jp212541y>.
- (360) Darré, L.; Machado, M. R.; Brandner, A. F.; González, H. C.; Ferreira, S.; Pantano, S. SIRAH: A Structurally Unbiased Coarse-Grained Force Field for Proteins with Aqueous Solvation and Long-Range Electrostatics. *J. Chem. Theory Comput.* 2015. <https://doi.org/10.1021/ct5007746>.
- (361) Thompson, A. P.; Aktulga, H. M.; Berger, R.; Bolintineanu, D. S.; Brown, W. M.; Crozier, P. S.; in 't Veld, P. J.; Kohlmeyer, A.; Moore, S. G.; Nguyen, T. D.; Shan, R.; Stevens, M. J.; Tranchida, J.; Trott, C.; Plimpton, S. J. LAMMPS - a Flexible Simulation

- Tool for Particle-Based Materials Modeling at the Atomic, Meso, and Continuum Scales. *Comput. Phys. Commun.* 2022. <https://doi.org/10.1016/j.cpc.2021.108171>.
- (362) Foley, B. L.; Tessier, M. B.; Woods, R. J. Carbohydrate Force Fields. *Wiley Interdisciplinary Reviews: Computational Molecular Science*. 2012. <https://doi.org/10.1002/wcms.89>.
- (363) Sundararajan, P. R.; Rao, V. S. R. Theoretical Studies on the Conformation of Aldopyranoses. *Tetrahedron* 1968. [https://doi.org/10.1016/0040-4020\(68\)89030-1](https://doi.org/10.1016/0040-4020(68)89030-1).
- (364) Melberg, S.; Rasmussen, K. Potential Energy Function for Calculation of Structures, Vibrational Spectra and Thermodynamic Functions of Alkanes, Alcohols, Ethers and Carbohydrates. *J. Mol. Struct.* 1979. [https://doi.org/10.1016/0022-2860\(79\)80248-3](https://doi.org/10.1016/0022-2860(79)80248-3).
- (365) Melberg, S.; Rasmussen, K. Conformations of Disaccharides by Empirical, Force-Field Calculations: Part III,  $\beta$ -Gentiobiose. *Carbohydr. Res.* 1980. [https://doi.org/10.1016/0008-6215\(80\)90001-4](https://doi.org/10.1016/0008-6215(80)90001-4).
- (366) Tvaroška, I. An Attempt to Derive the Potential Function for Evaluation of the Energy Associated with the Exo-Anomeric Effect. *Carbohydr. Res.* 1984. [https://doi.org/10.1016/0008-6215\(84\)85150-2](https://doi.org/10.1016/0008-6215(84)85150-2).
- (367) Ott, K. H.; Meyer, B. Parametrization of GROMOS Force Field for Oligosaccharides and Assessment of Efficiency of Molecular Dynamics Simulations. *J. Comput. Chem.* 1996. [https://doi.org/10.1002/\(SICI\)1096-987X\(199606\)17:8<1068::AID-JCC14>3.0.CO;2-A](https://doi.org/10.1002/(SICI)1096-987X(199606)17:8<1068::AID-JCC14>3.0.CO;2-A).
- (368) Grootenhuys, P. D. J.; Haasnoot, C. A. G. A Charmm Based Force Field for Carbohydrates Using the Cheat Approach: Carbohydrate Hydroxyl Groups Represented by Extended Atoms. *Mol. Simul.* 1993. <https://doi.org/10.1080/08927029308022160>.
- (369) Homans, S. W. A Molecular Mechanical Force Field for the Conformational Analysis of Oligosaccharides: Comparison of Theoretical and Crystal Structures of Man $\alpha$ 1-3Man $\beta$ 1-4GlcNAc. *Biochemistry* 1990. <https://doi.org/10.1021/bi00491a003>.
- (370) Woods, R. J.; Dwek, R. A.; Edge, C. J.; Fraser-Reid, B. Molecular Mechanical and Molecular Dynamical Simulations of Glycoproteins and Oligosaccharides. I. GLYCAM\_93 Parameter Development. *J. Phys. Chem.* 1995. <https://doi.org/10.1021/j100011a061>.
- (371) Nagarajan, B.; Sankaranarayanan, N. V.; Desai, U. R. Perspective on Computational Simulations of Glycosaminoglycans. *Wiley Interdisciplinary Reviews: Computational Molecular Science*. 2019. <https://doi.org/10.1002/wcms.1388>.
- (372) Winter, W. T.; Smith, P. J. C.; Arnott, S. Hyaluronic Acid: Structure of a Fully Extended 3-Fold Helical Sodium Salt and Comparison with the Less Extended 4-Fold Helical Forms. *J. Mol. Biol.* 1975. [https://doi.org/10.1016/S0022-2836\(75\)80142-2](https://doi.org/10.1016/S0022-2836(75)80142-2).
- (373) Verli, H.; Guimarães, J. A. Molecular Dynamics Simulation of a Decasaccharide Fragment of Heparin in Aqueous Solution. *Carbohydr. Res.* 2004. <https://doi.org/10.1016/j.carres.2003.09.026>.
- (374) Schüttelkopf, A. W.; Van Aalten, D. M. F. PRODRG: A Tool for High-Throughput Crystallography of Protein-Ligand Complexes. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 2004. <https://doi.org/10.1107/S0907444904011679>.
- (375) Gandhi, N. S.; Mancera, R. L. Can Current Force Fields Reproduce Ring Puckering in 2-O-Sulfo- $\alpha$ -l-Iduronic Acid? A Molecular Dynamics Simulation Study. *Carbohydr. Res.* 2010. <https://doi.org/10.1016/j.carres.2009.12.020>.
- (376) Cilpa, G.; Hyvönen, M. T.; Koivuniemi, A.; Riekkola, M. L. Atomistic Insight into Chondroitin-6-Sulfate Glycosaminoglycan Chain through Quantum Mechanics Calculations and Molecular Dynamics Simulation. *J. Comput. Chem.* 2010. <https://doi.org/10.1002/jcc.21453>.
- (377) Scott, J. E.; Cummings, C.; Brass, A.; Chen, Y. Secondary and Tertiary Structures of Hyaluronan in Aqueous Solution, Investigated by Rotary Shadowing-Electron Microscopy and Computer Simulation. Hyaluronan Is a Very Efficient Network-Forming Polymer. *Biochem. J.* 1991. <https://doi.org/10.1042/bj2740699>.
- (378) Almond, A.; Brass, A.; Sheehan, J. K. Deducing Polymeric Structure from Aqueous Molecular Dynamics Simulations of Oligosaccharides: Predictions from Simulations of Hyaluronan Tetrasaccharides Compared with Hydrodynamic and X-Ray Fibre Diffraction Data. *J. Mol. Biol.* 1998. <https://doi.org/10.1006/jmbi.1998.2245>.
- (379) Almond, A.; Brass, A.; Sheehan, J. K. Dynamic Exchange between Stabilized Conformations Predicted for Hyaluronan Tetrasaccharides: Comparison of Molecular Dynamics Simulations with Available NMR Data. *Glycobiology* 1998. <https://doi.org/10.1093/glycob/8.10.973>.
- (380) Almond, A.; Sheehan, J. K.; Brass, A. Molecular Dynamics Simulations of the Two Disaccharides of Hyaluronan in Aqueous Solution. *Glycobiology* 1997. <https://doi.org/10.1093/glycob/7.5.597>.
- (381) Wiegel, D.; Kaufmann, J.; Arnold, K. Polar Interactions of Chondroitinsulfate: Surface Free Energy and Molecular Dynamics Simulations. *Colloids Surfaces B Biointerfaces* 1999. [https://doi.org/10.1016/S0927-7765\(98\)00115-5](https://doi.org/10.1016/S0927-7765(98)00115-5).
- (382) Kaufmann, J.; Möhle, K.; Hofmann, H. J.; Arnold, K. Molecular Dynamics of a Tetrasaccharide Subunit of Chondroitin 4-Sulfate in Water. *Carbohydr. Res.* 1999. [https://doi.org/10.1016/S0008-6215\(99\)00091-9](https://doi.org/10.1016/S0008-6215(99)00091-9).
- (383) Ng, C.; Nandha Premnath, P.; Guvench, O. Rigidity and Flexibility in the Tetrasaccharide Linker of Proteoglycans from Atomic-Resolution Molecular Simulation. *J. Comput. Chem.* 2017. <https://doi.org/10.1002/jcc.24738>.
- (384) Paiardi, G.; Milanesi, M.; Wade, R. C.; D'ursi, P.; Rusnati, M. A Bittersweet Computational Journey among Glycosaminoglycans. *Biomolecules* 2021. <https://doi.org/10.3390/biom11050739>.
- (385) Kirschner, K. N.; Woods, R. J. Solvent Interactions Determine Carbohydrate Conformation. *Proc. Natl. Acad. Sci. U. S. A.* 2001. <https://doi.org/10.1073/pnas.191362798>.
- (386) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. GLYCAM06: A Generalizable Biomolecular Force Field. Carbohydrates. *J. Comput. Chem.* 2008. <https://doi.org/10.1002/jcc.20820>.
- (387) Forster, M. J.; Mulloy, B. Molecular Dynamics Study of Iduronate Ring Conformation. *Biopolymers* 1993. <https://doi.org/10.1002/bip.360330407>.
- (388) Muñoz-García, J. C.; Solera, C.; Carrero, P.; De Paz, J. L.; Angulo, J.; Nieto, P. M. 3D Structure of a Heparin Mimetic Analogue of a FGF-1 Activator. A NMR and Molecular Modelling Study. *Org. Biomol. Chem.* 2013. <https://doi.org/10.1039/c3ob41789a>.
- (389) Oka, N.; Mori, S.; Ikegaya, M.; Park, E. Y.; Miyazak, T. Crystal Structure and Sugar-Binding Ability of the C-Terminal Domain of N-Acetylglucosaminyltransferase IV Establish a New Carbohydrate-Binding Module Family. *Glycobiology* 2022. <https://doi.org/10.1093/glycob/ewac058>.
- (390) Muñoz-García, J. C.; López-Prados, J.; Angulo, J.; Díaz-Contreras, I.; Reichardt, N.; De Paz, J. L.; Martín-Lomas, M.; Nieto, P. M. Effect of the Substituents of the Neighboring Ring in the Conformational Equilibrium of Iduronate in Heparin-like Trisaccharides. *Chem. - A Eur. J.* 2012. <https://doi.org/10.1002/chem.201202770>.
- (391) Rodríguez-Carvajal, M. A.; Imberty, A.; Pérez, S. Conformational Behavior of Chondroitin and Chondroitin Sulfate in Relation to Their Physical Properties as Inferred by Molecular Modeling. *Biopolymers* 2003. <https://doi.org/10.1002/bip.10304>.
- (392) Singh, A.; Tessier, M. B.; Pederson, K.; Wang, X.; Venot, A. P.; Boons, G.-J.; Prestegard, J. H.; Woods, R. J. Extension and Validation of the GLYCAM Force Field Parameters for Modeling Glycosaminoglycans. *Can. J. Chem.* 2016. <https://doi.org/10.1139/cjc-2015-0606>.
- (393) Hsieh, P. H.; Thieker, D. F.; Guerrini, M.; Woods, R. J.; Liu, J. Uncovering the Relationship between Sulphation Patterns and Conformation of Iduronic Acid in Heparan Sulphate. *Sci. Rep.* 2016. <https://doi.org/10.1038/srep29602>.

- (394) 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. <https://doi.org/10.3390/ijms20205021>.
- (395) Ruiz-Gómez, G.; Vogel, S.; Möller, S.; Pisabarro, M. T.; Hempel, U. Glycosaminoglycans Influence Enzyme Activity of MMP2 and MMP2/TIMP3 Complex Formation - Insights at Cellular and Molecular Level. *Sci. Rep.* 2019. <https://doi.org/10.1038/s41598-019-41355-2>.
- (396) Bojarski, K. K.; Karczyńska, A. S.; Samsonov, S. A. Role of Glycosaminoglycans in Procathepsin B Maturation: Molecular Mechanism Elucidated by a Computational Study. *J. Chem. Inf. Model.* 2020. <https://doi.org/10.1021/acs.jcim.0c00023>.
- (397) Perez, S.; Makshakova, O. Multifaceted Computational Modeling in Glycoscience. *Chemical Reviews.* 2022. <https://doi.org/10.1021/acs.chemrev.2c00060>.
- (398) Molinero, V.; Goddard, W. A. M3B: A Coarse Grain Force Field for Molecular Simulations of Malto-Oligosaccharides and Their Water Mixtures. *J. Phys. Chem. B* 2004. <https://doi.org/10.1021/jp0354752>.
- (399) Bathe, M.; Rutledge, G. C.; Grodzinsky, A. J.; Tidor, B. A Coarse-Grained Molecular Model for Glycosaminoglycans: Application to Chondroitin, Chondroitin Sulfate, and Hyaluronic Acid. *Biophys. J.* 2005. <https://doi.org/10.1529/biophysj.104.058800>.
- (400) Sattelle, B. M.; Shakeri, J.; Almond, A. Does Microsecond Sugar Ring Flexing Encode 3D-Shape and Bioactivity in the Heparanome? *Biomacromolecules* 2013. <https://doi.org/10.1021/bm400067g>.
- (401) Sattelle, B. M.; Shakeri, J.; Cliff, M. J.; Almond, A. Proteoglycans and Their Heterogeneous Glycosaminoglycans at the Atomic Scale. *Biomacromolecules* 2015. <https://doi.org/10.1021/bm5018386>.
- (402) Kolesnikov, A. L.; Budkov, Y. A.; Nogovitsyn, E. A. Coarse-Grained Model of Glycosaminoglycans in Aqueous Salt Solutions. A Field-Theoretical Approach. *J. Phys. Chem. B* 2014. <https://doi.org/10.1021/jp503749a>.
- (403) López, C. A.; Bellesia, G.; Redondo, A.; Langan, P.; Chundawat, S. P. S.; Dale, B. E.; Marrink, S. J.; Gnanakaran, S. MARTINI Coarse-Grained Model for Crystalline Cellulose Microfibers. *J. Phys. Chem. B* 2015. <https://doi.org/10.1021/jp5105938>.
- (404) Lubecka, E. A.; Liwo, A. A General Method for the Derivation of the Functional Forms of the Effective Energy Terms in Coarse-Grained Energy Functions of Polymers. II. Backbone-Local Potentials of Coarse-Grained O 1 → 4 -Bonded Polyglucose Chains. *J. Chem. Phys.* 2017. <https://doi.org/10.1063/1.4994130>.
- (405) Zhang, W.; Howell, S. C.; Wright, D. W.; Heindel, A.; Qiu, X.; Chen, J.; Curtis, J. E. Combined Monte Carlo/Torsion-Angle Molecular Dynamics for Ensemble Modeling of Proteins, Nucleic Acids and Carbohydrates. *J. Mol. Graph. Model.* 2017. <https://doi.org/10.1016/j.jmgm.2017.02.010>.
- (406) Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S. Alchemical Free Energy Methods for Drug Discovery: Progress and Challenges. *Current Opinion in Structural Biology.* 2011. <https://doi.org/10.1016/j.sbi.2011.01.011>.
- (407) Abel, R.; Wang, L.; Harder, E. D.; Berne, B. J.; Friesner, R. A. Advancing Drug Discovery through Enhanced Free Energy Calculations. *Acc. Chem. Res.* 2017. <https://doi.org/10.1021/acs.accounts.7b00083>.
- (408) Kollman, P. Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* 1993. <https://doi.org/10.1021/cr00023a004>.
- (409) Perez, A.; Morrone, J. A.; Simmerling, C.; Dill, K. A. Advances in Free-Energy-Based Simulations of Protein Folding and Ligand Binding. *Current Opinion in Structural Biology.* 2016. <https://doi.org/10.1016/j.sbi.2015.12.002>.
- (410) Zwanzig, R. W. High $\beta$ ;  $\frac{1}{2}$ Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J. Chem. Phys.* 1954. <https://doi.org/10.1063/1.1740409>.
- (411) Wang, L.; Chambers, J.; Abel, R. Protein–Ligand Binding Free Energy Calculations with FEP+. In *Methods in Molecular Biology*; 2019. [https://doi.org/10.1007/978-1-4939-9608-7\\_9](https://doi.org/10.1007/978-1-4939-9608-7_9).
- (412) Bennett, C. H. Efficient Estimation of Free Energy Differences from Monte Carlo Data. *J. Comput. Phys.* 1976. [https://doi.org/10.1016/0021-9991\(76\)90078-4](https://doi.org/10.1016/0021-9991(76)90078-4).
- (413) Kollman, P. A.; 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, D. A.; Cheatham, T. E. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* 2000. <https://doi.org/10.1021/ar000033j>.
- (414) Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate-DNA Helices. *J. Am. Chem. Soc.* 1998. <https://doi.org/10.1021/ja981844+>.
- (415) Yamato, T.; Lapr evote, O. Normal Mode Analysis and Beyond. *Biophysics and physcobiology.* 2019. [https://doi.org/10.2142/biophysico.16.0\\_322](https://doi.org/10.2142/biophysico.16.0_322).
- (416) Rodgers, J. M.; Hemley, R. J.; Ichiye, T. Quasi-harmonic Analysis of Protein Energy Landscapes from Pressure-Temperature Molecular Dynamics Simulations. *J. Chem. Phys.* 2017. <https://doi.org/10.1063/1.5003823>.
- (417) Åqvist, J.; Medina, C.; Samuelsson, J. E. A New Method for Predicting Binding Affinity in Computer-Aided Drug Design. *Protein Eng. Des. Sel.* 1994. <https://doi.org/10.1093/protein/7.3.385>.
- (418) Frederick, S.; Chu, Z. T.; Michael, B.; Warshel, A. Calculations of Antibody-Antigen Interactions: Microscopic and Semi-Microscopic Evaluation of the Free Energies of Binding of Phosphorylcholine Analogs to McPC603. *Protein Eng. Des. Sel.* 1992. <https://doi.org/10.1093/protein/5.3.215>.
- (419) Reif, M. M.; Zacharias, M. Improving the Potential of Mean Force and Nonequilibrium Pulling Simulations by Simultaneous Alchemical Modifications. *J. Chem. Theory Comput.* 2022. <https://doi.org/10.1021/acs.jctc.1c01194>.
- (420) Bojarski, K. K.; Becher, J.; Riemer, T.; Lemmnitzer, K.; Möller, S.; Schiller, J.; Schnabelrauch, M.; Samsonov, S. A. Synthesis and in Silico Characterization of Artificially Phosphorylated Glycosaminoglycans. *J. Mol. Struct.* 2019. <https://doi.org/10.1016/j.molstruc.2019.07.064>.
- (421) Roux, B. The Calculation of the Potential of Mean Force Using Computer Simulations. *Comput. Phys. Commun.* 1995. [https://doi.org/10.1016/0010-4655\(95\)00053-1](https://doi.org/10.1016/0010-4655(95)00053-1).
- (422) Jafari, S.; Ryde, U.; Irani, M. Two-Substrate Glyoxalase I Mechanism: A Quantum Mechanics/Molecular Mechanics Study. *Inorg. Chem.* 2020. <https://doi.org/10.1021/ACS.INORGCHEM.0C02957>.
- (423) Khandelwal, A.; Lukacova, V.; Comez, D.; Kroll, D. M.; Raha, S.; Balaz, S. A Combination of Docking, QM/MM Methods, and MD Simulation for Binding Affinity Estimation of Metalloprotein Ligands. *J. Med. Chem.* 2005. <https://doi.org/10.1021/jm049050v>.
- (424) Möbius, K.; Nordsieck, K.; Pichert, A.; Samsonov, S. A.; Thomas, L.; Schiller, J.; Kalkhof, S.; Teresa Pisabarro, M.; Beck-Sickingler, A. G.; Huster, D. Investigation of Lysine Side Chain Interactions of Interleukin-8 with Heparin and Other Glycosaminoglycans Studied by a Methylation-NMR Approach. *Glycobiology* 2013. <https://doi.org/10.1093/glycob/cwt062>.
- (425) Samsonov, S. A.; Pisabarro, M. T. Importance of IdoA and IdoA2S) Ring Conformations in Computational Studies of Glycosaminoglycan-Protein Interactions. *Carbohydr. Res.* 2013, 381, 133–137. <https://doi.org/10.1016/j.carres.2013.09.005>.
- (426) Zsila, F.; Samsonov, S. A. Molecular Interactions of the Anticancer Agent Ellipticine with Glycosaminoglycans by in Silico Analysis. *Carbohydr. Res.* 2018, 462, 28–33. <https://doi.org/10.1016/j.carres.2018.03.014>.
- (427) Hintze, V.; Samsonov, S. A.; Anselmi, M.; Moeller, S.; Becher, J.; Schnabelrauch, M.; Scharnweber, D.; Pisabarro, M. T. Sulfated

- Glycosaminoglycans Exploit the Conformational Plasticity of Bone Morphogenetic Protein-2 (BMP-2) and Alter the Interaction Profile with Its Receptor. *Biomacromolecules* 2014. <https://doi.org/10.1021/bm5006855>.
- (428) Salbach-Hirsch, J.; Samsonov, S. A.; Hintze, V.; Hofbauer, C.; Picke, A. K.; Rauner, M.; Gehrcke, J. P.; Moeller, S.; Schnabelrauch, M.; Scharnweber, D.; Pisabarro, M. T.; Hofbauer, L. C. Structural and Functional Insights into Sclerostin-Glycosaminoglycan Interactions in Bone. *Biomaterials* 2015. <https://doi.org/10.1016/j.biomaterials.2015.07.021>.
- (429) Cochran, S.; Li, C. P.; Bytheway, I. An Experimental and Molecular-Modeling Study of the Binding of Linked Sulfated Tetracyclitols to FGF-1 and FGF-2. *ChemBioChem* 2005. <https://doi.org/10.1002/cbic.200500089>.
- (430) Kogut, M. M.; Maszota-Zieleniak, M.; Marcisz, M.; Samsonov, S. A. Computational Insights into the Role of Calcium Ions in Protein-Glycosaminoglycan Systems. *Phys. Chem. Chem. Phys.* 2021. <https://doi.org/10.1039/d0cp05438k>.
- (431) Plazinski, W.; Knys-Dziewciuch, A. Interactions between CD44 Protein and Hyaluronan: Insights from the Computational Study. *Mol. Biosyst.* 2012. <https://doi.org/10.1039/c2mb05399c>.
- (432) Babik, S.; Samsonov, S. A.; Pisabarro, M. T. Computational Drill down on FGF1-Heparin Interactions through Methodological Evaluation. *Glycoconj. J.* 2017. <https://doi.org/10.1007/s10719-016-9745-4>.
- (433) Zuckerman, D. M. Equilibrium Sampling in Biomolecular Simulations. *Annu. Rev. Biophys.* 2011. <https://doi.org/10.1146/annurev-biophys-042910-155255>.
- (434) Anandakrishnan, R.; Drozdetski, A.; Walker, R. C.; Onufriev, A. V. Speed of Conformational Change: Comparing Explicit and Implicit Solvent Molecular Dynamics Simulations. *Biophys. J.* 2015. <https://doi.org/10.1016/j.bpj.2014.12.047>.
- (435) David, L.; Luo, R.; Gilson, M. K. Comparison of Generalized Born and Poisson Models: Energetics and Dynamics of HIV Protease. *J. Comput. Chem.* 2000. [https://doi.org/10.1002/\(SICI\)1096-987X\(200003\)21:4<295::AID-JCC5>3.0.CO;2-8](https://doi.org/10.1002/(SICI)1096-987X(200003)21:4<295::AID-JCC5>3.0.CO;2-8).
- (436) Feig, M. Kinetics from Implicit Solvent Simulations of Biomolecules as a Function of Viscosity. *J. Chem. Theory Comput.* 2007. <https://doi.org/10.1021/ct7000705>.
- (437) Tsui, V.; Case, D. A. Molecular Dynamics Simulations of Nucleic Acids with a Generalized Born Solvation Model. *J. Am. Chem. Soc.* 2000. <https://doi.org/10.1021/ja9939385>.
- (438) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. The GB/SA Continuum Model for Solvation. A Fast Analytical Method for the Calculation of Approximate Born Radii. *J. Phys. Chem. A* 1997. <https://doi.org/10.1021/jp961992r>.
- (439) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Pairwise Solute Descreening of Solute Charges from a Dielectric Medium. *Chem. Phys. Lett.* 1995. [https://doi.org/10.1016/0009-2614\(95\)01082-K](https://doi.org/10.1016/0009-2614(95)01082-K).
- (440) Onufriev, A.; Bashford, D.; Case, D. A. Modification of the Generalized Born Model Suitable for Macromolecules. *J. Phys. Chem. B* 2000. <https://doi.org/10.1021/jp994072s>.
- (441) Onufriev, A.; Bashford, D.; Case, D. A. Exploring Protein Native States and Large-Scale Conformational Changes with a Modified Generalized Born Model. *Proteins* 2004, 55 (2), 383–394. <https://doi.org/10.1002/prot.20033>.
- (442) Mongan, J.; Simmerling, C.; McCammon, J. A.; Case, D. A.; Onufriev, A. Generalized Born Model with a Simple, Robust Molecular Volume Correction. *J. Chem. Theory Comput.* 2007. <https://doi.org/10.1021/ct600085e>.
- (443) Nguyen, H.; Roe, D. R.; Simmerling, C. Improved Generalized Born Solvent Model Parameters for Protein Simulations. *J. Chem. Theory Comput.* 2013. <https://doi.org/10.1021/ct3010485>.
- (444) D.A. Case, R.M. Betz, W. Botello-Smith, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. M. Y. and P. A. K. M. Y.; Kollman, P. A. Amber 16. *University of California, San Francisco*. 2016. <https://doi.org/10.1002/jcc.23031>.
- (445) Geney, R.; Layten, M.; Gomperts, R.; Hornak, V.; Simmerling, C. Investigation of Salt Bridge Stability in a Generalized Born Solvent Model. *J. Chem. Theory Comput.* 2006. <https://doi.org/10.1021/ct050183l>.
- (446) Okur, A.; Wickstrom, L.; Simmerling, C. Evaluation of Salt Bridge Structure and Energetics in Peptides Using Explicit, Implicit, and Hybrid Solvation Models. *J. Chem. Theory Comput.* 2008. <https://doi.org/10.1021/ct7002308>.
- (447) Okur, A.; Wickstrom, L.; Layten, M.; Geney, R.; Song, K.; Hornak, V.; Simmerling, C. Improved Efficiency of Replica Exchange Simulations through Use of a Hybrid Explicit/Implicit Solvation Model. *J. Chem. Theory Comput.* 2006. <https://doi.org/10.1021/ct050196z>.
- (448) Roe, D. R.; Okur, A.; Wickstrom, L.; Hornak, V.; Simmerling, C. Secondary Structure Bias in Generalized Born Solvent Models: Comparison of Conformational Ensembles and Free Energy of Solvent Polarization from Explicit and Implicit Solvation. *J. Phys. Chem. B* 2007. <https://doi.org/10.1021/jp066831u>.
- (449) Medina, J. S.; Prossniti, R.; Villarreal, P.; Delgado-Barrio, G.; Winter, G.; González, B.; Alemán, J. V.; Collado, C. Molecular Dynamics Simulations of Rigid and Flexible Water Models: Temperature Dependence of Viscosity. *Chem. Phys.* 2011. <https://doi.org/10.1016/j.chemphys.2011.07.001>.
- (450) Dang, L. X.; Pettitt, B. M. Simple Intramolecular Model Potentials for Water. *J. Phys. Chem.* 1987. <https://doi.org/10.1021/j100296a048>.
- (451) Wu, Y.; Tepper, H. L.; Voth, G. A. Flexible Simple Point-Charge Water Model with Improved Liquid-State Properties. *J. Chem. Phys.* 2006. <https://doi.org/10.1063/1.2136877>.
- (452) Yesylevskyy, S. O.; Schäfer, L. V.; Sengupta, D.; Marrink, S. J. Polarizable Water Model for the Coarse-Grained MARTINI Force Field. *PLoS Comput. Biol.* 2010. <https://doi.org/10.1371/journal.pcbi.1000810>.
- (453) Nada, H.; Van Der Eerden, J. P. J. M. An Intermolecular Potential Model for the Simulation of Ice and Water near the Melting Point: A Six-Site Model of H<sub>2</sub>O. *J. Chem. Phys.* 2003. <https://doi.org/10.1063/1.1562610>.
- (454) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* 1983. <https://doi.org/10.1063/1.445869>.
- (455) 37949 citations on 26.05.2023  
[https://scholar.google.com/scholar?hl=pl&as\\_sdt=0%2C5&q=Comparison+of+Simple+Potential+Functions+for+Simulating+Liquid+Water&btnG=](https://scholar.google.com/scholar?hl=pl&as_sdt=0%2C5&q=Comparison+of+Simple+Potential+Functions+for+Simulating+Liquid+Water&btnG=)
- (456) Sapay, N.; Cabannes, E.; Petitou, M.; Imberty, A. Molecular Modeling of the Interaction between Heparan Sulfate and Cellular Growth Factors: Bringing Pieces Together. *Glycobiology* 2011. <https://doi.org/10.1093/glycob/cwr052>.
- (457) Nagarajan, B.; Sankaranarayanan, N. V.; Patel, B. B.; Desai, U. R. A Molecular Dynamics-Based Algorithm for Evaluating the Glycosaminoglycan Mimicking Potential of Synthetic, Homogenous, Sulfated Small Molecules. *PLoS One* 2017. <https://doi.org/10.1371/journal.pone.0171619>.
- (458) Singh, A.; Kett, W. C.; Severin, I. C.; Agyekum, I.; Duan, J.; Amster, I. J.; Proudfoot, A. E. I.; Coombe, D. R.; Woods, R. J. The Interaction of Heparin Tetrasaccharides with Chemokine CCL5 Is Modulated by Sulfation Pattern and PH. *J. Biol. Chem.* 2015. <https://doi.org/10.1074/jbc.M115.655845>.
- (459) Roy, R.; Jonniya, N. A.; Kar, P. Effect of Sulfation on the Conformational Dynamics of Dermatan Sulfate Glycosaminoglycan: A Gaussian Accelerated Molecular Dynamics Study. *J. Phys. Chem. B* 2022, 126 (21), 3852–3866. <https://doi.org/10.1021/acs.jpcc.2c01807>.




- (460) Lang, E. J. M.; Baker, E. G.; Woolfson, D. N.; Mulholland, A. J. Generalized Born Implicit Solvent Models Do Not Reproduce Secondary Structures of de Novo Designed Glu/Lys Peptides. *J. Chem. Theory Comput.* 2022, 18 (7), 4070–4076. <https://doi.org/10.1021/acs.jctc.1c01172>.
- (461) Samsonov, S.; Teyra, J.; Pisabarro, M. T. A Molecular Dynamics Approach to Study the Importance of Solvent in Protein Interactions. *Proteins Struct. Funct. Genet.* 2008. <https://doi.org/10.1002/prot.22076>.
- (462) Marcisz, M.; Gaardl0s, M.; Bojarski, K. K.; Siebenmorgen, T.; Zacharias, M.; Samsonov, S. A. Explicit Solvent Repulsive Scaling Replica Exchange Molecular Dynamics (RS-REMD) in Molecular Modeling of Protein Glycosaminoglycan Complexes. *J. Comput. Chem.* 2022. <https://doi.org/10.1002/jcc.26965>.
- (463) Izadi, S.; Onufriev, A. V. Accuracy Limit of Rigid 3-Point Water Models. *J. Chem. Phys.* 2016. <https://doi.org/10.1063/1.4960175>.
- (464) Izadi, S.; Anandakrishnan, R.; Onufriev, A. V. Building Water Models: A Different Approach. *J. Phys. Chem. Lett.* 2014. <https://doi.org/10.1021/jz501780a>.
- (465) Neamtu, A.; Tamba, B.; Patras, X. Cellulose Chemistry and Technology Molecular Dynamics Simulations of Chondroitin Sulfate in Explicit Solvent: Point Charge Water Models Compared. *Cellul. Chem. Technol.* 2013.
- (466) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U. S. A.* 2002. <https://doi.org/10.1073/pnas.202427399>.
- (467) Ostermeir, K.; Zacharias, M. Hamiltonian Replica-Exchange Simulations with Adaptive Biasing of Peptide Backbone and Side Chain Dihedral Angles. *J. Comput. Chem.* 2014. <https://doi.org/10.1002/jcc.23476>.
- (468) Maszota-Zieleniak, M.; Marcisz, M.; Kogut, M. M.; Siebenmorgen, T.; Zacharias, M.; Samsonov, S. A. Evaluation of Replica Exchange with Repulsive Scaling Approach for Docking Glycosaminoglycans. *J. Comput. Chem.* 2021. <https://doi.org/10.1002/jcc.26496>.
- (469) Sankaranarayanan, N. V.; Nagarajan, B.; Desai, U. R. Combinatorial Virtual Library Screening Study of Transforming Growth Factor-B2–Chondroitin Sulfate System. *Int. J. Mol. Sci.* 2021. <https://doi.org/10.3390/ijms22147542>.
- (470) Roberts, B. C.; Mancera, R. L. Ligand - Protein Docking with Water Molecules. *J. Chem. Inf. Model.* 2008. <https://doi.org/10.1021/ci700285e>.
- (471) Thilagavathi, R.; Mancera, R. L. Ligand-Protein Cross-Docking with Water Molecules. *J. Chem. Inf. Model.* 2010. <https://doi.org/10.1021/ci900345h>.
- (472) van Dijk, A. D. J.; Bonvin, A. M. J. J. Solvated Docking: Introducing Water into the Modelling of Biomolecular Complexes. *Bioinformatics* 2006. <https://doi.org/10.1093/bioinformatics/btl395>.
- (473) Baron, R.; Setny, P.; Andrew McCammon, J. Water in Cavity-Ligand Recognition. *J. Am. Chem. Soc.* 2010. <https://doi.org/10.1021/ja1050082>.
- (474) Kimberley, F. C.; Van Bostelen, L.; Cameron, K.; Hardenberg, G.; Marquart, J. A.; Hahne, M.; Medema, J. P. The Proteoglycan (Heparan Sulfate Proteoglycan) Binding Domain of APRIL Serves as a Platform for Ligand Multimerization and Cross-linking. *FASEB J.* 2009. <https://doi.org/10.1096/fj.08-124669>.
- (475) Baert, L.; Manfroi, B.; Casez, O.; Sturm, N.; Huard, B. The Role of APRIL - A Proliferation Inducing Ligand - In Autoimmune Diseases and Expectations from Its Targeting. *Journal of Autoimmunity.* 2018. <https://doi.org/10.1016/j.jaut.2018.10.016>.
- (476) Kimberley, F. C.; Medema, J. P.; Hahne, M. APRIL in B-Cell Malignancies and Autoimmunity. *Results Probl. Cell Differ.* 2009. [https://doi.org/10.1007/400\\_2008\\_19](https://doi.org/10.1007/400_2008_19).
- (477) Baert, L.; Benkhoucha, M.; Popa, N.; Ahmed, M. C.; Manfroi, B.; Boutonnat, J.; Sturm, N.; Raguenez, G.; Tessier, M.; Casez, O.; Marignier, R.; Ahmadi, M.; Broisat, A.; Ghezzi, C.; Rivat, C.; Sonrier, C.; Hahne, M.; Baeten, D.; Vives, R. R.; Lortat-Jacob, H.; Marche, P. N.; Schneider, P.; Lassmann, H. P.; Boucraut, J.; Lalive, P. H.; Huard, B. A Proliferation-Inducing Ligand-Mediated Anti-Inflammatory Response of Astrocytes in Multiple Sclerosis. *Ann. Neurol.* 2019. <https://doi.org/10.1002/ana.25415>.
- (478) Marcisz, M.; Huard, B.; Lipska, A. G.; Samsonov, S. A. Further Analyses of APRIL/APRIL-Receptor/Glycosaminoglycan Interactions by Biochemical Assays Linked to Computational Studies. *Glycobiology* 2021. <https://doi.org/10.1093/glycob/cwab016>.
- (479) Hendriks, J.; Planelles, L.; de Jong-Odding, J.; Hardenberg, G.; Pals, S. T.; Hahne, M.; Spaargaren, M.; Medema, J. P. Heparan Sulfate Proteoglycan Binding Promotes APRIL-Induced Tumor Cell Proliferation. *Cell Death Differ.* 2005. <https://doi.org/10.1038/sj.cdd.4401647>.
- (480) Ingold, K.; Zumsteg, A.; Tardivel, A.; Huard, B.; Steiner, Q. G.; Cachero, T. G.; Qiang, F.; Gorelik, L.; Kalled, S. L.; Acha-Orbea, H.; Rennert, P. D.; Tschopp, J.; Schneider, P. Identification of Proteoglycans as the APRIL-Specific Binding Partners. *J. Exp. Med.* 2005. <https://doi.org/10.1084/jem.20042309>.
- (481) Moreaux, J.; Sprynski, A. C.; Dillon, S. R.; Mahtouk, K.; Jourdan, M.; Ythier, A.; Moine, P.; Robert, N.; Jourdan, E.; Rossi, J. F.; Klein, B. APRIL and TACI Interact with Syndecan-1 on the Surface of Multiple Myeloma Cells to Form an Essential Survival Loop. *Eur. J. Haematol.* 2009. <https://doi.org/10.1111/j.1600-0609.2009.01262.x>.
- (482) Bischof, D.; Elsawa, S. F.; Mantchev, G.; Yoon, J.; Michels, G. E.; Nilson, A.; Sutor, S. L.; Platt, J. L.; Ansell, S. M.; Von Bulow, G.; Bram, R. J. Selective Activation of TACI by Syndecan-2. *Blood* 2006. <https://doi.org/10.1182/blood-2005-01-0256>.
- (483) Sakurai, D.; Hase, H.; Kanno, Y.; Kojima, H.; Okumura, K.; Kobata, T. TACI Regulates IgA Production by APRIL in Collaboration with HSPG. *Blood* 2007. <https://doi.org/10.1182/blood-2006-08-041772>.
- (484) Kowalczyk-Quintas, C.; Willen, D.; Willen, L.; Golob, M.; Schuepbach-Mallepell, S.; Peter, B.; Eslami, M.; Vigolo, M.; Broly, H.; Samy, E.; Yalkinoglu, 0.; Schneider, P. No Interactions between Heparin and Atacept, an Antagonist of B Cell Survival Cytokines. *Br. J. Pharmacol.* 2019. <https://doi.org/10.1111/bph.14811>.

# APPENDIX

## Publication D1

Evaluation of replica exchange with  
repulsive scaling approach for docking  
glycosaminoglycans

# Evaluation of replica exchange with repulsive scaling approach for docking glycosaminoglycans

Martyna Maszota-Zieleniak<sup>1</sup> | Mateusz Marcisz<sup>1</sup> | Małgorzata M. Kogut<sup>1</sup> |  
Till Siebenmorgen<sup>2</sup> | Martin Zacharias<sup>2</sup> | Sergey A. Samsonov<sup>1</sup> 

<sup>1</sup>Faculty of Chemistry, University of Gdańsk, Gdańsk, Poland

<sup>2</sup>Physics Department, Technical University of Munich, Garching, Germany

## Correspondence

Sergey A. Samsonov, Faculty of Chemistry, University of Gdańsk, ul. Wita Stwosza 63, 80-308 Gdańsk, Poland.  
Email: sergey.samsonov@ug.edu.pl

## Funding information

Narodowe Centrum Nauki, Grant/Award Number: UMO-2018/30/E/ST4/00037; Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 863 – A10 – 111166240

## Abstract

Glycosaminoglycans (GAGs), long linear periodic anionic polysaccharides, are key molecules in the extracellular matrix (ECM). Therefore, deciphering their role in the biologically relevant context is important for fundamental understanding of the processes ongoing in ECM and for establishing new strategies in the regenerative medicine. Although GAGs represent a number of computational challenges, molecular docking is a powerful tool for analysis of their interactions. Despite the recent development of GAG-specific docking approaches, there is plenty of room for improvement. Here, replica exchange molecular dynamics with repulsive scaling (REMD-RS) recently proved to be a successful approach for protein–protein complexes, was applied to dock GAGs. In this method, effective pairwise radii are increased in different Hamiltonian replicas. REMD-RS is shown to be an attractive alternative to classical docking approaches for GAGs. This work contributes to setting up of GAG-specific computational protocols and provides new insights into the nature of these biological systems.

## KEYWORDS

glycosaminoglycan modeling, glycosaminoglycan–protein interactions, molecular docking, molecular dynamics, replica exchange with repulsive scaling

## 1 | INTRODUCTION

Glycosaminoglycans (GAGs) are a class of long linear periodic anionic polysaccharides made up of repetitive disaccharide units consisting of a hexosamine and an uronic acid.<sup>1</sup> Monosaccharide composition, uronic acid epimerization states, glycosidic linkage types as well as the sulfation pattern, which is often referred to as a sulfation code,<sup>2</sup> contribute to the vast chemical variety of GAGs.<sup>3</sup> In total, six classes of mammalian GAGs, heparin (HP), heparan sulfate (HS), chondroitin sulfate (CS), dermatan sulfate (DS), keratan sulfate (KS) and hyaluronic acid (HA) consisting of 202 unique disaccharide combinations including 48 HS disaccharide building blocks.<sup>4</sup> Furthermore, the organization of HS chains into domains with specific sulfation patterns as the ones

with predominance of NS- or S-domains (corresponding to the sulfation of the amino- or hydroxy-groups, respectively) increases the structural diversity of these molecules even more.<sup>5</sup> From the biological point of view, GAGs play a key role in the ongoing processes in the extracellular matrix of the cell (ECM) including cell signaling,<sup>6</sup> angiogenesis,<sup>7,8</sup> tissue regeneration,<sup>9</sup> and related pathologies such as cancer,<sup>10</sup> neurodegenerative diseases as prionic,<sup>11</sup> Alzheimer's,<sup>12</sup> and Parkinson's diseases,<sup>13</sup> and interactions between host and pathogen.<sup>14</sup> Recently, it has been shown that HP could be involved in the binding of SARS-CoV-2 Spike S1 protein in the receptor binding region.<sup>15</sup> The participation of GAGs in the mentioned processes is carried out via direct intermolecular interactions with protein targets as chemokines,<sup>16</sup> growth factors,<sup>17</sup> morphogens,<sup>18</sup> membrane receptors integrins<sup>19</sup> and lipoproteins<sup>20</sup> mostly in ECM but also in plasma membranes, exosomes and lysosomes.<sup>21</sup> All this renders natural and

Martyna Maszota-Zieleniak and Mateusz Marcisz contributed equally to this work.

artificially sulfated GAGs to be attractive targets for designing novel strategies in regenerative medicine as well as in treatment of disorders induced by the molecular mechanisms involving GAGs.<sup>22</sup> Due to their physicochemical nature, there are many challenges for both experimental and computational approaches to properly investigate molecular systems where these polysaccharides are involved.<sup>23</sup> One class of computational approaches, namely molecular docking, proved to be valuable to understand protein–GAG interactions in numerous computational and interdisciplinary studies where docking was a complementary tool to the experimental approaches.<sup>24</sup> However, despite recent successful attempts to dock GAGs, the docking performance in these systems is still limited in comparison to the docking performance on small molecule ligands.<sup>25</sup> The reason for this is high flexibility of GAGs,<sup>26</sup> the crucial impact of solvent-mediated interactions with proteins,<sup>27</sup> multipose binding due to their periodicity,<sup>28,29</sup> binding on the protein surface without well-defined binding pockets (in contrast, docking GAGs to GAG-specific enzymes is essentially an easier task)<sup>25</sup> and often unspecific/purely electrostatics-driven nature of their interactions<sup>30</sup> mediated by long and flexible positively charged residues.<sup>31</sup> Although GAG docking could be powerful when it is applied together with the restraints of experimental origin as mass spectroscopy,<sup>32</sup> nuclear magnetic resonance<sup>33</sup> or mutagenesis,<sup>34</sup> in practice docking is rather used when no information about the binding site and/or binding pose is available. Due to the decisive role of the charge–charge interactions in protein–GAG complexes, electrostatic potential calculations were shown to be effective in prediction of GAG binding regions.<sup>25</sup> In contrast, our evaluation of six docking programs on the nonredundant dataset of all protein–GAG structures available in the PDB and eight other programs on a dataset containing 28 complexes with GAG longer than dp3 (dp stands for the degree of polymerization) suggested that though the ligand binding poses (placement) could be correctly predicted by many approaches, the ranks of the docking poses (scoring) are often poorly assigned.<sup>25,35</sup> Only free accessible Autodock 3,<sup>36</sup> Dock<sup>37</sup> and commercial Glide<sup>38</sup> yielded the results of quality that meets the needs required for protein–GAG analysis in practice.<sup>35</sup> Besides the standard approaches, there were several attempts to develop GAG specific docking software as integration of HP parameters into ClusPro.<sup>39,40</sup> In 2014, we developed a dynamic molecular docking approach, where targeted molecular dynamics (MD) was used to force the initially unbound GAG ligand to slowly approach a priori known GAG binding region.<sup>41</sup> The advantage of this approach is that both a protein (receptor) and a GAG (ligand) are fully flexible, and the simulations are performed in the explicit solvent. The drawback of the method is being too expensive in terms of use of computational resources in comparison to conventional docking. In another attempt to develop a GAG-specific docking tool, Griffith et al.<sup>42</sup> created GAG-Dock, a method which allowed to successfully dock several protein–GAG systems, which however, did not represent complicated targets for other docking approaches either. Furthermore, we developed a fragment-based docking protocol for GAGs, which represented a set of scripts allowing to assemble long GAG chains from the dp3 fragments pre-docked by Autodock 3.<sup>43</sup> The aim of this method is to dock longer

GAG molecules more effectively than the classical docking approaches for which the increase of the GAG length and, consequently, the number of the degrees of freedom considered in the calculations, dramatically affects the performance. Although this method in general allows for docking GAGs of any length without increasing the need for computational resources, it is dependent on the results obtained from Autodock 3 for GAG dp3 fragments, and, therefore, is not able to extend the length of the GAG chain if some regions on the protein receptor surface are electrostatically too pronouncedly positive in comparison to their surroundings.

The actual work represents a next step to deal with the aforementioned challenges of GAG docking by applying the recently published Replica Exchange Molecular Dynamics with Repulsive Scaling (REMD-RS)<sup>44</sup> and to examine its performance for this particular molecular system type. While conventional MD approaches are supposed to be able to sample all possible conformations of the ligand on the receptor's surface, in reality such search would be too costly and, therefore, impractical because the molecular systems are often prone to be trapped into local minima in the free energy landscape.<sup>45</sup> In general, the biased potentials as implemented in Hamiltonian Replica Exchange approaches can assist to solve such a problem.<sup>46</sup> In REMD-RS, the additional potential is introduced by increasing effective pairwise van der Waals radii without affecting other types of inter- and intramolecular interactions for the systems in the implicit solvent. This method proved to be highly successful for a dataset of protein–protein complexes.<sup>44</sup> In protein–GAG complexes, electrostatic interactions are very strong, and dissociation/dramatic change of the GAG binding pose is hindered by the formation of strong charge–charge interactions in a conventional MD simulation. Therefore, this method could be particularly appropriate for these systems to avoid local minima, while the increase of the van der Waals radii in replicas would allow the ligand to sample the protein surface more extensively. Similarly, to the dynamic molecular docking<sup>41</sup> the fact that the docking by REMD-RS is performed by the application of an MD-based protocol implies that interacting molecules are flexible. GAGs are essentially smaller than proteins and, therefore, have less degrees of freedom implying that probably this method could yield promising results but also to be computationally even more effective than for protein–protein complexes. As in the fragment-based approach,<sup>43</sup> REMD-RS applicability is not limited to any length of a GAG ligand as it is the case for Autodock 3, for example. The number of allowed degrees of freedom in the ligand does not render the calculations to be significantly more expensive since the ligand is essentially smaller than the receptor, which is also simulated. In our work, we evaluate the performance of original REMD-RS method developed by Siebenmorgen et al.<sup>44</sup> for 21 protein–GAG complexes with GAGs of the length dp5–dp7 used as the fragment-based docking approach benchmark including a complex with HP dp4 containing Ca<sup>2+</sup> ions in the GAG binding site and the complex of alginate lyase enzyme AlgE4 with the alginate oligosaccharide, another linear anionic polysaccharide made up of two building blocks, mannuronic acid and guluronic acid. The reason for considering systems including Ca<sup>2+</sup> ions is that Ca<sup>2+</sup> ions and other divalent ions could be crucial for the mediation of protein–GAG interactions,<sup>47,48</sup> while AlgE4 is a highly negatively charged protein molecule binding its negatively charged substrate in a well-defined positively charged groove.<sup>49</sup> The data

obtained in this work suggest that the use of REMD-RS for protein–GAG systems where the GAG is not bound in well-defined groove/cavity, is promising for blind prediction of a GAG binding site and a direction of the GAG chain, and, therefore, can be used along with other conventional docking and specific GAG docking protocols.

## 2 | METHODS

### 2.1 | Benchmark

Fifteen X-ray experimental structures from the PDB were used as a benchmark dataset: 2AXM (FGF-1 with HP dp6, 2.2 Å; protein monomeric form was used),<sup>50</sup> 1BFC (FGF2 with HP dp6, 2.9 Å),<sup>51</sup> 1FQ9 (FGF-2/FGFR-1 with HP dp6, 3.0 Å),<sup>52</sup> 1GMN (NK1 [HGF] with HP dp5, 2.3 Å; protein dimeric form was used),<sup>53</sup> 1RID (VCP with HP dp7, 2.1 Å),<sup>54</sup> 1XMN (thrombin with HP dp6, 1.9 Å; protein monomeric form was used),<sup>55</sup> 2HYV (annexin 2A with HP dp5, 2.3 Å),<sup>56</sup> 2JCQ (CD44 with HA dp6, 1.3 Å),<sup>57</sup> 3C9E (cathepsin K with CS dp6, 1.8),<sup>58</sup> 4N8W (cathepsin K with CS dp6, 2.02),<sup>59</sup> 3INA (heparinase I with HP dp7, 1.9 Å),<sup>60</sup> 3MKP (NK1 with HP dp6, 2.8 Å),<sup>61</sup> 4AK2 (BT4661 with HP dp6, 1.4 Å),<sup>62</sup> 4C4N (hedgehog morphogen with HP dp6, 2.4 Å),<sup>63</sup> 1GN5 (annexin V with HP dp4, 1.9 Å),<sup>64</sup> 2PYH (AlgE4 with alginate dp3, 2.70 Å),<sup>49</sup> 2NWG (CXCL12 with two HP dp2, 2.07 Å; protein dimeric form was used).<sup>65</sup> In case there was a sandwich structure with the ligand being symmetrically bound between two monomeric units of the protein (2AXM, 1XMN), only a monomeric protein receptor was considered. When the ligand binding pose to a dimeric protein was not symmetric in relation to its monomeric units (1GMN, 2NWG), a dimeric receptor was considered. The criteria for the structures selection to be included into the benchmark was the length of a GAG ligand which was longer or equal to dp5. The structure of annexin V with HP dp4 (1GN5) was also considered in the calculations since it is one of two (together with 2HYV) nonredundant protein–GAG systems with Ca<sup>2+</sup> in the protein–GAG binding interface. In case of Ca<sup>2+</sup>-containing protein–GAG complexes (1GN5 and 2HYV) the ions were kept on the protein surface before REMD-RS docking procedure. The structure of CXCL12 with two HP dp2 (2NWG) was analyzed because it is the only known protein–GAG complex where there are two HP molecules simultaneously bound to high and low affinity HP binding sites.<sup>65</sup> For all these structures, the ligands present in the experimental structures were used for docking. For some of the systems, longer GAGs were also docked (in this case, the length of a ligand is additionally provided in the tables in the Results section and discussed explicitly in the text). The structure of the protein receptor from AlgE4 complex with dp3 (2PYH) was used to dock a dp15 alginate made up of β-D-mannuronic repetitive acid units.

### 2.2 | REMD-RS

The protocol used in the original work of Siebenmorgen et al.<sup>44</sup> was applied to the protein–GAG complexes almost minor modifications.

The ff14SBonlysc force field parameters for protein<sup>66</sup> and the GLY-CAM06<sup>67</sup> for GAGs were used, respectively. The protonation of the residues corresponded to pH 7: amino groups were protonated, carboxyl groups were deprotonated and histidines were protonated at ε-nitrogen (HIE residue library in AMBER). The oligosaccharide ligand was placed at the opposite side of the protein in respect to the experimentally known binding site. In case of 1FQ9, the initial location of the ligand was on the side of this symmetric tertiary complex. The structural differences between the GAG at the beginning of the REMD-RS simulation and in the experimental structure are shown in Table 1 (column 2). MD simulations were performed in implicit solvent with the model igb = 8<sup>68</sup> with an infinite cutoff for nonbonded interactions. The minimization was performed by 3000 steps of steepest descent and 3000 steps of conjugate gradient. It was followed by heating to 300 K for 10 ps with a Langevin thermostat ( $\gamma = 5 \text{ ps}^{-1}$ ). The harmonic restraints of 0.05 kcal/mol/Å<sup>2</sup> were applied on all heavy atoms of protein (and Ca<sup>2+</sup> if present) in the production run. In addition, in order to avoid ligand dissociation too far away from the receptor, the half parabolic distance restraints of 1.0 kcal/mol/Å<sup>2</sup> between the center of mass (COM) of the receptor and ligand was applied. The distance between COMs was calculated as a sum of the maximum distances between the COMs and atoms of the surface of the receptor and ligand, respectively, increased by 10 Å, to allow the free movement of the ligand on the complete surface of the protein. At the same time, the distance restraints are designed in the way that they do not allow the ligand to dissociate too far (further than a certain cut-off) from the surface of the protein and, therefore, keep the ligand so close to the protein surface that protein and ligand still “see” each other. Otherwise, when the ligand is not trying to move too far from the protein surface, its movements are free of any restraints, and the restraints do not affect the results of the REMD-RS docking. 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 receptor and ligand molecules. As in the original work<sup>44</sup> the parameter *d* adjusting the effective van der Waals radius (in Å) was set to 0.00, 0.01, 0.02, 0.04, 0.08, 0.12, 0.16, 0.20, 0.24, 0.28, 0.32, 0.38, 0.44, 0.50, 0.58, 0.68, and a factor *e* changing the LJ potential well depths (unitless) was assigned to and 1.00, 0.99, 0.98, 0.97, 0.96, 0.94, 0.92, 0.90, 0.88, 0.86, 0.84, 0.82, 0.80, 0.78, 0.76, 0.74, respectively. In the production run, 50,000 MD exchange steps between neighboring replicas were attempted yielding 500 ns per replica in total. For each particular system, the total time of the structural convergence was different (as it is discussed further in the Results and Discussion section). Trajectory frames were saved each 200 ps.

### 2.3 | Free energy estimation

For the calculations of binding free energy we used the total energy values from the AMBER output directly and molecular mechanics-generalized born surface area (MM-BSA) model igb = 2<sup>69</sup> as well as its components for the first replica (the one with unmodified LJ

**TABLE 1** Evaluation of REMD-RS methodology performance on a protein–GAG dataset before and after the refinement for the best 1, 10, and 100 binding poses

Complex, PDB ID	RMSatd, Å <sup>a</sup> Start	RMSatd1, Å <sup>b</sup> Before	RMSatd1, Å <sup>c</sup> After	RMSatd10, Å <sup>b</sup> Before	RMSatd10, Å <sup>c</sup> After	RMSatd100, Å <sup>b</sup> Before	RMSatd100, Å <sup>c</sup> After	$d_{\text{conv}}$ , ns
2AXM	45.1	5.4	2.7	5.8 ± 0.3	5.4 ± 1.3	8.0 ± 0.5	4.9 ± 1.3	40
1BFC	42.3	6.5	4.8	7.3 ± 0.5	6.7 ± 5.3	9.7 ± 1.4	8.7 ± 6.1	10
1FQ9 dp6	45.5	13.6	14.7	18.2 ± 2.0	14.4 ± 1.8	23.1 ± 2.2	17.8 ± 2.1	1
1FQ9 dp8	67.6	16.6	9.4	17.7 ± 0.6	10.0 ± 1.2	20.6 ± 1.4	11.6 ± 1.7	2
1FQ9 dp14	49.5	12.1	5.2	13.0 ± 0.6	5.4 ± 1.1	14.7 ± 0.8	5.6 ± 0.9	6
1GMN	58.8	4.9	4.6	6.2 ± 1.1	4.9 ± 0.9	12.1 ± 3.4	8.1 ± 2.9	20
1G5N	49.1	12.5	13.1	15.4 ± 1.8	13.3 ± 5.5	24.9 ± 4.5	22.5 ± 6.9	26
1RID	66.1	15.3	14.5	17.3 ± 1.4	18.6 ± 3.8	52.2 ± 26.4	52.0 ± 26.7	7
1XMN	57.7	11.6	6.5	14.6 ± 1.5	17.1 ± 9.0	34.3 ± 12.6	27.3 ± 12.5	28
2HYV	57.7	4.5	5.7	4.8 ± 0.2	6.3 ± 1.3	6.2 ± 0.7	7.1 ± 1.3	10
2JCQ	27.8	2.0	3.2	2.4 ± 0.2	3.3 ± 1.0	5.3 ± 1.9	5.9 ± 2.5	220
3C9E	50.1	6.9	6.5	7.2 ± 0.1	6.8 ± 0.4	13.3 ± 5.2	12.7 ± 5.5	8
3C9E dp12	48.5	6.8	3.5	8.6 ± 0.9	6.6 ± 2.9	14.9 ± 3.3	13.1 ± 4.2	4
4N8W	54.3	3.3	4.7	4.5 ± 0.6	6.0 ± 1.4	7.2 ± 1.4	7.4 ± 1.8	8
4N8W dp12	41.3	3.4	4.7	4.1 ± 0.4	4.8 ± 1.0	6.3 ± 1.2	6.5 ± 1.7	4
3INA	37.8	21.6	19.8	23.8 ± 1.0	19.6 ± 2.4	29.2 ± 2.8	24.9 ± 4.4	24
3MKP dp6	31.6	5.8	4.8	6.5 ± 0.5	6.0 ± 1.0	9.3 ± 1.5	7.4 ± 2.0	1
3MKP dp12	31.8	5.7	4.1	7.4 ± 0.7	5.0 ± 1.4	9.7 ± 1.2	6.2 ± 1.8	10
4AK2	47.5	11.6	8.7	18.8 ± 4.6	18.2 ± 5.0	44.0 ± 11.0	42.9 ± 11.8	No found
4CNC	47.7	8.5	12.1	13.0 ± 1.7	12.4 ± 2.9	18.6 ± 3.1	17.7 ± 4.3	60

<sup>a</sup>RMSatd at the starting frame of the REMD-RS simulation with the reference to the X-ray structure.

<sup>b</sup>RMSatd at the last frame of the REMD-RS simulation before the refinement with the reference to the X-ray structure.

<sup>c</sup>RMSatd after the refinement with the reference to the X-ray structure.

<sup>d</sup> $t_{\text{conv}}$ : time of the simulation convergence; In bold are the values where the refinement improved the results in comparison to the not refined docking solutions.

parameters). In addition, the total binding energies from the AMBER output were analyzed.

## 2.4 | Parameters used for docking performance evaluation

As a metric for structural similarity, we used RMSatd (root mean squared atom type deviation) which is defined as the root-mean-square of pairwise atomic distances while pairing up the spatially closest atoms of the same type. This metric is supposed to be physically more appropriate for GAG-containing systems than classical RMSD per atomic identifier due to the periodic nature of GAGs.<sup>41</sup> This distance metric is sounder for GAG analysis than a classical RMSD while it accounts for the periodicity of functional groups in GAGs by considering two GAGs shifted by  $n$  periodic units as structurally similar, whereas classical RMSD (root mean squared deviation with identical atom-based matching) ignores this GAG property. RMSatd can be used for comparison of the structures with a different number of atoms (e.g., GAGs of different lengths), and it finds the best match accounting for the GAG periodicity. In contrast, RMSD could be used

for such a comparison only if a particular part of a longer GAG is pre-defined before the comparison. Nevertheless, in case two ligand structures of the same length are compared, and there is not shift along the GAG chain, the metrics yield almost indistinguishable results. The only difference originates from the fact that, for example, for  $-\text{SO}_3-$  group, the closest oxygen atoms between each other in two structures will be considered for calculations when RMSatd is used, while the a priori numbered oxygen atoms will be considered in the calculations of RMSD. In the second case, any rotation of the  $-\text{SO}_3-$  group during a docking or MD run, in which one of the physically indistinguishable but fixed by its number oxygen takes place of another one, leads to the increase of RMSD because the atoms are numbered in that way but not in RMSatd, which is physically more correct. For the calculation of RMSatd, all atoms in one structure are compared with all in a reference structure but the pairing happens for the ones with the lowest distance between each other if they are of the same type. Taking into account the linearity of a GAG, such a metric allows for a physically relevant comparison between two structures. At the same time, RMSatd is even more physically correct for describing the interactions with proteins: i.e., when a sulfate group occupies the same position as a different sulfate group in a reference

structure, the interaction established with the protein is equivalent, which is reflected by the RMSatd but not by the corresponding RMSD value. Several RMSatd-based descriptors of the MD simulation were used in our study to evaluate the REMD-RS docking prediction power: time of RMSatd convergence; RMSatd of the best poses (the lowest RMSatd with the reference to the corresponding experimental structure); RMSatd of the top poses (the RMSatd with the reference to the corresponding experimental structure for the structures with the best score/free energy value); the relationship between RMSatd and scores/free energy values; RMSatd values of best and top poses after the refinement. We also calculated classical ligand RMSD.

## 2.5 | Refinement

Refinement of the structures obtained in the REMD-RS step was performed for 1, 10, and 100 structures for each complex from the first replica corresponding to the first 1, 10, and 100 best (the closest structures to the experimental reference in terms of RMSatd) and 1, 10, and 100 top (the lowest total binding energies from the AMBER output) docking solutions. In this procedure, 5 ns of unrestrained MD simulation in explicit solvent was performed. The obtained docking solutions were solvated in a truncated octahedron TIP3P periodic box of 8 Å water layer from the solute to the box's border. Na<sup>+</sup> or Cl<sup>-</sup> counterions were used. Energy minimization was carried out in two steps: 500 cycles of steepest descent and 1000 of conjugate gradient with 100 kcal/mol/Å<sup>2</sup> harmonic force restraints, continued with 3000 steepest descent cycles and 3000 conjugate gradient cycles without restraints. The system was heated up from 0 to 300 K for 10 ps with harmonic force restraints of 100 kcal/mol/Å<sup>2</sup>. Then, the system was equilibrated at 300 K and 1 atm in isothermal isobaric ensemble for 5 ns.

## 2.6 | Data analysis

Statistical analysis of the data and plot were prepared in R.<sup>70</sup>

## 2.7 | Visualization

Structures and trajectories were analyzed in Pymol<sup>71</sup> and VMD.<sup>72</sup>

# 3 | RESULTS AND DISCUSSION

## 3.1 | The REMD-RS docking workflow for protein-GAG complexes

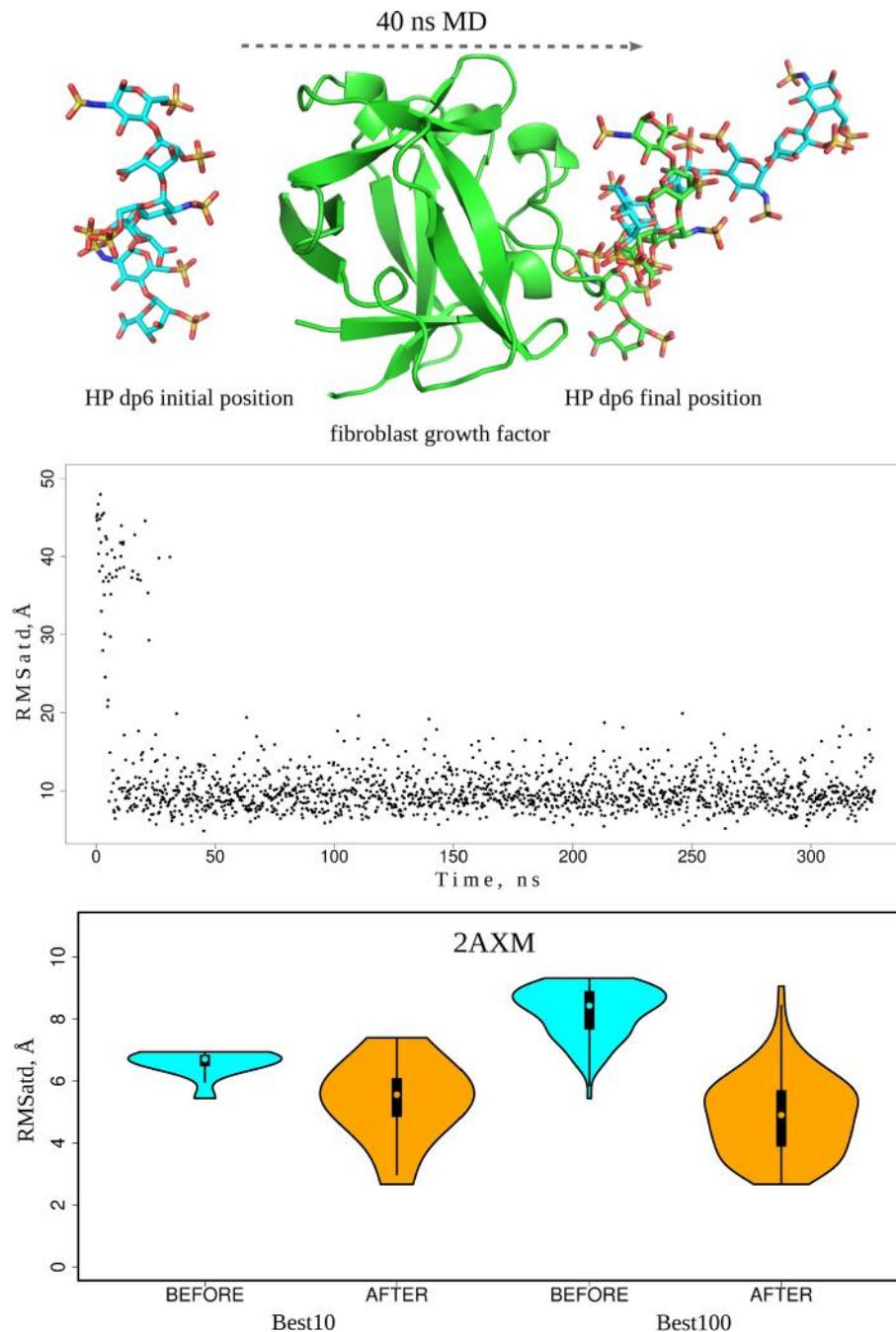
In this subsection we describe an example of REMD-RS docking results for the complex formed between fibroblast growth factor 1 (FGF-1) and HP dp6, which is one of the most studied protein-GAG complexes both experimentally and computationally.<sup>73,74</sup> Previously,

several docking programs have succeeded in predicting a proper binding pose in this complex.<sup>25,35</sup> At the beginning of the simulation, HP was placed as far as possible from the experimental ligand position on the protein surface (Figure 1). During the initial phase of the REMD-RS MD simulation the ligand searched the entire surface of the protein, but did not dock at any site for more than 1–2 ns. After 40 ns of the simulation HP docked on the protein surface in the binding site corresponding to the X-ray structure and remained in this position for the rest of simulation time (total MD time was 326 ns). Forty nanoseconds is, therefore, referred as a time of ligand convergence and is similarly evaluated for other simulated complexes (Table 1).

After the REMD-RS MD step the FGF-1/HP dp6 structures generated from the first replica (corresponding to the unmodified vdW values) were compared to the X-ray structure using RMSatd metric. Based on this, for the 100 complex structures with the best (lowest) RMSatd values refinement procedure was applied (Figure 2). As it is seen in the Table 1 and Figure 1, in each case (best 1, 10, and 100 solutions), the RMSatd values decreased which suggests that the refinement improved the results for the best-placed docking solutions. The best poses characterize the placement performance of the docking approach. Another important feature of the docking performance is scoring. We compared several scoring schemes to estimate the correlation between the RMSatd of the docking solutions and free energy component values obtained from MM-GBSA (total, electrostatic in vacuo and with the consideration of GB reaction field and van der Waals) as well as the total energy from the AMBER output, which corresponds to the full potential and kinetic energy of the system in the implicit solvent. Pearson correlation describes the ability of the scoring scheme to quantitatively differentiate the solutions with more favorable scores based on their structural differences, while Spearman correlation shows how a scoring scheme can properly rank (qualitatively differentiate) the solutions corresponding to certain structural differences. The correlations were calculated for the frames prior to the pose convergence considered together with the same number of frames after the convergence in order to have the consistency and comparability of these correlations among different complexes (Table 2). In particular, for this complex, 80 ns (400 frames) were used for the correlation analysis. For FGF-1/HP dp6 the highest Pearson correlation (0.93) was obtained for the electrostatic component of MM-GBSA in vacuo, followed by full electrostatics (0.73) and total MM-GBSA energy (0.70). The lowest correlation for an MM-GBSA component (0.25) was observed for the van der Waals one, while no correlation was obtained for the total energy from the AMBER output (Table S1). Such results are in agreement with the established view on the crucial role of electrostatics in protein-GAG interactions in general<sup>75</sup> and particularly for this complex.<sup>25</sup> Spearman correlations are significantly lower than the Pearson ones for all the components except for the van der Waals component. In comparison to other 14 tested docking programs, where the highest observed correlations were obtained for Autodock 3 (0.61 and 0.62 for Pearson and Spearman correlations, respectively)<sup>25</sup> and Dock (0.76 and 0.84 for Pearson and Spearman correlations, respectively),<sup>35</sup> REMD-RS in a combination with MM-GBSA clearly overperforms them. Furthermore, we analyzed whether the top solutions (based on the total energy from the



**FIGURE 1** Top panel: The initial and final structure of the FGF-1/HP dp6 complex in REMD-RS simulation (docking run). Green color—X-ray structure of the complex (PDB ID: 2AXM); blue color—Position of the HP dp6 at the beginning and at the end of the REMD-RS simulation. Middle panel: RMSatd of the HP in the docking run with the reference to the X-ray structure. Bottom panel: RMSatd for best 10 and 100 docking solutions before and after the refinement procedure presented as violin plots. The probability density of the RMSatd is shown smoothed by a kernel density estimator, while median, minimum, maximum, the first and the third quartile are depicted as an open circle, lower, upper point of the vertical line, lower and the upper borders of the black box, respectively



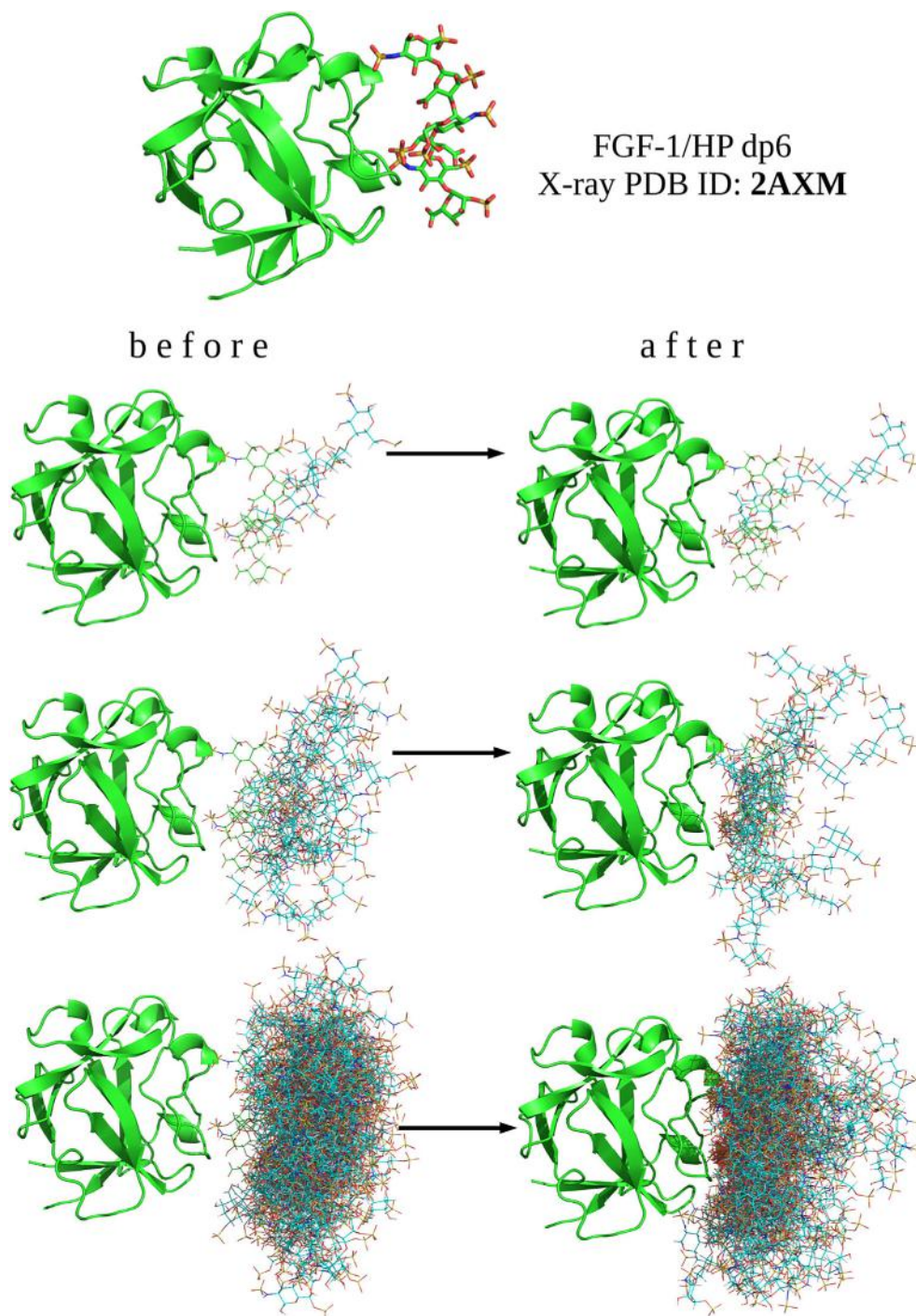
AMBER output) could be improved by the refinement. Interestingly, despite the weak scoring performance of this scoring scheme, the first top solution improved its RMSatd from 7.7 to 2.7 Å for 2AMX, while only the slight improvement for top 100 solutions was observed.

### 3.2 | General performance of the REMD-RS docking approach for protein–GAG complexes

The aforementioned procedure was applied to 21 protein–GAG complexes and one protein–alginate complex. For 19 of 21 systems, experimentally observed binding sites were correctly predicted by the

REMD-RS approach (Table 1), and the experimental GAG orientation in the binding site was within the highly scored solutions. However, many solutions were shifted along the reference GAG chain or an antiparallel binding pose was predicted.

Among protein–GAG complexes, only for the complex of SusE-like surface polysaccharide binding protein with HP dp6 (PDB ID: 4AK2) the binding site was identified incorrectly (the lowest RMSatd was 11.6 and 8.7 Å before and after the refinement, respectively). In most of the cases the properly bound binding sites corresponded to the protein surface patches with the most positive electrostatic potential (Figure S1) as it was previously described.<sup>25</sup> For the AlgE4–alginate complex, the ligand dissociated immediately from the protein



**FIGURE 2** The best 1, 10, and 100 docked structures (from top to bottom) of the FGF-1/HP dp6 complex before and after the 5 ns of the MD refinement procedure. Green color—X-ray structure of the complex (PDB ID: 2AXM); blue color—The best HP dp6 docking positions

and did not bind to it at any protein surface region. This is related to the fact that AlgE4 protein has a negative net charge of  $-34$  with a well-defined positively charged groove in the enzymatic binding site (Figure S2). In the course of the REMD-RS simulation, the negatively charged alginate is repelled from the protein (also when its starting conformation is in the close proximity to the binding groove), while due to the narrowness of the binding groove, it cannot associate with the protein. This happens because its volume is just too big in the higher replicas to come close enough to the binding groove and to fit there. For another complex of the enzymatic protein in the dataset,

heparinase with HP dp7 (PDB ID: 3INA), the method finds the binding site properly after 24 ns of the simulation but similarly to the scenario with AlgE4. However, due to the nature of the method when  $d > 0$  and  $\epsilon < 1.0$  are used, HP ligand does not succeed to enter the binding groove. These results suggest that REMD-RS docking methodology with these particular parameters can be successfully applied to the nonenzymatic complexes of GAGs with proteins, where the binding is not maintained in a groove/pocket. Fortunately, neither of nonenzymatic protein-GAG complexes represent a pocket-type of binding.<sup>25</sup> Instead, GAGs tend to bind on the positively charged and

**TABLE 2** Pearson and Spearman correlation between RMSatd and MM-GBSA binding free energy components

Complex	$\Delta G_{\text{total}}$		$\Delta G_{\text{Ele}}$		$\Delta G_{\text{Ele}} + \Delta G_{\text{GB}}$		$\Delta G_{\text{vdw}}$	
	Pearson	Spearman	Pearson	Spearman	Pearson	Spearman	Pearson	Spearman
2AXM	0.70	0.44	0.93	0.65	0.72	0.38	0.25	0.39
1BFC	0.69	0.80	0.81	0.82	0.71	0.77	0.36	0.48
1FQ9 dp6	0.80	0.45	0.88	0.62	0.76	0.41	0.34	0.64
1FQ9 dp8	0.91	0.87	0.96	0.88	0.96	0.86	0.61	0.86
1FQ9 dp14	0.92	0.91	0.96	0.96	0.93	0.91	0.62	0.76
1GMN	0.68	0.72	0.80	0.78	0.70	0.69	0.36	0.49
1G5N	0.52	0.40	0.50	0.41	0.52	0.39	0.19	0.26
1RID	-0.33	-0.44	-0.33	-0.44	-0.32	-0.44	-0.21	-0.19
1XMN	0.35	0.38	0.41	0.50	0.33	0.35	0.29	0.56
2HYV	-0.24	-0.36	0.97	0.95	-0.43	-0.39	0.58	0.77
2JCQ	0.65	0.67	0.61	0.62	-0.59	-0.55	0.71	0.71
3C9E	0.39	0.13	0.64	0.25	-0.22	-0.23	0.50	0.31
3C9E dp12	0.61	0.59	0.84	0.80	0.37	0.35	0.35	0.33
4N8W	0.50	0.37	0.85	0.59	0.23	0.29	0.29	0.26
4N8W dp12	0.68	0.66	0.90	0.84	0.53	0.57	0.27	0.35
3INA	0.81	0.81	0.85	0.84	0.79	0.79	0.38	0.74
3MKP dp6	0.62	0.61	0.85	0.74	0.54	0.47	0.48	0.52
3MKP dp12	0.63	0.37	0.76	0.42	0.78	0.44	0.01	0.33
4AK2	0.63	0.64	-0.36	-0.33	-0.35	-0.36	0.00	-0.06
4CNC	-0.04	0.01	0.00	0.06	-0.06	-0.01	0.16	0.27

solvent exposed protein surface patches. For enzyme-GAG systems, where a GAG binds tightly packed into the shape-complementary binding pose to the groove/pocket,  $d < 0$  and  $e > 1.0$  should be rather used instead in REMD-RS approach to be potentially able to yield good quality results. At the same time, a number of conventional and computationally less demanding docking approaches are quite successful in docking GAGs to enzymes.<sup>25</sup>

The quality of the placement of the approach is reflected by RMSatd values of the best poses (Table 1) suggesting that for more than half of the complexes poses with RMSatd values lower than 10 Å were abundant. Since before the refinement higher replicas with the increased van der Waals radii do not allow for close stabilized contacts between the receptor and the ligand, these values are expected to be improved by the refinement (see the next subsection). It should be noted that in this approach completely blind/global docking is carried out, while most of the previously applied docking programs performed local docking in predefined binding sites. Other docking programs typically yield at best 3–4 Å difference for predicted protein-GAG complexes in terms of RMSatd to the experimental structure.<sup>25,35</sup> However, in case of the referred work, local docking approaches to rigid receptors were evaluated. In case the flexibility of the receptor was allowed, leading to high fluctuations of the long positively charged GAG binding residues and so increasing the RMSatd for the ligand, the best RMSatd values were about 5–7 Å, which is similar to the described here RMSD-RS performance.<sup>41</sup> Moreover, a

10 μs MD study of FGF1-HP dp6 X-ray structure, which is one of the easiest for all docking approaches,<sup>25,35</sup> showed very high (about 4 Å of RMSD) fluctuations of the GAG ligand in the course of the simulation<sup>76</sup> rendering it a priori challenging to dock a GAG with much higher accuracy than these fluctuation values.

The amount of time needed for the equilibration of different systems are in a range from 1 ns (for HP dp5 in 1MKP complex and for HP dp6 in 1FQ9 complex, e.g.), to 220 ns (for HA dp6 in 2JCQ complex). However, for half of the systems the time of convergence was shorter than 10 ns, while for 17 systems, the convergence time was less than 30 ns. Interestingly, CD44/HA dp6 complex was previously computationally analyzed rigorously by 50 μs MD simulation demonstrating how challenging this particular system could be for in silico analysis.<sup>77</sup> Our method allowed to find the best 1, 10, and 100 docking poses with RMSatd of 2.0 Å, 2.4 ± 0.2 Å and 5.3 ± 1.9 Å, respectively. In addition, while CD44/HA dp6 (PDB ID: 2JCQ) represents a complex where electrostatics play a less important role<sup>27</sup> than for most of protein-GAG complexes, and where the GAG (hyaluronic acid) is the least charged GAG, more charged GAGs do not move significantly on the protein surface in nonbiased MD simulations of a feasible length. Even when a starting conformation of a GAG is far from the one corresponding to the global free energy minimum, the presence of several positively charged residues in its neighborhood does not allow the GAG to dissociate or to move away from these positively charged “anchors”.

Due to the high importance of the electrostatics in protein–GAG complexes, pH corresponding histidine residues protonation state could potentially affect the results of the MD simulation as well as the obtained MM-GBSA free energy values obtained from it.<sup>25</sup> In the analyzed benchmark dataset, there were three complexes, which were obtained under the experimental conditions with acidic pH at which histidine residues are supposed to be doubly protonated: 4N8W (pH 4.5), 3C9E (pH 4.5), 2HYV (pH 5). Whereas for 4N8W the histidine residue is far from the GAG binding site, in 3C9E and 2HYV it is approximately in the middle of the binding site, representing one of the anchors for binding the middle of a GAG chain when protonated, which is reflected in significantly more favorable binding energies for these complexes at acidic pH.<sup>25</sup> The binding pose, however, did not alter when different protonation states of these histidine residues were considered.<sup>25</sup> Since both histidine residues in the binding sites of 3C9E and 2HYV have this very central location concerning the bound GAG chain, they would contribute to the higher positive electrostatic potential in the region and could only assist REMD-RS to find the proper binding site and binding pose, which in this study was already well predicted for these complexes at neutral pH. This means that the protonation state of the histidines should be taken with care when REMD-RS or any other docking approaches are applied for protein–GAG complexes. The fact that histidine protonation could be affected by the protein local environment apart from the pH contributes to the general complexity of appropriate pH consideration in such simulations.

### 3.3 | Refinement

The refinement of the obtained docking poses allowed the relaxation of the system that is inaccessible in the REMD-RS simulation due to the bigger van der Waals radii in the simulated replicas. The closer and stronger contacts between the protein and the ligand were expected to be maintained upon the refinement procedure. Indeed, the refinement improved the results for 13, 14, and 16 systems out of 20 for 1, 10, and 100 best poses, respectively (Table 1). This clearly suggests that the refinement procedure could be beneficial in the docking pipeline. We also applied the same refinement procedure for the top 1, 10, and 100 solutions (based on the AMBER total energy). Despite the RMSatd values were significantly higher for these docking solutions in comparison to the best docking solutions, the refinement improved the results for 19, 15, and 18 systems out of 20 for the top 1, 10, and 100 solutions, respectively (Table S1). However, these results for the top poses should be interpreted carefully since in this case most of the docking poses were far from being close to the X-ray structures. There are several reasons why the refinement was performed in the explicit solvent, whereas the REMD-RS step was done in the implicit solvent: 1. the simulation in the implicit solvent is computationally much less expensive, which suggests more effective sampling on the full surface of the protein; 2. the electrostatic interaction that is the driving force in protein–GAG complexes is not screened too dramatically as it would be the case for explicit solvent, especially for the replicas with the biggest van der Waals radii, which also leads

to a more efficient “electrostatic complementarity” of the ligand to the proper binding site; 3. since the water-mediated interactions are very important in protein–GAG systems,<sup>27</sup> the use of the explicit solvent in the refinement procedure results in obtaining the structures where the contacts between protein and GAG residues are partially water-mediated which is more appropriate for these systems.

For the starting points as well as for the best 1 and 10 binding poses before and after the refinement we compared the values of RMSatd and classical RMSD (Table S2). RMSD values are essentially higher than corresponding RMSatd values, although they are well correlated: 0.88, 0.83, 0.73, 0.92, and 0.91 Pearson correlation coefficients were obtained for the starting points, the best 1 pose before and after the refinements and the best 10 poses before and after the refinements, respectively. The sufficient visual overlap between the obtained docking solutions and the X-ray structure and these high RMSD values underline that the RMSatd metric is more useful metric compared to RMSD as explained above. This statement can be justified by considering two principally important properties of GAG binding not taken into account when RMSD is applied. The first one is the GAG periodicity: if a docked structure is shifted by a dimeric unit in comparison to the X-ray reference structure, an overlapping part of a GAG could be still docked correctly (which would be reflected in a low RMSatd) but its RMSD to the X-ray structure is already higher than 8–10 Å which are typical distances between the atoms in repetitive equivalent units (e.g., based on the NMR structure of the unbound heparin, PDB ID: 1HPN). The second reason for such high RMSD values is that it is very challenging to distinguish two different orientations (or polarity) of GAG binding in the same binding sites by the computational methods since the binding strength of the antiparallel orientations is very similar.<sup>78</sup> Whereas two well overlapping antiparallel orientations are very similar in terms of RMSatd, they will have a very high RMSD (Figure S3). In terms of the physical properties of the system, the disposition of the functional groups, in particular sulfate and carboxyl groups that are keys for the interactions with proteins<sup>25</sup> are very similar in the antiparallel orientations. Therefore, the results of this work and the studies where RMSatd is used should be interpreted with care: low RMSatd values do not strictly mean high similarity in terms of the distance between the identical atoms in the docked and the reference structure but suggests that a shift by periodic unit or antiparallel orientation of the docking solution are considered in this description of the docking results, which is apparently not the case in for classical RMSD.

### 3.4 | Scoring

Furthermore, REMD-RS scoring schemes were evaluated in terms of RMSatd Pearson and Spearman correlations with the score values (Table 2 and Table S1). The highest correlations were obtained for  $\Delta G_{Eie}$ : for most of the complexes, the corresponding correlation coefficients were above 0.8 for Pearson and slightly lower for Spearman correlation. The outliers were: 1G5N, 3C9E (see the next subsection for details) and 1RID, 4AK2, 4CNC, 1XMN, 2JCQ. For 1RID, 4AK2, 4CNC, the reason for this was that the predicted binding pose was essentially

different from the experimental one. For 1XMN, two distinct binding poses (one of them was similar to the experimental one) were predicted to be similarly favorable leading to the lower correlation when taking into account RMSatd for both poses coexisting in dynamic equilibrium. For 2JCQ complex, electrostatic-driven interactions were previously shown to be less important than van der Waals ones.<sup>27</sup> In general, the obtained correlations with  $\Delta G_{\text{total}}$  and  $\Delta G_{\text{Ele}} + \Delta G_{\text{GB}}$  are lower than for  $\Delta G_{\text{Ele}}$ . However, for most of the complexes they are still meaningful (higher than 0.6). The decrease of the correlation coefficients could be related to the implicit solvent model imperfections since the proper description of solvent is especially important for protein-GAG interaction analysis.<sup>27</sup> Even lower correlation obtained for  $\Delta G_{\text{vdW}}$  supports the dominant role of electrostatics in most of the analyzed systems. The only clear exception from this rule is 2JCQ (both Pearson and Spearman correlation are 0.71), where the van der Waals contribution is more decisive than the electrostatic one<sup>27</sup> and where the GAG (hyaluronic acid) is the least charged in the whole dataset. At the same time, all MM-GBSA scoring schemes employed with REMD-RS approach including  $\Delta G_{\text{vdW}}$  revealed clear advantage over the ones implemented in 14 previously analyzed docking programs when compared in terms of the correlations between the similarity to the experimental structure and scores.<sup>25,35</sup> This suggests that considering the descent placement performance together with the use of  $\Delta G_{\text{Ele}}$ , in particular, could lead to reliable docking results. In contrast, when using AMBER total energy from the output (Table S1), there was only one complex (2HYV), for which meaningful correlations were obtained. These results are especially interesting since the key differences between these two scoring schemes are (in the order of decreasing significance): (1) the consideration of the kinetic energy component by AMBER total energy; (2) employment of the different implicit solvent models; (3) minimization of the structure for each frame in the MM-GBSA approach.

In a recent study of Siebenmorgen and Zacharias conducted in parallel to our work, MBAR free binding energy-based scoring Scheme<sup>79</sup> was applied to a dataset of 36 protein-protein complexes docked by REMD-RS method.<sup>80</sup> In this work, the Spearman correlation with the experimental values were 0.77 and 0.55 for the simulations in explicit and implicit solvent models, respectively. These results suggested that especially in case of the charge-driven interactions, the use of explicit solvent model can be advantageous. In the future steps of REMD-RS tests for protein-GAG systems where experimental binding data are accessible (which are very few for the available experimental structures<sup>25</sup>), it is worth to compare different scoring schemes and solvent models for different classes of ligands including proteins, peptides, nucleic acids and polysaccharides.

### 3.5 | Particular cases of protein-GAG complexes

#### 3.5.1 | Protein-GAG systems with Ca<sup>2+</sup> ion in the complex interface (PDB ID: 1G5N, 2HYV)

For both systems containing annexins II and V (2HYV and 1G5N, respectively) with HP, Ca<sup>2+</sup> ions in the complex interface directly

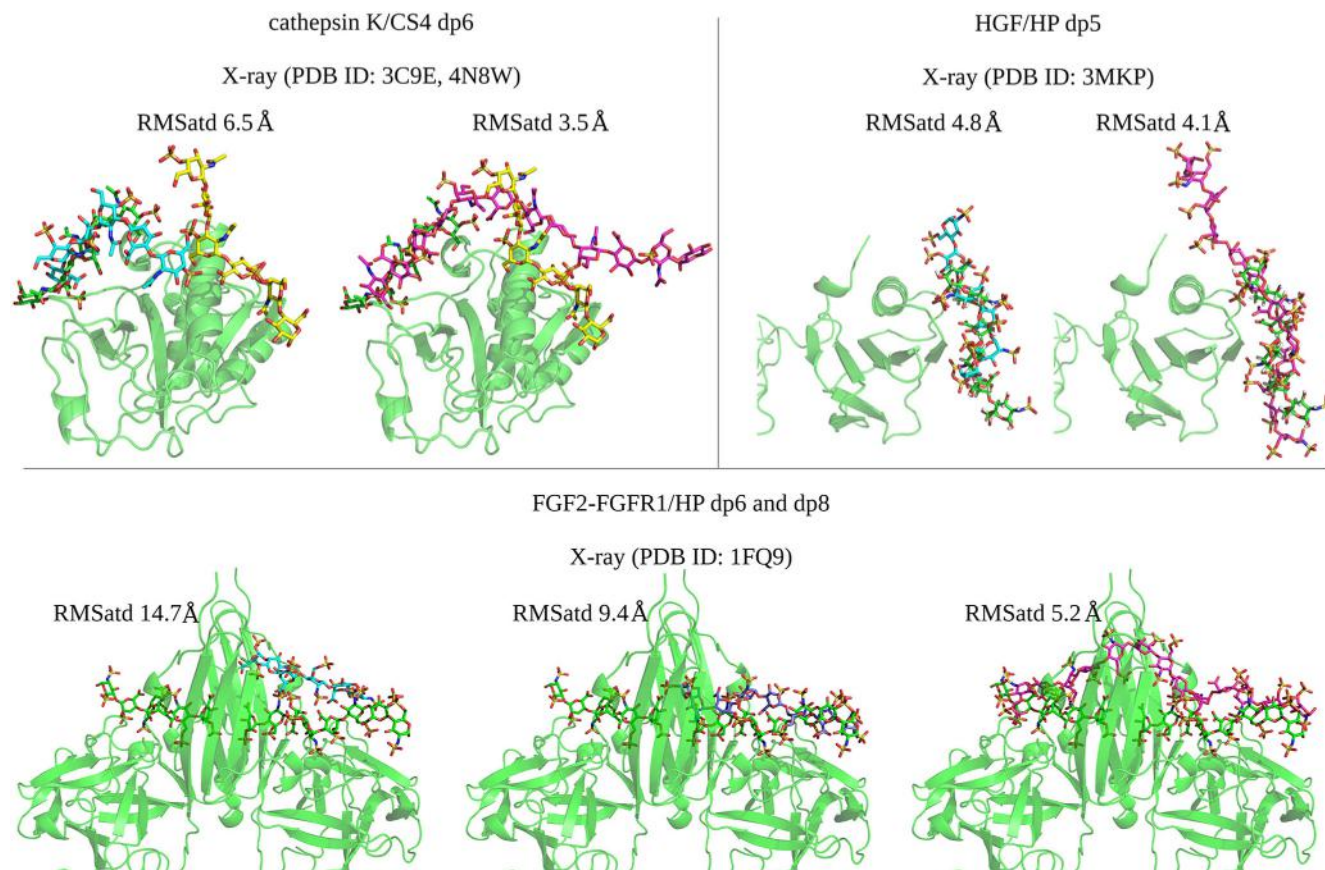
participate in establishing protein-GAG interactions. For both complexes, REMD-RS succeeded in predicting binding sites correctly and, in general, the predictive ability of this approach for these complexes did not stand out from the rest of the complexes (without ions participation) in the dataset. The experimental binding poses were accurately reproduced for 2HYV (Table 1) but not for 1G5N, where the most favorable binding pose was predicted to be only in a partial overlap with the experimental one. As a consequence, all the correlations between scores and RMSatd for 1G5N are lower than for the rest of the studied complexes on average (Table 2). For 2HYV, only correlations for  $\Delta G_{\text{Ele}}$  and  $\Delta G_{\text{vdW}}$  are meaningful, which agrees with our previous data regarding inability of MM-GBSA to account for ions in the  $\Delta G_{\text{GB}}$  term and, therefore, in  $\Delta G_{\text{total}}$ .<sup>81</sup> Instead, AMBER total energy correlated well with the structural difference between the docked structures for 2HYV and its experimental reference structure. It is worth mentioning that despite the applied restraints Ca<sup>2+</sup> ions moved significantly during the REMD-RS simulation. The scarcity of the subset of protein-Ca<sup>2+</sup>-GAG structures does not allow to draw any conclusion whether such complexes are more challenging for the approach than other complexes. As we showed recently, for these two Ca<sup>2+</sup>-containing protein-GAG complexes, the presence of Ca<sup>2+</sup> ions on the protein surface essentially affects the performance of computational analysis of these systems, including conventional and REMD-RS docking performance.<sup>82</sup>

#### 3.5.2 | Cathepsin K/CS4 dp6 (PDB ID: 3C9E, 4N8W)

Cathepsin K/CS4 is a unique system in our dataset. It is the only known protein-GAG complex where a GAG can bind to two different sites and its preferences are dictated by the experimental conditions.<sup>58,59</sup> REMD-RS found both binding sites but with a clear tendency of the 4N8W in terms of both accuracy of the best docking poses and correlations between RMSatd with free energy components (Tables 1 and 2, Figure 3). This is in agreement with the study of Lecaille et al. where the binding pose from 4N8W is a preferred one not only in the experimental study but also in the MD-based analysis.<sup>83</sup> When the docked GAG is elongated to dp12, it connects both binding poses from the X-ray structures improving the correlation coefficients for each of them. For this system, correlations with  $\Delta G_{\text{Ele}}$  and  $\Delta G_{\text{total}}$  are both significant. Moreover, this system revealed serious challenges for docking programs tested previously with regard to placement algorithms as well as for scoring performance.<sup>25,35</sup>

#### 3.5.3 | HGF/HP dp6 (PDB ID: 3MKP)

In the complex of HGF/HP dp6, a great performance of REMD-RS was observed (Tables 1 and 2, Figure 3). In this case, correlations of all MM-GBSA components were high. The results did not change significantly upon the elongation of HP dp6 to dp12. The longer GAG was docked in the way that one terminal part of the



**FIGURE 3** REMD-RS docking results (best pose) for cathepsin K/CS4 dp6 (top, left: CS4 from 3C9E is in green, CS4 from 4N8W is in yellow; best pose for dp6 is shown in blue and for dp12 in magenta); HGF/HP dp5 (top, right; the best pose for dp6 is in blue, the best pose for dp12 is in magenta, the X-ray structure is in green); FGF2-FGFR1-HP/dp6 and dp8 systems (bottom from left to right docking results for HP dp6, dp8 and dp14; X-ray structures are in green, the docked structures are in: Blue—dp6, dark blue—dp8, magenta—dp14). RMSatd values with the reference to the experimental structures are provided

docked ligand overlapped with the X-ray ligand structure for dp6. Therefore, REMD-RS could be also used to rank the GAG subsites regarding their importance/energy contribution for complex formation, while docking GAGs of different lengths could be useful for defining the minimal binding units in protein–GAG systems. The independence between performance and GAG length in REMD-RS was found to be a major advantage compared to conventional docking methods.

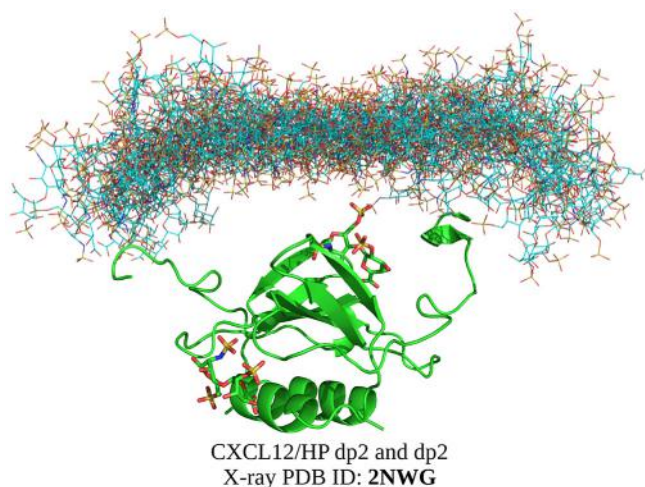
### 3.5.4 | FGF-2/FGFR1/HP dp6 and dp8 (PDB ID: 1FQ9)

FGF-2/FGFR1 with HP is another example of how GAG of different lengths could be used for the characterization and comparison of GAG binding subsites. In the crystal structure, HP dp6 and HP dp8 are resolved.<sup>52</sup> Therefore, we docked HP dp6, HP dp8, and HP dp14 separately. Our data show (Tables 1 and 2, Figure 3) that both HP dp6 and HP dp8 are preferably docked to the subsite where HP dp8 in the X-ray structure is resolved. Docking of HP dp14 yields the structure that overlaps with both experimentally resolved ligands. Interestingly,

the docking approach performs even better for HP dp14 than for shorter GAGs which is the opposite to the results obtained by conventional docking, where the performance significantly drop upon the GAG elongation.<sup>35</sup>

### 3.5.5 | CXCL-12/HP (PDB ID: 2NWG)

Finally, we docked a HP dp14 to a CXCL12 which binds two HP dp2 in the X-ray structure 2NWG in two different binding sites, characterized by high affinity and low affinity to HP.<sup>65</sup> Both NMR and molecular modeling performed with conventional docking proposed that the high affinity binding site is the first anchoring point for a GAG chain, which could be then elongated into the lower affinity binding site.<sup>64</sup> REMD-RS produced an ensemble of docked structures that indeed covered the high affinity binding site (Figure 4). The direction of GAGs in the ensemble allows its involvement in the binding to low affinity binding site after the refinement. Therefore, the new docking approach was able to properly distinguish several binding sites of different affinities on the same protein.



**FIGURE 4** Ensemble of docked structures for HP dp14 (thin blue sticks) in comparison to the X-ray structures for HP dp2 (thick green sticks) in the complex of CXCL12/HP (PDB ID: 2NWG). Each 2 ns from 60 ns of the MD simulation after the convergence (reached in 30 ns) were used for the visualization

## 4 | CONCLUSIONS

We evaluated the predictive power of the REMD-RS approach for docking GAGs to proteins on a dataset of 22 molecular systems which included also one protein–alginate complex. The starting location of the ligand in MD simulations was defined as far as possible from the experimentally known binding site. For the 19 out of 21 protein–GAG complexes, the experimental oligosaccharide binding site was found correctly. For the complex of AlgE4 with the alginate, dissociation from the protein surface was observed due to the high net negative charge of the protein. Depending on a particular complex, time of the simulation convergence ranged from 1 to 220 ns with 16 complex structures being converged in less than 30 ns. For most of the complexes, the binding poses were predicted correctly in terms of their orientations in the predicted binding site, shown by an overlap of the structural ensembles of the obtained solutions and the corresponding experimental structure. However, due to the fact that in REMD-RS simulations all replicas besides the first one contain bigger van der Waals volumes of the ligand and receptor atoms than the ones in the unmodified force field, the method struggles to dock well into the pocket/groove type of binding sites. Also, the method does not allow to distinguish GAG binding poses in the opposite polarity. Refinement of the docked structures in explicit solvent, nevertheless, essentially improves the quality of docking predictions. Strong correlation between MM-GBSA free binding energy components (especially its electrostatic component in vacuo) and the structural similarity to the experimental structure suggest that the use of MM-GBSA free energy components represent an effective scoring scheme to be used with this approach. The results for the complexes including ions in the binding interface did not stand out from the results for the rest of the systems. The performance of the approach is not affected by the increase of the GAG ligand length as it was the case for conventional

docking approaches making it especially attractive in docking long GAGs. To summarize, the method proposed by Siebenmorgen<sup>44</sup> originally aimed at protein–protein docking, when applied for protein–GAG systems, yielded promising results demonstrating the following advances over other docking methodologies previously applied to the protein–GAG systems: (1) allows full flexibility of both receptor side chains and ligand; (2) provides a reliable scoring scheme; (3) is totally independent of the GAG ligand length rendering it feasible to evaluate the role of GAG binding subsites and prediction of GAG minimal binding units. Our study contributes to the protein–GAG specific tool set, which application allows for an improvement of theoretical analysis quality of these challenging and biologically relevant molecular systems.

## ACKNOWLEDGMENTS

This research was funded by National Science Centre of Poland, grant number UMO-2018/30/E/ST4/00037 and Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 863 – A10 – 111166240. The molecular dynamics simulations were performed on the PROMETHEUS cluster provided by Polish Grid Infrastructure (PL-GRID, gpuaprilgags, plgagaprilgag2 and plggagstr2gpu) as well as on the local “piasek” cluster.

## DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## ORCID

Sergey A. Samsonov  <https://orcid.org/0000-0002-5166-4849>

## REFERENCES

- [1] A. Varki, E. Editor, R. D. Cummings, J. D. Esko, P. Stanley, G. W. Hart, M. Aebi, A. G. Darvill, T. Kinoshita, N. H. Packer, J. H. Prestegard, R. L. Schnaar, P. H. Seeberger, *Essentials of Glycobiology*, 3rd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY) **2015**.
- [2] H. Habuchi, O. Habuchi, K. Kimata, *Glycoconjugate J.* **2004**, *21*, 47.
- [3] C. I. Gama, L. C. Hsieh-Wilson, *Curr. Opin. Chem. Biol.* **2005**, *9*, 609.
- [4] O. Clerc, J. Mariethoz, A. Rivet, F. Lisacek, S. D. PCrossed Signrez, S. Ricard-Blum, *Glycobiology* **2019**, *29*, 36.
- [5] J. Gallagher, *Int. J. Exp. Pathol.* **2015**, *96*, 203.
- [6] N. K. Karamanos, Z. Piperigkou, A. D. Theocharis, H. Watanabe, M. Franchi, S. Baud, S. Brézillon, M. Götte, A. Passi, D. Vigetti, S. Ricard-Blum, R. D. Sanderson, T. Neill, R. V. Iozzo, *Chem. Rev.* **2018**, *118*, 9152.
- [7] P. Chioldelli, A. Bugatti, C. Urbinati, M. Rusnati, *Molecules* **2015**, *20*, 6342.
- [8] F. M. Spinelli, D. L. Vitale, G. Demarchi, C. Cristina, L. Alaniz, *Clin. Transl. Immunol.* **2015**, *4*, e52.
- [9] M. C. Goh, Y. Hwang, G. Tae, *Carbohydr. Polym.* **2016**, *147*, 251.
- [10] S. Morla, *Int. J. Mol. Sci.* **2019**, *20*, 1963.
- [11] M. Yamada, T. Hamaguchi, *J. Biol. Chem.* **2018**, *27*, 10841.
- [12] Z. Scholefield, E. A. Yates, G. Wayne, A. Amour, W. McDowell, J. E. Turnbull, *J. Cell Biol.* **2003**, *163*, 97.
- [13] S. Mehra, D. Ghosh, R. Kumar, M. Mondal, L. G. Gadhe, S. Das, A. Anoop, N. N. Jha, R. S. Jacob, D. Chatterjee, S. Ray, N. Singh, A. Kumar, S. K. Maji, *J. Biol. Chem.* **2018**, *293*, 12975.
- [14] R. S. Aquino, Y. H. F. Teng, P. W. Park, *Biochem. Soc. Trans.* **2018**, *46*, 371.

- [15] S. Y. Kim, W. Jin, A. Sood, D. W. Montgomery, O. C. Grant, M. M. Fuster, L. Fu, J. S. Dordick, R. J. Woods, F. Zhang, R. J. Linhardt, *Antiviral Res.* **2020**, *181*, 104873.
- [16] A. E. I. Proudfoot, *Front. Immunol.* **2015**, *6*, 246.
- [17] J. Shute, *Handb Exp Pharmacol.* **2012**, *207*, 307.
- [18] J. Wigén, L. Elowsson-Rendin, L. Karlsson, E. Tykesson, G. Westergren-Thorsson, *Stem. Cells Dev.* **2019**, *28*, 823.
- [19] C. Faye, C. Moreau, E. Chautard, R. Jetne, N. Fukai, F. Ruggiero, M. J. Humphries, B. R. Olsen, S. Ricard-Blum, *J. Biol. Chem.* **2009**, *284*, 22029.
- [20] D. Xu, J. D. Esko, *Annu. Rev. Biochem.* **2014**, *83*, 129.
- [21] S. D. Vallet, O. Clerc, S. Ricard-Blum, *J. Histochem. Cytochem.* **2020**, *69*, 93.
- [22] D. Scharnweber, L. Hübner, S. Rother, U. Hempel, U. Anderegg, S. A. Samsonov, M. T. Pisabarro, L. Hofbauer, M. Schnabelrauch, S. Franz, J. Simon, V. Hintze, *J. Mater. Sci. Mater. Med.* **2015**, *26*, 232.
- [23] A. Almond, *Curr. Opin. Struct. Biol.* **2018**, *50*, 58.
- [24] N. V. Sankaranarayanan, B. Nagarajan, U. R. Desai, *Curr. Opin. Struct. Biol.* **2018**, *50*, 91.
- [25] S. A. Samsonov, M. T. Pisabarro, *Glycobiology* **2016**, *26*, 850.
- [26] M. Mobli, M. Nilsson, A. Almond, *Glycoconjugate J.* **2008**, *25*, 401.
- [27] S. A. Samsonov, J. Teyra, M. T. Pisabarro, *J. Comput.-Aided Mol. Des.* **2011**, *25*, 477.
- [28] P. R. B. Joseph, P. D. Mosier, U. R. Desai, K. Rajarathnam, *Biochem. J.* **2015**, *472*, 121.
- [29] S. Rother, S. A. Samsonov, T. Hofmann, J. Blaszkiewicz, S. Köhling, S. Moeller, M. Schnabelrauch, J. Rademann, S. Kalkhof, M. von Bergen, M. T. Pisabarro, D. Scharnweber, V. Hintze, *Acta Biomater.* **2016**, *9*, 9539.
- [30] L. Kjellén, U. Lindahl, *Curr. Opin. Struct. Biol.* **2018**, *50*, 101.
- [31] K. Möbius, K. Nordsieck, A. Pichert, S. A. Samsonov, L. Thomas, J. Schiller, S. Kalkhof, M. Teresa Pisabarro, A. G. Beck-Sickinger, D. Huster, *Glycobiology* **2013**, *23*, 1260.
- [32] T. Hofmann, S. A. Samsonov, A. Pichert, K. Lemmnitzer, J. Schiller, D. Huster, M. T. Pisabarro, M. von Bergen, S. Kalkhof, *Methods* **2015**, *89*, 45.
- [33] S. Köhling, G. Künze, K. Lemmnitzer, M. Bermudez, G. Wolber, J. Schiller, D. Huster, J. Rademann, *Chemistry* **2016**, *22*, 5563.
- [34] K. Nordsieck, A. Pichert, S. A. Samsonov, L. Thomas, C. Berger, M. T. Pisabarro, D. Huster, A. G. Beck-Sickinger, *ChemBioChem* **2012**, *13*, 2558.
- [35] U. Uciechowska-Kaczmarzyk, I. de Beauchene, S. A. Samsonov, *J. Mol. Graphics Modell.* **2019**, *90*, 42.
- [36] G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, A. J. Olson, *J. Comput. Chem.* **1998**, *19*, 1639.
- [37] R. R. J. D. T. Moustakas, P. T. Lang, S. Pegg, E. Pettersen, I. D. Kuntz, N. Brooijmans, *J. Comput. Aided Mol.* **2006**, *20*, 601.
- [38] R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis, P. S. Shenkin, *J. Med. Chem.* **2004**, *2*, 84.
- [39] D. Kozakov, D. Beglov, T. Bohnuud, S. E. Mottarella, B. Xia, D. R. Hall, S. Vajda, *Proteins Struct. Funct. Bioinform.* **2013**, *81*, 2159.
- [40] D. Kozakov, D. R. Hall, B. Xia, K. A. Porter, D. Padhorny, C. Yueh, D. Beglov, S. Vajda, *Nat. Protoc.* **2017**, *12*, 255.
- [41] S. A. Samsonov, J. P. Gehrcke, M. T. Pisabarro, *J. Chem. Inf. Model.* **2014**, *55*, 114.
- [42] A. R. Griffith, C. J. Rogers, G. M. Miller, R. Abrol, L. C. Hsieh-Wilson, W. A. Goddard, *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 13697.
- [43] S. A. Samsonov, M. Zacharias, I. C. de Beauchene, *J. Comput. Chem.* **2019**, *40*, 1429.
- [44] T. Siebenmorgen, M. Engelhard, M. Zacharias, *J. Comput. Chem.* **2020**, *41*, 1436.
- [45] A. Laio, M. Parrinello, *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562.
- [46] K. Ostermeir, M. Zacharias, *J. Comput. Chem.* **2014**, *35*, 150.
- [47] S. Ricard-Blum, O. Féraud, H. Lortat-Jacob, A. Rencurosi, N. Fukai, F. Dkhissi, D. Vittet, A. Imberty, B. R. Olsen, M. Van Der Rest, *J. Biol. Chem.* **2004**, *279*, 2927.
- [48] J. Wang, T. Yang, J. Tian, T. Zeng, X. Wang, J. Yao, J. Zhang, Z. Lei, *Carbohydr. Polym.* **2014**, *136*, 527.
- [49] H. J. Rozeboom, T. M. Bjerkan, K. H. Kalk, H. Ertesvåg, S. Holtan, F. L. Aachmann, S. Valla, B. W. Dijkstra, *J. Biol. Chem.* **2008**, *283*, 23819.
- [50] A. D. DiGabriele, I. Lax, D. I. Chen, C. M. Svahn, M. Jaye, J. Schlessinger, W. A. Hendrickson, *Nature* **1998**, *393*, 812.
- [51] S. Faham, R. E. Hileman, J. R. Fromm, R. J. Linhardt, D. C. Rees, *Science (80-)* **1996**, *273*, 18617.
- [52] J. Schlessinger, A. N. Plotnikov, O. A. Ibrahim, A. V. Eliseenkova, B. K. Yeh, A. Yayon, R. J. Linhardt, M. Mohammadi, *Mol. Cell* **2000**, *6*, 743.
- [53] D. Lietha, D. Y. Chirgadze, B. Mulloy, T. L. Blundell, E. Gherardi, *EMBO J.* **2001**, *20*, 5543.
- [54] V. K. Ganesh, S. A. Smith, G. J. Kotwal, K. H. M. Murthy, *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 8924.
- [55] W. J. Carter, E. Cama, J. A. Huntington, *J. Biol. Chem.* **2005**, *280*, 2745.
- [56] C. Shao, F. Zhang, M. M. Kemp, R. J. Linhardt, D. M. Waisman, J. F. Head, B. A. Seaton, *J. Biol. Chem.* **2006**, *281*, 31689.
- [57] S. Banerji, A. J. Wright, M. Noble, D. J. Mahoney, I. D. Campbell, A. J. Day, D. G. Jackson, *Nat. Struct. Mol. Biol.* **2007**, *14*, 234.
- [58] Z. Li, M. Kienetz, M. M. Cherney, M. N. G. James, D. Brömme, *J. Mol. Biol.* **2008**, *383*, 78.
- [59] A. H. Aguda, P. Panwar, X. Du, N. T. Nguyen, G. D. Brayer, D. Brömme, *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 17474.
- [60] Y. H. Han, M. L. Garron, H. Y. Kim, W. S. Kim, Z. Zhang, K. S. Ryu, D. Shaya, Z. Xiao, C. Cheong, Y. S. Kim, R. J. Linhardt, Y. H. Jeon, M. Cygler, *J. Biol. Chem.* **2009**, *284*, 34019.
- [61] J. Herrou, C. Bompard, R. Wintjens, E. Dupré, E. Willery, V. Villeret, C. Lochet, R. Antoine, F. Jacob-Dubuisson, *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 17351.
- [62] A. Cartmell, E. C. Lowe, A. Baslé, S. J. Firbank, D. A. Ndeh, H. Murray, N. Terrapon, V. Lombard, B. Henrissat, J. E. Turnbull, M. Czjzek, H. J. Gilbert, D. N. Bolam, *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 7037.
- [63] D. M. Whalen, T. Malinauskas, R. J. C. Gilbert, C. Siebold, *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 16420.
- [64] I. Capila, M. J. Hernáiz, Y. D. Mo, T. R. Mealy, B. Campos, J. R. Dedman, R. J. Linhardt, B. A. Seaton, *Structure* **2001**, *9*, 57.
- [65] J. W. Murphy, Y. Cho, A. Sachpatzidis, C. Fan, M. E. Hodsdon, E. Lolis, *J. Biol. Chem.* **2007**, *282*, 10018.
- [66] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, *Proteins: Struct. Funct. Genet.* **2006**, *65*, 712.
- [67] K.-P. Hopfner, L. Craig, G. Moncalian, R. A. Zinkel, T. Usui, B. A. Owen, A. Karcher, B. Henderson, J.-L. Bodmer, C. T. McMurray, J. P. Carney, J. H. J. Petrini, J. A. Tainer, *Nature* **2002**, *418*, 562.
- [68] H. Nguyen, D. R. Roe, C. Simmerling, *J. Chem. Theory Comput.* **2013**, *9*, 2020.
- [69] A. Onufriev, D. Bashford, D. A. Case, *Proteins* **2004**, *55*, 383.
- [70] R Core Team, (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org>
- [71] Schrodinger, LLC, PyMOL Mol. Graph. Syst. Version 1.3r1 **2010**.
- [72] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graph.* **1996**, *14*, 33.
- [73] J. C. Muñoz-García, M. J. García-Jiménez, P. Carrero, Á. Canales, J. Jiménez-Barbero, M. Martín-Lomas, A. Imberty, D. Paz, J. L. J. Angulo, H. Lortat-Jacob, P. M. Nieto, *Glycobiology* **2014**, *24*, 1004.
- [74] S. Babik, S. A. Samsonov, M. T. Pisabarro, *Glycoconjugate J.* **2017**, *34*, 427.
- [75] A. Imberty, H. Lortat-Jacob, S. Pérez, *Carbohydr. Res.* **2007**, *342*, 430.
- [76] K. K. Bojarski, A. K. Sieradzan, S. A. Samsonov, *Biopolymers* **2019**, *110*, e23252.
- [77] J. Vuorio, I. Vattulainen, H. Martinez-Seara, *PLoS Comput. Biol.* **2017**, *13*, e1005663.
- [78] K. K. Bojarski, S. A. Samsonov, *J. Chem. Inf. Model.* **2020**, *60*, 2247.



- [79] M. R. Shirts, J. D. Chodera, *J. Chem. Phys.* **2008**, *129*, 124105.
- [80] T. Siebenmorgen, M. Zacharias, *J. Chem. Inf. Model.* **2020**, *60*, 5552.
- [81] J. Potthoff, K. K. Bojarski, G. Kohut, A. G. Lipska, A. Liwo, E. Kessler, S. Ricard-Blum, S. A. Samsonov, *Int. J. Mol. Sci.* **2019**, *20*, 5021.
- [82] M. M. Kogut, M. Maszota-Zieleniak, M. Marcisz, S. Samsonov, *Phys. Chem. Chem. Phys.* **2021**, *23*, 3519.
- [83] F. Lecaillon, T. Chazeirat, K. K. Bojarski, J. Renault, A. Saidi, V. G. N. V. Prasad, S. Samsonov, G. Lalmanach, *Biochim. Biophys. Acta Proteins Proteomics* **2020**, *1868*, 140318.
- [84] N. Panitz, S. Theisgen, S. A. Samsonov, J. P. Gehrcke, L. Baumann, K. Bellmann-Sickert, S. Köhling, M. Teresa Pisabarro, J. Rademann, D. Huster, A. G. Beck-Sickinger, *Glycobiology* **2016**, *26*, 1209.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

**How to cite this article:** 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. <https://doi.org/10.1002/jcc.26496>

## Publication D2

Explicit solvent repulsive scaling replica  
exchange molecular dynamics (RS-REMD)  
in molecular modeling of protein-  
glycosaminoglycan complexes

## RESEARCH ARTICLE

# Explicit solvent repulsive scaling replica exchange molecular dynamics (RS-REMD) in molecular modeling of protein-glycosaminoglycan complexes

Mateusz Marcisz<sup>1,2</sup> | Margrethe Gaardl s<sup>1</sup> | Krzysztof K. Bojarski<sup>1,3</sup> |  
Till Siebenmorgen<sup>4</sup> | Martin Zacharias<sup>5</sup> | Sergey A. Samsonov<sup>1</sup> 

<sup>1</sup>Faculty of Chemistry, University of Gdańsk, Gdańsk, Poland

<sup>2</sup>Intercollegiate Faculty of Biotechnology, University of Gdańsk and Medical University of Gdańsk, Gdańsk, Poland

<sup>3</sup>Department of Physical Chemistry, Gdańsk University of Technology, Gdańsk, Poland

<sup>4</sup>Institute of Structural Biology, Helmholtz Zentrum M nchen, Neuherberg, Germany

<sup>5</sup>Physics Department, Technical University of Munich, Garching, Germany

## Correspondence

Martin Zacharias, Physics Department, Technical University of Munich, James-Frank-Strasse 1, 85748 Garching, Germany.  
Email: zacharias@tum.de

Sergey A. Samsonov, Faculty of Chemistry, University of Gdańsk, ul. Wita Stwosza 63, 80-308 Gdańsk, Poland.  
Email: sergey.samsonov@ug.edu.pl

## Funding information

Narodowe Centrum Nauki, Grant/Award Number: UMO-2018/30/E/ST4/00037

## Abstract

Glycosaminoglycans (GAGs), linear anionic periodic polysaccharides, are crucial for many biologically relevant functions in the extracellular matrix. By interacting with proteins GAGs mediate processes such as cancer development, cell proliferation and the onset of neurodegenerative diseases. Despite this eminent importance of GAGs, they still represent a limited focus for the computational community in comparison to other classes of biomolecules. Therefore, there is a lack of modeling tools designed specifically for docking GAGs. One has to rely on existing docking software developed mostly for small drug molecules substantially differing from GAGs in their basic physico-chemical properties. In this study, we present an updated protocol for docking GAGs based on the Repulsive Scaling Replica Exchange Molecular Dynamics (RS-REMD) that includes explicit solvent description. The use of this water model improved docking performance both in terms of its accuracy and speed. This method represents a significant computational progress in GAG-related research.

## KEYWORDS

explicit solvent, glycosaminoglycan modeling, molecular docking, molecular dynamics, repulsive scaling replica exchange

## 1 | INTRODUCTION

Glycosaminoglycans (GAGs) are heterogeneous long linear periodic anionic polysaccharides that consist of repeating disaccharide units.<sup>1</sup> The components of these disaccharide units may exhibit different sulfation patterns, which alter their conformational properties and binding characteristics.<sup>2,3</sup> GAGs are important players of the extracellular matrix where they are crucial for a variety of the biological functions via their interactions with protein targets. In particular, GAGs influence processes such as tissue regeneration,<sup>4,5</sup> cancer development,<sup>6,7</sup> infection,<sup>8,9</sup> cardiovascular diseases,<sup>10</sup> cell maturation<sup>11</sup> and proliferation,<sup>12</sup> Alzheimer's and Parkinson's diseases<sup>13</sup> and inflammatory response.<sup>14</sup> Two major groups of proteins that interact with GAGs are growth factors<sup>15–17</sup> and chemokines.<sup>18–20</sup> While some protein-GAG interactions are considered highly specific<sup>21</sup>

and selective,<sup>22</sup> many other are predominantly driven by non-specific electrostatic interactions.<sup>23,24</sup>

Considering GAG structural properties and the fact that GAG binding regions on the protein surface are not always well-defined, docking of GAGs may be immensely challenging.<sup>24</sup> Additionally, computational GAG research is still lagging behind that of small drug molecules, peptides and proteins as only a limited group of scientists are involved in GAG-related studies. Although there is a number of docking software that can be applied for docking GAGs, most of them do not perform at satisfactory level for these systems.<sup>25,26</sup> One of the conventional docking tools, Autodock3 (AD3),<sup>27</sup> was proven to work effectively in several GAG-protein complexes.<sup>25,26</sup> However, AD3 has an important limitation, the maximum number of torsional degrees of freedom considered in the calculations, which narrows effective docking to GAGs with a degree of polymerization (dp) of up to 8. To tackle

this challenge, a few new docking tools emerged to account for long GAG flexibility. One of them is Dynamic Molecular Docking (DMD),<sup>28</sup> which is a steered MD-based method where the GAG is moved from a distant position towards the binding site on the protein surface by application of an additional potential defined by the distance between two groups of specific atoms in the protein and in the GAG. DMD allows for fully flexible docking of long GAG molecules, but it is a local docking approach and it is very demanding in terms of required computational resources. Another docking tool created to face the challenge of flexible docking of long molecules is Replica Exchange Molecular Dynamics with Repulsive Scaling (RS-REMD)<sup>29,30</sup> which was originally designed for protein–protein docking by Siebenmorgen et al. It was later implemented for GAG docking to proteins and proved to be successful at this task,<sup>31</sup> even for GAGs of dp24 and dp48.<sup>32</sup> In RS-REMD, van der Waals radii are scaled in different replicas (without influencing other types of interactions), which allows for an extensive and robust search of the binding sites and proper sampling of GAG conformations on the protein surfaces. Additionally, both protein sidechains and GAGs remain flexible. This method is especially suited for GAGs, because electrostatics play a crucial role in protein–GAG complexes by establishing strong charge–charge interactions. Therefore, affecting van der Waals radii in different replicas allows for a drastically faster scanning of the protein surface, avoiding potential local minima and providing effective prediction of GAG binding sites defined mainly by electrostatics.

In this work, we propose an improved RS-REMD protocol that includes an explicit water model that is expected to be superior to the implicit model in terms of docking quality due to more realistic description of the interactions between solvent and solutes.<sup>33</sup> To test this, three complexes were rigorously analyzed in terms of docking performance: Acidic Fibroblast Growth Factor (FGF1)–heparin (HP) dp6, Basic Fibroblast Growth Factor (FGF-2)–HP dp6, and Antithrombin III (ATIII)–HP dp8. We observed an improvement in the performance of docking GAG molecules in these complexes: RMSatd (root mean squared atom type deviation; described in details in Materials and Methods) for the 10 most energetically favorable poses decreased from  $6.7 \pm 5.3$ ,  $5.4 \pm 1.3$ , and  $10.9 \pm 8.1$  Å to  $5.4 \pm 1.1$ ,  $4.4 \pm 0.5$ , and  $2.5 \pm 0.2$  Å when comparing results obtained with implicit and explicit solvent model, respectively. Therefore, the improvement of the docking results was by 1.3, 1.0, and 8.4 Å for the tested systems. It was also found that despite using a more accurate explicit water model there was no observed increase in computational expenses, and thus this upgraded protocol is not more demanding in terms of computational resources than the previously described one deploying an implicit solvent model. Since all the steps are performed with the explicit water model, after the production runs there was no need for the additional refinement required in the RS-REMD protocols with implicit solvent. This new approach contributes to a pool of new docking programs targeted for protein–GAG interactions that can gear the progress in the field of GAG-driven drug design and regenerative medicine, as well as cancer research.

## 2 | MATERIALS AND METHODS

### 2.1 | Structures and parameters

All the experimental structures used for this work are obtained from the Protein Data Bank (PDB): 1BFC (Basic Fibroblast Growth Factor complexed with HP dp6)<sup>16</sup>; 2AXM (dimer of Fibroblast Growth Factor complexed with HP dp6),<sup>17</sup> monomeric protein form was used as in original paper for the RS-REMD GAG docking method with the implicit solvent<sup>31</sup>; 1E03 (plasma alpha Antithrombin III complexed with HP dp5 derivative, in this work HP dp8 was used).<sup>34</sup> ff14SBonlysc force field parameters<sup>35</sup> were used for the proteins. Literature data for the sulfate groups<sup>36</sup> and GLYCAM06<sup>37</sup> force field parameters were used for GAGs.

### 2.2 | RS-REMD docking simulations

All the simulations have been performed in AMBER20 package.<sup>38</sup> The original protocol used for protein–protein docking was described by Siebenmorgen et al.<sup>29</sup> Implemented protocol for the GAG to protein docking was described in details in the work of Maszota-Zieleniak et al.<sup>31</sup> In brief, the ff14SBonlysc<sup>39</sup> force field parameters for protein and the GLYCAM06<sup>37</sup> for GAGs were used. GAGs were placed at a random position on the protein surface far from the experimentally known binding site. In the original protocol,<sup>31</sup> MD docking runs were performed in implicit solvent using igb model 8<sup>40</sup> and infinite cutoff for the nonbonded interactions. 3000 steps of steepest descent and 3000 steps of conjugate gradient were performed as a minimization. This was followed by heating up the system to 300 K using Langevin thermostat. In the main RS-REMD simulations the harmonic restrains ( $0.05 \text{ kcal/mol/Å}^2$ ) were used on a protein's heavy atoms and additionally the half parabolic distance restraints of  $1.0 \text{ kcal/mol/Å}^2$  were used between COM of the ligand and receptor to avoid complete dissociation from the protein. In 16 replicas, Lennard–Jones (LJ) parameters were altered. Parameter  $d$  affecting the effective van der Waals radius was set to 0.00, 0.01, 0.02, 0.04, 0.08, 0.12, 0.16, 0.20, 0.24, 0.28, 0.32, 0.38, 0.44, 0.50, 0.58, 0.68 Å, while the factor  $e$  adjusting the LJ well depths was assigned to 1.00, 0.99, 0.98, 0.97, 0.96, 0.94, 0.92, 0.90, 0.88, 0.86, 0.84, 0.82, 0.80, 0.78, 0.76, 0.74, respectively. After the RS-REMD docking using implicit solvent was completed additional refinement using explicit solvent was performed to achieve more appropriate ligand conformations with a more advanced water model and to properly relax the system which is hindered in this RS-REMD simulation due to the combination of bigger van der Waals radii in high replicas with the implicit solvent model. A 100 energetically best poses were chosen, and short 5 ns MD unrestrained simulations were performed.<sup>31</sup>

In the current work, the protocol is adapted to explicit solvent using the explicit water model TIP3P.<sup>41</sup> Due to the introduction of the explicit water model, prior to the RS-REMD runs, minimization and equilibration steps were also performed using explicit water model. An 8 Å layer as a truncated octahedron made of TIP3P water molecules was added to solvate complexes. To neutralize the charge of the

system,  $\text{Na}^+/\text{Cl}^-$  counterions were added.<sup>42</sup> Preceding RS-REMD runs energy minimization was performed consisting of 500 steepest descent cycles and  $10^3$  conjugate gradient cycles with 100 kcal/mol/ $\text{\AA}^2$  harmonic force restraint on solute atoms followed by  $3 \times 10^3$  steepest descent cycles and  $3 \times 10^3$  conjugate gradient cycles without any restraints. This was continued by heating of the system to 300 K for 10 ps with harmonic force restraints of 100 kcal/mol/ $\text{\AA}^2$  on solute atoms. Later, the system was equilibrated at 300 K and  $10^5$  Pa in an isothermal isobaric ensemble for 500 ps. All other details and steps not mentioned here were the same as in the original version of the method.<sup>31</sup> Examples of the new input files including all parameters used with the explicit water model are provided in the Supplementary Materials.

Five independent RS-REMD simulations in explicit solvent were performed for the complexes IBFC and 2AXM, and results from implicit solvent simulations of these complexes were reported earlier.<sup>31</sup> 1E03 with HP dp8 was subjected to RS-REMD simulations for the first time, and in addition to five independent explicit solvent simulations, a simulation in implicit solvent was also performed for this complex.

We also performed five independent RS-REMD simulations for FGF-2-HP dp6 system (using 1BFC pdb structure as reference) in explicit solvent with additional potentials applied to HP glycosidic linkages. These potentials were defined by the following expressions:

$$E(\phi) = k_1[1 + \cos(n_1\phi - m_1)]$$

$$E(\psi) = k_2[1 + \cos(n_2\psi - m_2)] \quad (1)$$

where  $\phi$  and  $\psi$  are dihedral angles describing glycosidic linkages, as it was defined in the work of Satelle et al.<sup>43</sup>  $k$  is the force constant,  $n$  is periodicity and  $m$  is phase shift. The parameters for these potentials are based on the glycosidic linkage energy maps<sup>44</sup> and their respective values are summarized in Supplementary Materials (Supplementary Table 2). These parameters were fitted in a way that for each replica, the maxima of the potential function corresponded to the glycosidic linkage energy minima. For these RS-REMD simulations we applied the protocol described above.

### 2.3 | Binding free energy calculations

MM/GBSA (molecular mechanics generalized born surface area) model  $\text{igb} = 2$ <sup>45</sup> from AMBER20<sup>38</sup> was used for per residue decomposition and free energy calculations on the trajectories obtained from RS-REMD simulations. These energy values should be understood as approximate binding free energy values.

### 2.4 | RMSD and RMSatd evaluation

RMSD (root mean square deviation) and RMSatd (root mean square atom type deviation) were used as metrics for the structural similarity between obtained results and the crystal structure. RMSatd is defined

as the root-mean square of pairwise atomic distances while pairing up the spatially closest atoms of the same type. Due to the periodic nature of the GAGs, this method should be more appropriate than classical RMSD analysis.<sup>28</sup> RMSatd accounts for the periodicity of the functional groups in the GAG molecules by claiming that two GAGs shifted by several periodic units are structurally similar. In the same case, the classical RMSD distance metric would consider those two structures significantly different. Additionally, RMSatd can be used for comparison of molecules with different numbers of atoms. This is very important in the analysis of GAG docking site due to the fact that often only a fragment of the longer periodic GAG molecule used for crystallization is visible in the crystal structures.

## 3 | RESULTS AND DISCUSSION

In the protocol prepared by Maszota-Zieleniak et al.<sup>31</sup> an implicit solvent model was used during RS-REMD docking, followed by a short 5 ns refinement with an explicit TIP3P water model. The new protocol presented in the current study uses TIP3P water through the whole RS-REMD docking protocol, thus there is no need for additional refinement for the best poses after the docking. Additionally, the new procedure is not more time consuming in terms of computational resources and the wall-clock time. Usually, docking is stopped when convergence is observed: ligand remains near one particular binding site and does not essentially move away from it. In this study, however, we continued the simulations after achieving convergence to assess the computational performance of the method and to prove that no changes in the binding pose occur. The RS-REMD simulations, when using Nvidia Tesla K40d gpu cards (16 per job, one per replica), typically finished after less than 3 days (65–69 h, the range means that the calculation times were slightly different for independent RS-REMD runs), yielding 500 ns for the 1BFC and 2AXM complexes. For the 1E03 complex, 220–240 ns, which was a sufficient time for convergence, were calculated in 72 h. With the implicit solvent it took 72 h on the same GPUs to complete around 300 ns of the simulations for the 1BFC and 2AXM and around 190–200 ns for the 1E03. The reason for this could be that in the case of implicit solvent we used additional restraints to keep proper puckering of GAG rings that are often distorted during MD simulation in implicit solvent, which essentially affected the time of computations. Arguably, the new method can be even faster when a user aims to analyze many best scored poses due to the absence of additional refinement that is required when using the implicit solvent RS-REMD docking. In the latter, a standard procedure for one hundred best poses requires running one hundred 5 ns refinement runs in explicit solvent, which is time consuming both in terms of file preparation efforts and computational resources.

### 3.1 | RS-REMD docking performance

RMSatd values for the docking poses obtained with the new RS-REMD procedure with explicit solvent for the analyzed complexes

**TABLE 1** RMSatd values (Å) for the energetically best poses obtained in the docking by RS-REMD with explicit solvent in comparison to the corresponding data obtained with the implicit solvent RS-REMD approach by Maszota-Zieleniak et al.<sup>31</sup> For 1E03, implicit solvent simulations were performed in the current study.

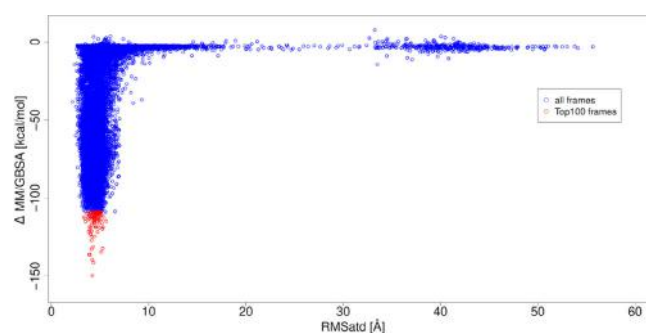
Simulation Explicit	1BFC				2AXM				1E03			
	Top10	sd	Top100	sd	Top10	sd	Top100	sd	Top10	sd	Top100	sd
1	4.9	0.6	5.1	0.9	4.5	0.8	5.0	0.7	2.8	0.5	3.0	1.0
2	5.8	0.6	5.9	0.9	4.1	0.5	4.1	0.5	2.4	0.1	2.5	0.2
3	4.0	0.6	5.1	1.5	5.1	1.0	4.5	0.8	2.4	0.3	2.5	0.5
4	7.1	0.3	6.9	0.4	3.9	0.9	3.9	0.9	2.3	0.1	2.4	0.2
5	5.2	0.7	5.1	0.7	4.4	0.5	4.5	0.5	2.7	0.2	2.9	0.6
Average	5.4	0.6	5.6	0.9	4.4	0.7	4.4	0.7	2.5	0.2	2.7	0.3
Implicit	Top10	sd	Top100	sd	Top10	sd	Top100	sd	Top10	sd	Top100	sd
Average	6.7	5.3	8.7	6.1	5.4	1.3	4.9	1.3	10.9	8.1	21.6	6.4

Abbreviation: sd, standard deviation.

(Table 1) were compared to the values obtained with the method with implicit solvent. First, we performed MM/GBSA analysis for all the frames from RS-REMD simulation. Those docking results were ranked in terms of their free binding energies. Then, as in the original paper<sup>31</sup> on GAG docking with RS-REMD method with implicit solvent, the 10 and 100 energetically best poses, defined as Top10 and Top100 poses, respectively, were chosen. Afterwards, RMSatd was calculated for those Top10/100 poses. Average RMSatd values for the 1BFC, 2AXM and 1E03 were  $5.4 \pm 1.1$ ,  $4.4 \pm 0.5$ , and  $2.5 \pm 0.2$  Å, respectively, in case of Top10 poses using the explicit solvent model, which were lower than RMSatd obtained with the implicit solvent model. With the implicit solvent Top10 yielded following values for the 1BFC, 2AXM and 1E03:  $7.3 \text{ Å} \pm 0.5 \text{ Å}$  and  $6.7 \pm 5.3 \text{ Å}$ ,<sup>31</sup>  $5.8 \pm 0.3 \text{ Å}$  and  $5.4 \pm 1.3 \text{ Å}$ ,<sup>31</sup>  $22.4 \pm 7.8 \text{ Å}$  and  $10.9 \pm 8.1 \text{ Å}$  before and after refinement, respectively. The following Top100 data were obtained from RS-REMD docking for the 1BFC, 2AXM and 1E03:  $5.6 \pm 0.9 \text{ Å}$ ,  $4.4 \pm 0.6 \text{ Å}$ ,  $2.7 \pm 0.2 \text{ Å}$  in case of explicit solvent. Top100 with the implicit solvent for the 1BFC, 2AXM and 1E03 yielded following results:  $9.7 \pm 1.4 \text{ Å}$  and  $8.7 \pm 6.1 \text{ Å}$ ,<sup>31</sup>  $8.0 \pm 0.5 \text{ Å}$  and  $4.9 \pm 1.3 \text{ Å}$ ,<sup>31</sup>  $21.6 \pm 6.4 \text{ Å}$  and  $11.5 \pm 8.8 \text{ Å}$  before and after refinement, respectively.

Alternatively, instead of selecting top-scored poses, one can cluster the structures obtained in the MD simulations after the convergence is reached, and further choose the representatives of the clusters for classical MD simulations with the MM/GBSA free energy calculations. This potentially could provide more accurate free energy estimates of the particular binding poses since in such calculations multiple frames would be used in comparison to a single frame corresponding to each of the structures within Top10 and Top100 solutions. However, such a “classical” approach for docking scoring would require extensively more computational resources.

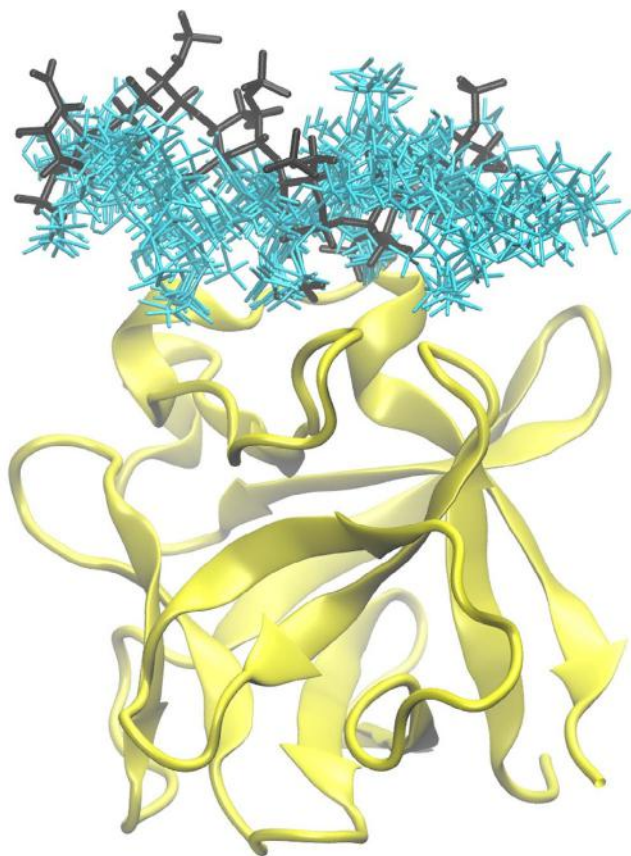
Additionally, we compared our results with AD3 docking results obtained by Samsonov and Pisabarro for 50 top scored solutions from 1000 AD3 docking poses.<sup>46</sup> In case of 1BFC RS-REMD method yielded slightly worse results than AD3 ( $5.4 \pm 1.1 \text{ Å}$  vs.  $4.8 \pm 1.4 \text{ Å}$ ). However, in case of 2AXM and 1E03 RS-REMD allowed for much



**FIGURE 1** MM/GBSA binding free energies in the protein-GAG complex 2AXM and RMSatd values of the GAG molecule (with the reference to the crystal structure).

better quality of docking when compared with AD3:  $4.4 \pm 0.5 \text{ Å}$  vs.  $7.3 \pm 2.5 \text{ Å}$  and  $2.5 \pm 0.2 \text{ Å}$  vs.  $4.6 \pm 1.4 \text{ Å}$ , respectively. It should be noticed that in the work of the Samsonov and Pisabarro the AD3 docking was not performed for the whole protein but only for the part of the protein that included the experimentally known binding site.

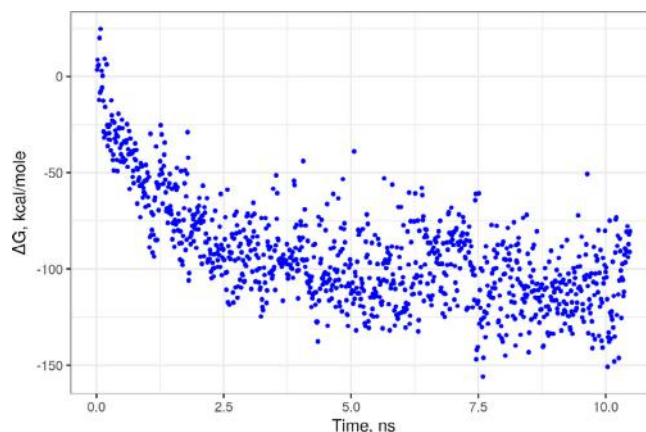
Importantly, all the obtained docking poses with high RMSatd are energetically unfavorable (Figures 1 and S1). Therefore, in a practical case when there are no *a priori* data on the GAG binding site, no poses would be selected for further analysis that are far outside the docking region obtained using MM/GBSA evaluation. However, it should be admitted that some poses with very low RMSatd were not among the top-scored. The reason for this could be the fact that binding energies in higher replicas are much higher than in lower replicas for the same or similar ligand poses. Due to increased atomic van der Waals radii in higher replicas, after the exchange from the lower replica some ligand atoms can be too close to protein atoms and cause unfavorable spatial overlap. For the case when slightly higher RMSatd values corresponding to favorable energetic values are observed, the explanation could be that in the PDB structures only a small fragment of the periodic GAG molecule is resolved. This means that some fragments of the GAG molecule are not resolved, and this results in decreased



**FIGURE 2** Top 10 binding poses (in cyan) from the RS-REMD docking for the 2AXM complex. HP dp6 from the X-ray structure (in black) for the reference is visualized in the binding site.

reliability of the RMSD/RMSatd analysis. We believe that when we observe especially favorable energetic interactions for poses that are not top-rated in terms of the RMSatd, these poses could correspond to the ones adopted by longer GAG molecules (made up of more repetitive periodic units). Therefore, RS-REMD could suggest how the extension of the binding site, beyond what was captured in the crystal structure with its X-ray method limitations, could look like to accommodate longer GAG oligosaccharides. Furthermore, GAGs display high flexibility, periodicity and a particular property that is termed multipose binding.<sup>47</sup> This means that they can bind multiple sites on the protein surface and/or in multiple poses in one binding region (e.g., at different angles between the GAG chains in those poses). This phenomenon, depending on the range of flexibility and conformational freedom of the GAG upon binding to the protein, may be also called either plastic or non-selective binding.<sup>24</sup>

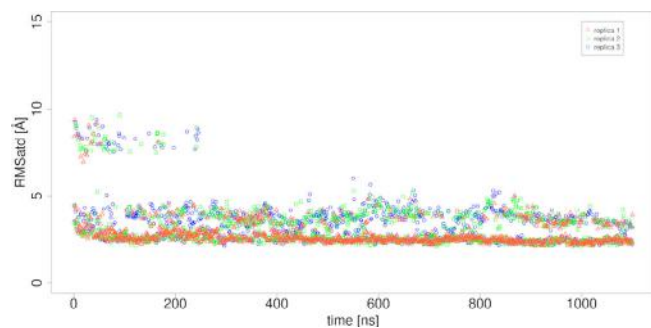
RS-REMD simulations were also performed for 1BFC system with additional potential functions imposed on dihedral angles that compose glycosidic linkages, to check if additionally allowed flexibility of the GAG chains could lead to the improvement of the docking results. With subsequently higher replicas, the height of the energy barrier decreased. This potential function was included in order to force GAG to change its conformation on higher replicas, which, in turn could improve the search of its conformational space. The



**FIGURE 3** Binding energy values obtained with MM/GBSA calculations in the course of the MD simulation for the first replica from the RS-REMD docking.

results obtained with this approach are comparable with the data discussed above (Table S3). In particular, for most replicas we observed that the Top10 structures based on their respective binding free energies yielded better results than Top100. The obtained results might suggest that glycosidic linkages are able to adapt their preferable conformation already in the unbiased RS-REMD simulations. The comparison of the two types of glycosidic linkages for the lowest and highest replicas, named type 1 (between GlcNS(6S)-IdoA(2S)) and type 2 (between IdoA(2S)-GlcNS(6S)), revealed differences in the preferable conformations (Figure S5). For type 1, in the lowest replica a few local minima were present, while for the highest, one well-defined minimum was defined. For type 2, for both the lowest and the highest replica, one minimum was present. However, for this glycosidic linkage type the difference in the position of this minimum is significant. The distribution of dihedral angles suggest that these additional potential function might increase the available conformational space during the RS-REMD simulation. These potentials might be essential in a scenario in which dihedral angles present in reference structure correspond to local minima where they are trapped.

The Top10 poses from each RS-REMD run were visually analyzed. The observed poses look very similar, and they all are located in the same specific region on the protein surfaces (Figures 2 and S2). Additionally, they correspond to low RMSD between themselves (4.2 Å, 3.9 Å, 1.4 Å for 1BFC, 2AXM and 1E03, respectively; Table S1) which suggests that they all belong to the same structural cluster. Considering that the GAGs observed in computational studies often reveal high mobility and plasticity of the binding on the protein surface<sup>24,46</sup> the results should be interpreted essentially differently than when analyzing binding of small drug molecules. During conventional MD runs starting from the crystal structure with the same force field parameters, observed GAG RMSD (using the crystal structure as a reference) was  $4.1 \text{ Å} \pm 0.6$  for the 1BFC complex,  $5.2 \pm 1.2 \text{ Å}$  for the 2AXM dimer complex and  $1.4 \pm 0.3 \text{ Å}$  for the 1E03 complex.<sup>46</sup> For the 2AXM monomeric variant, an RMSD of  $\sim 4 \text{ Å}$  was observed.<sup>44</sup> In



**FIGURE 4** GAG RMSatd values (with the reference to the crystal structure) for the first 3 replicas from the representative RS-REMD simulation of the 1E03 complex.

the work of Sankaranarayanan et al., they observed a characteristic GAG RMSD fluctuation of 2.83 Å for the Transforming Growth Factor- $\beta$ 2 with chondroitin sulfate dp6.<sup>48</sup> This indicates that the variance of the energetically most favored poses from RS-REMD docking can be even lower than natural movements of the GAG molecules for the MD starting from their X-ray binding pose, and the movements are associated with imperfections of the force fields. This suggests a remarkable specificity of the structures obtained in this study, which together with the accuracy of the method renders its performance superior to all other previously applied methods in the field of protein-GAG docking.

Additionally, we observed a decrease in RMSatd values and the free binding energy already after the first nanoseconds of the RS-REMD simulation (Figures 3 and 4, S3 and S4), suggesting a more favorable binding in comparison to the data from the implicit RS-REMD approach. This means that after a relatively short time of RS-REMD-based docking, the potential region for GAG binding could be detected.

There may be several reasons for the improvement of the docking performance with explicit solvent in comparison to the implicit one. Water is crucial for the interactions of GAGs and protein molecules. In the available experimental structures, the amount of the water molecules in protein-GAG complexes is about 10 times higher than in protein-protein complexes.<sup>49</sup> Furthermore, about a half of all protein-GAG residue contacts in the PDB are water-mediated,<sup>49</sup> and due to the importance of electrostatic interactions in protein-GAG complexes accurate modeling of water-mediated interactions is exceptionally important.<sup>33</sup> Many studies reported relevance of the dynamical behavior of the solvent surrounding on the saccharides conformations.<sup>50–52</sup> It was shown in multiple studies that explicit model improves molecular docking quality<sup>53–56</sup> and that the including explicit water model to MD simulations and docking may be immensely beneficial.<sup>33,57–60</sup> A few studies reported that water molecules bridge protein-GAG interactions and may function as structured water helping to recognize and stabilize the interactions.<sup>22,61–64</sup> Therefore, we believe that explicit solvent is a more advanced water model can more accurately mediate interactions between protein and GAG molecules. Additionally, TIP3P water model is widely accepted

and used in protein-GAG studies and proved to work well in this type of systems.<sup>65–67</sup>

To reveal potential effects of the solvent model on GAG conformational space we performed glycosidic linkage analysis of the octameric heparin from 1E03 complex for the first replica from the implicit solvent and explicit solvent RS-REMD simulations (Figure S6). The obtained data suggest that the glycosidic linkage sampling was very similar for both protocols. However, for type 1 linkages the implicit solvent runs show only one local minimum, whereas the explicit solvent sometimes samples two different minima. This suggests that a slightly more advanced conformational space sampling with the explicit solvent.

## 4 | CONCLUSIONS

Improvement of the RS-REMD method by introducing explicit solvent in the form of TIP3P water model is a leap forward in an already well-functioning docking tool. This new refined procedure both improves docking performance reflected by lower RMSatd values of the obtained structures, and decreases the computational time required for the simulation. Performance enhancement in terms of the structural similarity of the docked poses to the experimental structures was achieved by implementing a physically more relevant water model. An efficiency boost in terms of computational time was possible due to removal of restraints responsible for ring puckering, since their conformational space is not distorted within the TIP3P water model. Additionally, in simulations with the explicit model we used a nonbonded cut-off of 8 Å, while in the implicit solvent the cut-off is set to 999 Å, which also contributed to the gain in computational time. Therefore, in bigger systems with implicit solvent, more time to calculate protein-GAG interactions may be required. Importantly, unlike many other docking tools, this one does not suffer from increasing the size of the docked ligand and was proven working well for system with dp24 and dp48 GAGs before.<sup>32</sup> This is of great importance given that in nature GAG molecules are often of length of thousands units, and there is a lack of computational works that cover this area of interactions of protein and GAG molecules longer than dp6–10. We believe the new RS-REMD method with the explicit solvent should also be easily implemented to other systems with predominantly electrostatic driven interactions and long linear ligands, and could be used as a promising tool in the field of drug design with its further implication for GAG-involved regenerative medicine.

## ACKNOWLEDGMENTS

This research was funded by the National Science Centre of Poland, grant number UMO-2018/30/E/ST4/00037. The molecular dynamics simulations were performed on the PROMETHEUS cluster provided by Polish Grid Infrastructure (PL-GRID, plggpugagsoftware, plgprokat) as well as on the local “piasek” cluster.

## DATA AVAILABILITY STATEMENT

Data available on request from the authors.



## ORCID

Sergey A. Samsonov  <https://orcid.org/0000-0002-5166-4849>

## REFERENCES

- [1] C. Bu, L. Jin, *Front. Mol. Biosci.* **2021**, *8*, 646808. <https://doi.org/10.3389/fmolb.2021.646808>
- [2] H. Habuchi, O. Habuchi, K. Kimata, *Glycoconjugate J.* **2004**, *21*(1–2), 47.
- [3] A. Varki, R. D. Cummings, J. D. Esko, P. Stanley, G. W. Hart, M. Aebi, D. Mohnen, T. Kinoshita, N. H. Packer, J. H. Prestegard, R. L. Schnaar, P. H. Seeberger, *Essentials of Glycobiology*, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY) **2015**.
- [4] C. Paganini, R. Costantini, A. Superti-Furga, A. Rossi, *FEBS J.* **2019**, *286*, 3008.
- [5] J. Salbach, T. D. Rachner, M. Rauner, U. Hempel, U. Anderegg, S. Franz, J. C. Simon, L. C. Hofbauer, *J. Mol. Med.* **2012**, *90*, 625.
- [6] W. Risau, *Nature* **1997**, *386*, 671.
- [7] S. N. Ma, Z. X. Mao, Y. Wu, M. X. Liang, D. D. Wang, X. Chen, P. Chang, W. Zhang, J. H. Tang, *Cell Adhes. Migrat.* **2020**, *14*, 118.
- [8] N. K. Karamanos, Z. Piperigkou, A. D. Theocharis, H. Watanabe, M. Franchi, S. Baud, S. Brézillon, M. Götte, A. Passi, D. Vigetti, S. Ricard-Blum, R. D. Sanderson, T. Neill, R. V. Iozzo, *Chem. Rev.* **2018**, *118*, 9152.
- [9] D. Shi, A. Sheng, L. Chi, *Front. Mol. Biosci.* **2021**, *8*, 639666. <https://doi.org/10.3389/fmolb.2021.639666>
- [10] T. N. Wight, *Matrix Biol.* **2018**, *71–72*, 396.
- [11] M. Marcisz, B. Huard, A. G. Lipska, S. A. Samsonov, *Glycobiology* **2021**, *31*, 772.
- [12] A. Moustakas, S. Souchelnytskyi, C. H. Heldin, *J. Cell Sci.* **2001**, *114*, 4359.
- [13] M. B. Huynh, M. O. Ouidja, S. Chantepie, G. Carpentier, A. Maïza, G. Zhang, J. Vilares, R. Raisman-Vozari, D. Papy-Garcia, *PLoS One* **2019**, *14*, e0209573.
- [14] S. Morla, *Int. J. Mol. Sci.* **2019**, *20*, 1963. <https://doi.org/10.3390/ijms20081963>
- [15] U. Uciechowska-Kaczmarzyk, S. Babik, F. Zsila, K. K. Bojarski, T. Beke-Somfai, S. A. Samsonov, *J. Mol. Graphics Modell.* **2018**, *82*, 157.
- [16] S. Faham, R. E. Hileman, J. R. Fromm, R. J. Linhardt, D. C. Rees, *Science* **1996**, *80*, 1116.
- [17] A. D. DiGabriele, I. Lax, D. I. Chen, C. M. Svahn, M. Jaye, J. Schlessinger, W. A. Hendrickson, *Nature* **1998**, *393*(6687), 812.
- [18] R. Derler, B. Gesslbauer, C. Weber, E. Strutzmann, I. Miller, A. Kungl, *Int. J. Mol. Sci.* **2017**, *18*, 2605. <https://doi.org/10.3390/ijms18122605>
- [19] A. Penk, L. Baumann, D. Huster, S. A. Samsonov, *Glycobiology* **2019**, *29*, 715.
- [20] K. Nordsieck, L. Baumann, V. Hintze, M. T. Pisabarro, M. Schnabelrauch, A. G. Beck-Sickinger, S. A. Samsonov, *Biopolymers*, **2018**, *109*, e23103.
- [21] M. Petitou, B. Casu, U. Lindahl, *Biochimie* **2003**, *85*, 83.
- [22] K. M. Sepuru, B. Nagarajan, U. R. Desai, K. Rajarathnam, *J. Biol. Chem.* **2018**, *293*, 17817.
- [23] A. Imberty, H. Lortat-Jacob, S. Pérez, *Carbohydr. Res.* **2007**, *342*(3), 430.
- [24] B. Nagarajan, S. G. Holmes, N. V. Sankaranarayanan, U. R. Desai, *Curr. Opin. Struct. Biol.* **2022**, *74*, 102356.
- [25] U. Uciechowska-Kaczmarzyk, I. de Beauchene, S. A. Samsonov, *J. Mol. Graphics Modell.* **2019**, *90*, 42.
- [26] M. Frank, *Carbohydrates Drugs* **2014**, *12*(June), 53.
- [27] G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, A. J. Olson, *J. Comput. Chem.* **1998**, *19*, 1639.
- [28] S. A. Samsonov, J. P. Gehrcke, M. T. Pisabarro, *J. Chem. Inf. Model.* **2014**, *54*, 582.
- [29] T. Siebenmorgen, M. Engelhard, M. Zacharias, *J. Comput. Chem.* **2020**, *41*, 1436.
- [30] T. Siebenmorgen, M. Zacharias, *J. Chem. Inf. Model.* **2020**, *60*, 5552.
- [31] M. Maszota-Zieleniak, M. Marcisz, M. M. Kogut, T. Siebenmorgen, M. Zacharias, S. A. Samsonov, *J. Comput. Chem.* **2021**, *42*, 1040.
- [32] M. Marcisz, M. Maszota-Zieleniak, B. Huard, S. A. Samsonov, *Biomolecules* **2021**, *11*, 1349. <https://doi.org/10.3390/biom11091349>
- [33] S. A. Samsonov, J. Teyra, M. T. Pisabarro, *J. Comput.-Aided Mol. Des.* **2011**, *25*(5), 477.
- [34] A. J. McCoy, X. Y. Pei, R. Skinner, J. P. Abrahams, R. W. Carrell, *J. Mol. Biol.* **2003**, *326*, 823.
- [35] H. Nguyen, J. Maier, H. Huang, V. Perrone, C. Simmerling, *J. Am. Chem. Soc.* **2014**, *136*, 13959.
- [36] C. J. M. Huige, C. Altona, *J. Comput. Chem.* **1995**, *16*, 56.
- [37] K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. González-Outeiriño, C. R. Daniels, B. L. Foley, R. J. Woods, *J. Comput. Chem.* **2008**, *29*, 622.
- [38] D. A. Case, K. Belfon, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham, III, V. W. D. Cruzeiro, T. A. Darden, R. E. Duke, G. Giambasu, M. K. Gilson, H. Gohlke, A. W. Goetz, R. Harris, S. Izadi, S. A. Izmailov, K. Kasavajhala, A. Kovalenko, R. Krasny, T. Kurtzman, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, V. Man, K. M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, A. Onufriev, F. Pan, S. Pantano, R. Qi, D. R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. L. Simmerling, N. R. Skrynnikov, J. Smith, J. Swails, R. C. Walker, J. Wang, L. Wilson, R. M. Wolf, X. Wu, Y. Xiong, Y. Xue, D. M. York, P. A. Kollman. AMBER 2020, University of California, San Francisco, **2020**.
- [39] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, *Proteins: Struct., Funct., Genet.* **2006**, *65*, 712.
- [40] H. Nguyen, D. R. Roe, C. Simmerling, *J. Chem. Theory Comput.* **2013**, *9*, 2020.
- [41] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, *J. Chem. Phys.* **1998**, *79*, 926.
- [42] P. Li, L. F. Song, K. M. Merz Jr., *J. Chem. Theory Comput.* **2015**, *11*(4), 1645.
- [43] B. M. Sattelle, J. Shakeri, A. Almond, *Biomacromolecules* **2013**, *14*(4), 1149.
- [44] K. K. Bojarski, A. K. Sieradzan, S. A. Samsonov, *Biopolymers* **2019**, *110*(7), e23252.
- [45] A. Onufriev, D. A. Case, D. Bashford, *J. Comput. Chem.* **2002**, *23*, 1297.
- [46] S. A. Samsonov, M. T. Pisabarro, *Glycobiology* **2016**, *26*, 850.
- [47] M. M. Kogut, M. Marcisz, S. A. Samsonov, *Curr. Opin. Struct. Biol.* **2022**, *73*, 102332.
- [48] N. V. Sankaranarayanan, B. Nagarajan, U. R. Desai, *Int. J. Mol. Sci.* **2021**, *22*, 7542. <https://doi.org/10.3390/ijms22147542>
- [49] J. Teyra, S. A. Samsonov, S. Schreiber, M. T. Pisabarro, *BMC Bioinform.* **2011**, *12*, 398.
- [50] Q. Liu, J. W. Brady, *J. Am. Chem. Soc.* **1996**, *118*(46), 12276.
- [51] S. E. Pagnotta, S. E. McLain, A. K. Soper, F. Bruni, M. A. Ricci, *J. Phys. Chem. B.* **2010**, *114*(14), 4904.
- [52] L. Pol-Fachin, H. Verlin, *Carbohydr. Res.* **2008**, *343*(9), 1435.
- [53] B. C. Roberts, R. L. Mancera, *J. Chem. Inf. Model.* **2008**, *48*(2), 397.
- [54] R. Thilagavathi, R. L. Mancera, *J. Chem. Inf. Model.* **2010**, *50*(3), 415.
- [55] A. D. van Dijk, A. M. Bonvin, *Bioinformatics* **2006**, *22*(19), 2340.
- [56] R. Baron, P. Setny, J. A. McCammon, *J. Am. Chem. Soc.* **2010**, *132*(34), 12091.
- [57] S. A. Samsonov, J. Teyra, M. T. Pisabarro, *Proteins* **2008**, *73*(2), 515.
- [58] S. A. Samsonov, *Methods Mol. Biol.* **2018**, *1762*, 445.
- [59] A. Almond, J. K. Sheehan, A. Brass, *Glycobiology* **1997**, *7*(5), 597.
- [60] A. Almond, J. K. Sheehan, *Glycobiology* **2000**, *10*(3), 329.
- [61] M. Jana, S. Bandyopadhyay, *Phys. Chem. Chem. Phys.* **2012**, *14*(18), 6628.
- [62] A. Nurisso, B. Blanchard, A. Audfray, L. Rydner, S. Oscarson, A. Varrot, A. Imberty, *J. Biol. Chem.* **2010**, *285*(26), 20316.

- [63] K. M. Sepuru, B. Nagarajan, U. R. Desai, K. Rajarathnam, *J. Biol. Chem.* **2016**, 291(39), 20539.
- [64] D. Ganesh, P. Jain, C. D. Shanthamurthy, S. Toraskar, R. Kikkeri, *ACS Chem. Biol.* **2021**, 16(11), 2481.
- [65] N. Sapay, E. Cabannes, M. Petitou, A. Imbert, *Glycobiology* **2011**, 21(9), 1181.
- [66] B. Nagarajan, N. V. Sankaranarayanan, B. B. Patel, U. R. Desai, *PLoS One* **2017**, 12(2), e0171619.
- [67] A. Singh, W. C. Kett, I. C. Severin, I. Agyekum, J. Duan, I. J. Amster, A. E. I. Proudfoot, D. R. Coombe, R. J. Woods, *J. Biol. Chem.* **2015**, 290(25), 15421.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** M. Marcisz, M. Gaardl s, K. K. Bojarski, T. Siebenmorgen, M. Zacharias, S. A. Samsonov, *J. Comput. Chem.* **2022**, 43(24), 1633. <https://doi.org/10.1002/jcc.26965>

## Publication D3

Modeling Protein-Glycosaminoglycan  
Complexes: Does the Size Matter?

# Modeling Protein–Glycosaminoglycan Complexes: Does the Size Matter?

Mateusz Marcisz, Martin Zacharias, and Sergey A. Samsonov\*



Cite This: *J. Chem. Inf. Model.* 2021, 61, 4475–4485



Read Online

ACCESS |



Metrics & More

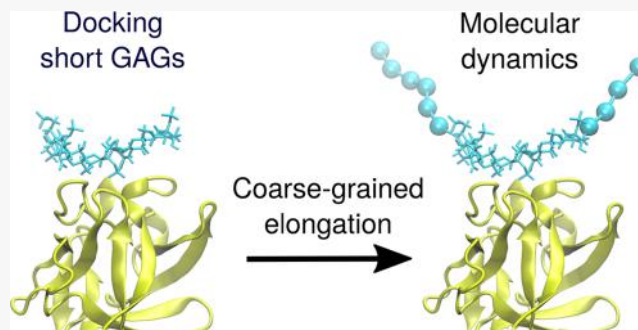


Article Recommendations



Supporting Information

**ABSTRACT:** Docking glycosaminoglycans (GAGs) has been challenging because of the complex nature of these long periodic linear and negatively charged polysaccharides. Although standard docking tools like Autodock3 are successful when docking GAGs up to hexameric length, they experience challenges to properly dock longer GAGs. Similar limitations concern other docking approaches typically developed for docking ligands of limited size to proteins. At the same time, most of more advanced docking approaches are challenging for a user who is inexperienced with complex *in silico* methodologies. In this work, we evaluate the binding energies of complexes with different lengths of GAGs using all-atom molecular dynamics simulations. Based on this analysis, we propose a new docking protocol for long GAGs that consists of conventional docking of short GAGs and further elongation with the use of a coarse-grained representation of the GAG parts not being in direct contact with its protein receptor. This method automated by a simple script is straightforward to use within the Autodock3 framework but also useful in combination with other standard docking tools. We believe that this method with some minor case-specific modifications could also be used for docking other linear charged polymers.



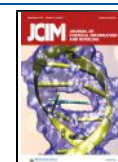
## INTRODUCTION

Human cells express multiple polymers that display a variety of functions. One particular class of those polymers are glycosaminoglycans (GAGs). They are long periodic linear and negatively charged polysaccharides that by interacting with proteins play an immense role in the extracellular matrix processes. Depending on their sulfation pattern and charge densities, GAGs manifest different conformational and binding properties.<sup>1</sup> GAGs are built of repeating disaccharide units. Each of them consists of an amino sugar and an uronic acid or galactose.<sup>2</sup> Depending on the sulfation pattern and monosaccharide composition, GAG disaccharide units can display 408 variants,<sup>3</sup> of which 202 can be found in mammals.<sup>4,5</sup> While some of the protein–GAG interactions are specific, most of them are considered as nonspecific and/or electrostatically driven due to the high negative charge of those polysaccharides directly correlating with the binding affinities.<sup>6</sup> Among many proteins, there are two major protein groups that GAGs can interact with. One of them are growth factors,<sup>7,8</sup> and the second group are chemokines.<sup>9–11</sup> In the case of growth factors, GAGs are able to influence the cell signaling and the activity of the proteins by changing their conformation or by oligomerization facilitation of their receptors by binding and clustering multiple fibroblast growth factors (FGFs) at the same time.<sup>12,13</sup> For example, in the case of vascular endothelial growth factor (VEGF), a key player in cancer, arthritis, angiogenesis, and regenerative processes,<sup>14</sup> global conforma-

tional changes induced by heparin (HP) binding influence its binding capability to its receptor on the cell membrane.<sup>7</sup> HP and heparan sulfate are also able to bind to transforming growth factor  $\beta$  (TGF- $\beta$ 1),<sup>15,16</sup> a protein that is responsible for the regulation of the proliferation, adhesion, differentiation, and cell migration.<sup>17</sup> Depending on the sulfation pattern, hyaluronan derivatives influence TGF- $\beta$ 1 activity and its binding to its receptor.<sup>18,19</sup> The second mentioned group of the protein that interacts with GAGs are chemokines.<sup>10,20</sup> This is mostly a proinflammatory group of proteins that belongs to cytokines. They may influence cells in different manners: some of the chemokines can alter metastasis tumor growth and angiogenesis by either promoting or inhibiting it.<sup>21</sup> GAGs by interacting with IL-8 can alter the ability to activate leukocytes.<sup>22–24</sup> GAGs can also affect pro-/anti-inflammatory functions of IL-10.<sup>25,26</sup> It was also shown that HP may interact with CXCL-14,<sup>11</sup> and by doing that, it increases migratory potential on monocytic THP-1 cells.<sup>27</sup> Many computational studies on GAGs show their promising potential in the

Received: June 14, 2021

Published: September 8, 2021



examination of the protein–GAG interactions. The following studies successfully investigated effects of the GAGs binding on a variety of different proteins, such as CXCL-14,<sup>11</sup> VEGF,<sup>7</sup> CXCL-8,<sup>9,24,28</sup> a Proliferation Inducing Ligand (APRIL),<sup>29</sup> IL-10,<sup>25,26,30</sup> CXCL-12,<sup>31</sup> acidic fibroblast growth factor (FGF-1),<sup>32</sup> or protein–ion–GAG complexes.<sup>33</sup>

Even though computational studies seem to be very successful and helpful in protein–GAG investigations, there are still a lot of challenges that have not been fully overcome yet. One of them is docking long GAG molecules. Usually, GAGs dp4 or 6 (dp stands for degree of polymerization) are used in molecular docking. This is caused by the fact that most of the docking software can handle only a limited number of torsional degrees of freedom for the docked molecules. The number of torsional degrees of freedom is often given by the number of rotatable bonds in the ligand. For example, when using Autodock3, which is the most accurate docking tool for GAG docking,<sup>34</sup> to dock GAGs of a higher degree of polymerization, a user cannot include all torsional degrees of freedom and needs to manually pick the most relevant ones not to overcome the limit of 33. The more the torsional degrees of freedom are active in a docking procedure, the more accurate docking results should theoretically be possible to obtain. Thus, using very long GAG molecules (e.g., dp10 or higher) heavily hampers docking performance and makes it unfeasible and/or unreliable. However, there are some ways to overcome this issue. In a fragment-based approach, trimeric GAG fragments are docked on the protein's surface, and afterward, they are assembled based on structural overlaps.<sup>35</sup> While this method is of great benefit for a number of protein–GAG complexes, it has some flaws, e.g., when GAG is located in a way that some of the oligosaccharide units are in close proximity to the negatively charged amino acid residues (contributing to unfavorable interactions), this method may fail to dock trimeric fragments nearby such residues and thus fail to produce properly docked longer GAG fragments. Perhaps the best method to dock long GAG molecules so far is replica exchange with repulsive scaling method.<sup>36,37</sup> This method is rather independent of the length of the GAG both in terms of docking predicting power and computational resources requirement (although, this method could demand heavy computational resources—no matter how long the GAG is—depending on the protein size in the complex). This method, while being promising for GAG docking in the vast majority of cases, may experience difficulties to dock GAG molecule into an enzymatic pocket of the protein.<sup>37</sup> One more argument in disfavor of the above-mentioned specific GAG docking approaches is the fact that they bring in some considerable complexity compared to standard docking methods and may be complicated to handle especially for nonexperts in the molecular modeling and researchers not familiar with the mentioned technical solutions.

Given all that, we aimed to propose a straightforward approach to dock longer GAG molecules without creating unnecessary technical complications while maintaining docking quality. The approach is based on four simple steps: (1) to dock a short (hexameric) GAG; (2) to add more GAG units in the coarse-grained (CG) representation to the previously docked ones manually, e.g., using programs that prepare molecular dynamics (MD) input files like LEaP program from the AMBER suite; (3) to run a molecular dynamics simulation to find an ensemble of GAG conformations for the whole GAG molecule; and (4) to calculate binding free energy. Combining

molecular docking with molecular dynamics approaches to predict a complex structure between a receptor and a ligand was previously shown to be a more powerful approach than the usage of the molecular docking alone for other molecular systems.<sup>38,39</sup> Moreover, the application of molecular dynamics approach allows for the scoring of docking poses with the use of more accurate free energy calculation schemes than it is usually done within molecular docking software and that, in addition, takes into account movements in the molecular system (this aspect is partially or completely neglected in classical docking scoring schemes). In particular, molecular mechanics/Poisson–Boltzmann surface area (MM/PBSA) and its approximation molecular mechanics/generalized Born surface area (MM/GBSA), both based on the use of the implicit solvent model,<sup>40</sup> showed previously to be able to rank experimental binding poses<sup>41</sup> and the modeled binding poses<sup>22,42</sup> for a number of protein–GAG systems in accordance with the experimental data. Apart from this, the per residue free energy decomposition scheme implemented within these methods allows us to dissect individual free energy contributions of the particular residues to the binding affinity allowed and properly rank the effects of point mutations on the binding affinity in the protein–GAG systems.<sup>43,44</sup> Also, recently, it was shown that the MM/GBSA scoring could be useful in distinguishing a native binding pose from other ones for this type of complexes.<sup>37</sup>

Therefore, our method combining molecular docking, molecular dynamics, and molecular dynamics-based free energy calculation schemes is expected to be more effective than classical docking approaches because of its conceptual superiority, in particular when applied to a GAG ligand that represents numerous challenges for conventional docking protocols.

The study consists of several parts. First, MM/PBSA and MM/GBSA methods to calculate binding free energies are applied to a dataset of protein–GAG experimental structures. The results for all-atom (AA) and coarse-grained (CG) GAGs modeled using previously obtained CG parameters that describe several GAG chemical moieties as different beads<sup>45</sup> are compared, and the general applicability of these free energy calculation approaches for a CG GAG model is justified. Furthermore, short GAGs from the X-ray structures available for two proteins and GAG docked poses obtained with three peptide receptors are elongated and simulated using a conventional AA approach and the corresponding binding energies are calculated. Then, a new, essentially more simplified, CG model of GAG is introduced. In this model, each GAG monosaccharide unit is represented just by a single pseudoatom. These pseudoatoms are used to substitute the parts of the GAG that are not in contact with the protein/peptide receptor based on the AA simulations. These systems with CG parts are simulated, and the differences between the obtained free binding energies in AA and CG simulations are discussed. Finally, we aimed to propose a model that allows us to calculate free binding energy of a GAG of a given length without simulating the GAG containing an elongated part explicitly using Coulomb and Hückel models of electrostatics. We also attempted to approach the interactions of these GAGs with the protein using only one CG bead to model the elongated part.

We believe that the method for modeling protein complexes with long GAGs proposed in the study with the introduction of some minor changes should also be applicable to most other

charged linear polysaccharides or biopolymers like, for example, nucleic acids.

## MATERIALS AND METHODS

**Comparing the Performances of MM/PBSA and MM/GBSA Free Energy Decomposition Calculations for GAG Ligands in AA and CG Representations Complexed with Proteins.** Short (10 ns) MD simulations (see the protocols in the *Molecular Dynamics* section) were performed for a dataset of nine protein–GAG X-ray structures obtained from the PDB with the following PDB IDs: 1GMN (receptor: NK1; ligand: HP dp5), 1HM2 (receptor: chondroitinase AC lyase; ligand: dermatan sulfate dp4), 1LOH (receptor: hyaluronate lyase; ligand: hyaluronic acid dp6), 1OFM (receptor: chondroitinase B; ligand: chondroitin sulfate-4 dp4), 2D8L (receptor: rhamnogalacturonyl hydrolase; ligand: desulfated chondroitin sulfate dp2), 2NWG (receptor: CXCL-12; ligand: two HP dp2 bound to two different binding sites), 3ANK (receptor: glucuronyl hydrolase mutant D175N; ligand: chondroitin sulfate-6 dp2), 3OGX (receptor: peptidoglycan recognition protein; ligand: HP dp2), 3OJV (receptor: FGF-1 in complex with the ectodomain of FGFR1c; ligand: HP dp6). The dataset included both enzymatic and nonenzymatic proteins previously shown to be characterized by significantly different binding properties<sup>46</sup> and GAGs of different types and lengths. Two series of the simulations were performed: in the first one, GAGs were described by all-atom model (AA), while in the second one, GAGs were simulated using the coarse-grained representation with the parameters obtained previously (CG).<sup>45</sup> In this model, specific GAG chemical groups were represented by pseudoatoms, spherical particles described by an integer charge corresponding to the charge of the respective chemical groups and Lennard-Jones parameters. In brief, in this CG representation constructed to be compatible with the AMBER package,<sup>47</sup> several pseudoatom types were selected to model the pyranose sugar ring (without hydroxyl group substitutes), *N*-acetyl, sulfate, and carboxyl groups, as well as glycosidic oxygen atoms. The bonded parameters (bonds, angles, dihedral angles) were obtained by the Boltzmann inversion approach from the corresponding AA simulations: the distributions of the parameters corresponding to the atomic groups defining pseudoatoms were analyzed, and the corresponding force field parameters fitting the distributions were extracted to define the new atomic types using the AMBER formalism. The charges were assigned empirically, while the Lennard-Jones potential parameters for pseudoatoms were calculated using the potential of mean force approach. Molecular mechanics/Poisson–Boltzmann surface area (MM/PBSA) calculations with default parameters for the whole trajectories of the binding free energies as well as per residue decomposition analysis was performed for the whole obtained trajectories.

Furthermore, the dynamic molecular docking approach (DMD)<sup>48</sup> was applied to the structures obtained from the PDB with the following PDB IDs: 1BFB (receptor: FGF-1; ligand: HP dp4), 1BFC (receptor: FGF-1; ligand: HP dp6), 2NWG (receptor: CXCL-12; ligand: HP dp2), 3C9E (receptor: cathepsin L; ligand: chondroitin sulfate-4 dp6), 2JQC (receptor: CD44; ligand: hyaluronic acid dp7). In these simulations, the GAG molecules were treated as CG, and the obtained results were compared with the AA DMD results for the same protein–GAG complexes from the original DMD work.<sup>48</sup> In brief, the DMD approach uses targeted molecular

dynamics protocol to dock a GAG ligand to a protein receptor by applying an additional potential to move a ligand from a distant starting position (beyond the cutoff of nonbonded interactions) to the predefined binding site on the receptor surface. DMD performance was compared for AA and CG ligand models of GAGs. The details for the applied protocols can be found in the original DMD work. The following parameters were included for this comparative analysis: RMSatd<sub>top</sub>: structural difference between the best scored docked structure and the corresponding experimental structure; RMSatd<sub>best</sub>: structural difference between the docked structure, which is the most similar structure to the corresponding experimental structure and the corresponding experimental structure; Rank<sub>best</sub>: rank of the docked structure, which is the most similar structure to the corresponding experimental structure; RMSatd: mean structural difference between all docked structures and the corresponding experimental structure; RMSatd<sub>top cluster</sub>: mean structural difference between all docked structures from the cluster of solutions with the highest scores and the corresponding experimental structure;  $r(\Delta G_{\text{total}} \sim \text{RMSatd})$ : Pearson correlation coefficient for total free binding energy and RMSatd of all docked structures;  $r(\Delta G_{\text{elect}} \sim \text{RMSatd})$ : Pearson correlation coefficient for *in vacuo* electrostatic free binding energy component and RMSatd of all docked structures; number of correctly predicted residues; number of correctly charged predicted residues; and number of correctly predicted uncharged polar residues were referenced to the 10 protein residues with the highest impacts on binding according to the per residue decomposition for the corresponding X-ray structures.

### Structures Used in the GAG Elongation Analysis.

**Protein Structures.** The following X-ray experimental structures from PDB was used in this work: 1AMX, 2AXM (FGF-1 with HP dp4 and dp6, respectively, monomeric form was used; dp stands for degree of polymerization),<sup>49</sup> 1BFB, 1BFC (FGF-2 with HP dp4 and dp6, respectively).<sup>13</sup>

**Peptide Structures.** The structure of the N-terminal fragment of the APRIL protein (ALA-VAL-LEU-THR-GLN-LYS-GLN-LYS-LYS-GLN) was adopted from Marcisz et al.<sup>29</sup> The structures of both peptides GLY-LYS-GLY-LYS-GLY and LYS-GLY-GLY-GLY-LYS (called InLYS and OutLYS, respectively) were constructed using xleap tool from AMBER suite.<sup>47</sup> Afterward, in the case of both peptides, 100 ns MD runs (described in the *Molecular Dynamics* section) were performed in AMBER to obtain most probable peptide conformations. The APRIL-derived peptide was chosen to represent a naturally existing GAG binding epitope, while InLYS and OutLYS, peptides were artificially constructed as short positively charged model peptides with the difference in the sequential and spatial distance between the GAG binding positively charged LYS side chains.

**GAG Structures.** All of the full-atom GAG structures—HP dp4 and dp6, dp10, dp16—were constructed from the building blocks of the sulfated GAG monomeric units' libraries<sup>22</sup> compatible with AMBER16 package. <sup>47</sup>GLYCAM06 force field<sup>50</sup> and literature data<sup>51</sup> were the sources of GAGs' charges.

**Molecular Docking.** Since there are no available experimental structures of the peptides with HP, for all three peptides, Autodock3<sup>52</sup> was used as it was previously described to yield the best results for protein–GAG complexes.<sup>34,41</sup> Entire peptides were covered using maximum gridbox size (126 Å × 126 Å × 126 Å) with a 0.375 Å grid step. The size of

**Table 1. MM/PBSA Free Binding Energy Analysis for Protein–GAG Complexes: Comparison of AA and CG GAG Representations<sup>a</sup>**

PDB ID	AA GAG model			CG GAG model		
	$\Delta G_{\text{elect}}$ (kcal mol <sup>-1</sup> )	$\Delta G_{\text{vdW}}$ (kcal mol <sup>-1</sup> )	$\Delta G_{\text{total}}$ (kcal mol <sup>-1</sup> )	$\Delta G_{\text{elect}}$ (kcal mol <sup>-1</sup> )	$\Delta G_{\text{vdW}}$ (kcal mol <sup>-1</sup> )	$\Delta G_{\text{total}}$ (kcal mol <sup>-1</sup> )
1GMN	-3354.6 ± 80.1	-42.2 ± 4.8	-92.6 ± 7.8	-3625.8 ± 84.3	-53.2 ± 4.7	-98.8 ± 9.0
1HM2	-458.6 ± 46.3	-47.2 ± 6.5	-22.4 ± 10.5	-539.4 ± 9.8	-61.7 ± 9.8	-80.6 ± 14.8
1LOH	-42.5 ± 34.1	-76.5 ± 6.6	-55.6 ± 11.7	-103.3 ± 6.3	-31.8 ± 34.4	-109.8 ± 9.7
1OFM	-746.5 ± 52.8	-27.7 ± 3.8	-42.1 ± 9.9	-767.2 ± 47.9	-27.2 ± 6.2	-50.9 ± 12.3
2D8L	-30.7 ± 21.2	-25.3 ± 3.9	-5.5 ± 9.9	-44.9 ± 35.7	-35.1 ± 5.3	-40.2 ± 10.7
2NWG	-1737.9 ± 102.4	-22.4 ± 5.2	-55.5 ± 18.5	-2334.1 ± 126.6	-33.5 ± 8.1	-94.4 ± 19.6
	-1096.7 ± 57.7	-21.5 ± 2.9	-25.1 ± 6.6	-1158.5 ± 106.6	-35.4 ± 7.3	-57.9 ± 12.8
3ANK	3.9 ± 45.1	-41.0 ± 4.5	-22.1 ± 7.0	-83.7 ± 48.4	-52.9 ± 6.5	-88.6 ± 16.9
3OGX	-1235.8 ± 35.7	-53.9 ± 4.3	-51.6 ± 8.7	-1351.7 ± 53.6	-54.3 ± 5.0	-57.3 ± 11.1
3OJV	-5701.5 ± 175.0	-86.0 ± 6.6	-194.9 ± 14.5	-5978.7 ± 148.1	-88.4 ± 6.2	-233.2 ± 15.6

<sup>a</sup> $\Delta G_{\text{elect}}$ ,  $\Delta G_{\text{vdW}}$ , and  $\Delta G_{\text{total}}$  are in vacuo electrostatic, van der Waals, and total MM/PBSA binding free energy values, respectively.

300 for the initial population and 10<sup>5</sup> generations for termination conditions were chosen. A total of 1000 independent runs with Lamarckian genetic algorithm was used, and 9995 × 10<sup>5</sup> energy evaluations were performed. DBSCAN algorithm<sup>53</sup> was used for clustering. RMSatd metric was used for clustering, which accounts for equivalence of the atoms of the same atomic type. This metric was reported to be more appropriate for GAG docking than classical root-mean-square deviation (RMSD) for periodic ligands.<sup>48</sup>

**Coarse-Grained Model Parameters for a Docked GAG Oligomer Elongation.** Obtained in this work, CG parameters compatible with AMBER format were obtained by the Boltzmann inversion approach and saved as the Parameter modification file (file.frcmod, see the Supporting Information). These parameters are described in the Results and Discussion section. These new parameters were obtained to be used for the MD simulations of the docked GAG in the AA representation that was further elongated by CG units. Each monomeric unit was represented by a single pseudoatom.

**Molecular Dynamics.** Experimental structures of protein–GAG, the docked structures of peptide–GAG complexes, and the corresponding structures with elongated GAGs were further analyzed by the MD approach. All of the MD simulations were performed using AMBER16 software package.<sup>47</sup> The ff14SB force field parameters were used for the protein and peptide molecules, while GLYCAM06j-1 parameters were used for GAGs. 8 Å water layer from solute to box's borders in shape of truncated octahedron was used to solvate complexes. Even in the case of HP dp16, this size of the layer was verified to be sufficient enabling the whole GAG molecule to always remain in the periodic box unit during the MD simulation. Na<sup>+</sup>/Cl<sup>-</sup> counterions were used to neutralize the net charge of the system. Preceding the production MD runs, energy minimization was made. A total of 500 steepest descent cycles and 10<sup>3</sup> conjugate gradient cycles with 100 kcal mol<sup>-1</sup> Å<sup>-2</sup> harmonic force restraint were performed. It continued with 3 × 10<sup>3</sup> steepest descent cycles and 3 × 10<sup>3</sup> conjugate gradient cycles without any restraints and followed by heating up the system to 300 K for 10 ps with harmonic force restraints of 100 kcal mol<sup>-1</sup> Å<sup>-2</sup> with the Langevin thermostat ( $\gamma = 5$  ps<sup>-1</sup>). Afterward, the system was equilibrated at 300 K and 10<sup>5</sup> Pa in isothermal isobaric ensemble for 500 ps with the Langevin thermostat ( $\gamma = 5$  ps<sup>-1</sup>) and Berendsen barostat (tau<sub>p</sub> = 1 ps). Then, the actual MD runs were carried out using the same isothermal isobaric ensemble for 100 ns. Particle mesh Ewald method for treating electrostatics and SHAKE

algorithm for all of the covalent bonds containing hydrogen atoms were implemented in the MD simulations. For both AA and CG simulations, the integration step of 2 fs was used.

Although we used short 10 ns MD simulations for a dataset of the experimental structures with short GAGs in the first part of our work (see the Comparing the Performances of MM/PBSA and MM/GBSA Free Energy Decomposition Calculations for GAG Ligands in AA and CG Representations Complexed with Proteins section), here we used 100 ns for all modeled complexes with elongated GAGs with the purpose of obtaining more proper sampling of the GAG conformational space when starting from a docked/modeled structures that cannot be verified by experimental data.

**Binding Free Energy Calculations.** For the free energy and per residue energy decomposition calculations, MM/GBSA (molecular mechanics generalized Born surface area) model igb = 2<sup>54</sup> from AMBER16 was used with default parameters on the whole trajectories (100 ns) obtained from MD simulations. Linear interaction energy (LIE) analysis was performed with a dielectric constant of 80 and noncalibrated weights (both  $\alpha$  and  $\beta$  were set to 1), performed by CPPTRAJ scripts on the same frames as the MM/GBSA.

## RESULTS AND DISCUSSION

**MM/PBSA Calculations for Protein–GAG Complexes: AA vs CG Representation of a GAG.** Prior to analyzing the elongated AA-GAG ligands bound to the proteins with the CG part, which represents the focus of this study, we performed MM/PBSA calculations of the binding free energies for nine nonredundant representative protein–GAG complexes where the full GAGs are modeled by with the AA and CG approaches. The aim of these calculations was to find out if the MM/PBSA method yields the results for a system containing a CG part that are in agreement with the data obtained for a conventional AA system. The CG parameters used to obtain the data provided in this subsection were described in detail in the work of Samsonov et al.<sup>45</sup> The data are summarized in Table 1. Pearson and Spearman correlations for  $\Delta G_{\text{elect}}$ ,  $\Delta G_{\text{vdW}}$ , and  $\Delta G_{\text{total}}$  are 0.997, 0.645, and 0.920; and 0.988, 0.503, and 0.758, respectively, suggesting that CG approximation, as it would be expected, affects van der Waals energy components but retains a very similar description of the systems in terms of the electrostatics. Since the electrostatic interactions are dominating in the protein–GAG systems, the total binding free energies were very similar as well. This suggests that the introduction of the CG part of a GAG that

Table 2. MM/GBSA Analysis of Binding HP of Different Lengths

	OutLys (kcal mol <sup>-1</sup> )	InLys (kcal mol <sup>-1</sup> )	APRIL peptide (kcal mol <sup>-1</sup> )	2AXM (kcal mol <sup>-1</sup> )	1BFC (kcal mol <sup>-1</sup> )
dp4	-24.2	-23.4	-25.6	-71.9	-65.7
dp6	-31.2	-27.6/19.6 <sup>a</sup>	-27.1	-84.8	-112.1
dp10	-35.9	-33.6	-42.8	-91.3	-126.2
dp16	-39.2	-36.9	-51.4	-86.6	-144.7

<sup>a</sup>In the case of one MD simulation, dissociation was observed. The first value indicates energies w/o mention of MD run, and the second value indicates those when taking it into account.

only interacts with the protein receptor via electrostatic interactions could be properly described by the MM/PBSA or MM/GBSA calculations compatibly with similar calculations for AA GAG representation. However, this conclusion should be taken with care: even if the effects of van der Waals description inaccuracies originated from the CG model do not directly affect electrostatic component of binding, they affect the general flexibility of the bound molecule. CG GAGs were shown to be indeed in general less flexible than the AA ones in the original work on this CG model.<sup>45</sup> Therefore, the introduction of the CG description affects GAG conformational space and, as a consequence, the whole structural organization of the bound GAG. This, in turn, results in the indirect effect of the modified van der Waals interactions on the electrostatics of the system influencing the binding affinity.

The relative mean differences between AA and CG absolute energy values (normalized by the AA corresponding values) are 30, 5, and 18% for  $\Delta G_{\text{elect}}$ ,  $\Delta G_{\text{vdW}}$ , and  $\Delta G_{\text{total}}$  respectively (clear outlier 3ANK is excluded). For all components, the values obtained with CG approach are overestimated in comparison to the ones from the AA approach. Per-residue free energy decomposition also shows systematic agreement for the AA and CG approaches when analyzing the individual impacts of the protein residues (Table S1). At the same time, there are no correlations in the per residue values obtained for the GAG residues. Furthermore, we compared the performances of the DMD docking approach using both AA and CG GAG representations (Table S2). The results obtained for the CG GAG model are slightly worse but, in general, quite similar to the ones obtained for the AA GAG model in the original DMD study.<sup>45</sup> All of these analyses suggest that the CG description of a GAG molecule complexed with a protein is consistent with the AA representation in terms of application of the MM/PBSA. This served as a premise for our further step in this study: in particular, for the proposition of even a more simplistic CG model for a GAG part that does not establish direct contact with a protein receptor. In this model, the interactions between this CG part of a GAG and the protein could be described as purely electrostatics-driven.

**All-Atom Simulations.** To obtain the reference data for the CG model, development and testing AA MD simulations were performed. For this, the available experimental structures of FGF-1 (PDB ID: 1AXM, 2AMX) and FGF-2 (1BFB, 1BFC) with HP dp4 and dp6 were used. These complexes could be successfully obtained by many conventional docking programs including AD3 (RMSD ~2.5 to 3.5 Å for the best scored docked poses).<sup>34,41</sup> Since the experimental structures with the peptides are not available, HP dp4 and dp6 were docked to all of the peptides: N-terminal part of the APRIL protein, InLys, and OutLys (all targets described in the Materials and Methods section). It is important to mention that in this work, we did not aim to improve the docking quality for short GAGs but to estimate the effect of the GAG elongation and to

understand if this elongation could be described properly using a mixed AA/CG GAG model. AA representation of GAGs was used as a reference for our analysis.

Since MM/PBSA and MM/GBSA approaches yielded essential correlation in protein–GAG systems (see an example in Figure S1), we further used only the MM/GBSA approach for these longer simulations since this approach is significantly faster.

We clearly observe that longer GAGs bind stronger independently of the analyzed system and the type of the receptor (protein or peptide) (Table 2). This is an expected net effect of the electrostatic interactions that become stronger with the increase of the GAG negative charge upon its elongation. Since the net charge of a GAG binding site on the protein/peptide surface always corresponds to the extent of the positive electrostatic potential,<sup>41</sup> an elongation of any GAG ligand bound to any of its receptors would render the interactions stronger. Although the specific binding unit of GAGs is relatively short according to the available PDB structures of protein–GAG complexes,<sup>41</sup> natural GAGs in the extracellular matrix are very long, reaching molecular weights up to over 100 kDa,<sup>5</sup> rendering the energetic effect originating in a GAG long chain to be important to take into account when the corresponding modeling is performed. Except the 2AXM, the difference between dp6 and dp16 in terms of binding free energy was 20% or higher (on average 24%). One more highlight of this comparison is that the energy discrepancy between dp6 and dp10 was 2 times higher than that between dp10 and dp16 despite addition of more sugar ring units in the case of dp10 to dp16 elongation. A very large increase in terms of binding strength was observed upon the elongation from dp4 to dp6, indicating that experiments with dp4 GAGs may strongly underestimate the binding strength of longer GAGs. Taken into account how often dp4/dp6 GAGs are used as models in computational studies, it is worth checking and rethinking those standards prior to applying dp4-based protocols to any new system.

**CG Parameters Obtained from All-Atom MD Simulations.** The new parameters described below were obtained from the AA MD simulations to be used for the CG elongation of the docked GAG in the AA representation as described in the following subsections. This new model was particularly developed for the purpose of elongating those parts of bound GAG chains that do not establish direct contact with the protein these GAGs are interacting with, and, therefore, it is thought to account only for electrostatics. Containing a single new atomic type corresponding to a whole GAG monomeric unit, this model is conceptually different and much more simple than the old one.<sup>45</sup> It is completely nonspecific for any chemical modifications of GAG residues since it is constructed to account primarily for electrostatic interactions and could be used for all negatively charged monosaccharide residues allowing for a straightforward modification of the residue



point charge when needed. This is not the case for the old model that, on the contrary, was developed to consider specific electrostatic and van der Waals interactions for particular monosaccharide units. In terms of the required computational expenses, MD simulations with the new model would be faster if only a CG GAG would be simulated. However, in the presence of a protein, an AA-GAG part, and explicit water molecules, the benefit in terms of the computational time reduction is rather negligible.

The new Z1 atomic type constructed corresponds to a CG pseudoatom describing a complete residue unit (monosaccharide unit with the charge of  $-2$ ) and, therefore, glycosidic linkages are omitted between monosaccharide units in this CG model.

**Bonded Parameters.** Bonded parameters (bonds, angles, and dihedral angles) were obtained from the AA MD simulations. For the calculations of equilibrium values and harmonic constants for bonds, angles, and dihedral angles (Tables 3–5), the Boltzmann inversion approach was used.<sup>55</sup>

**Table 3. Z1 Pseudoatom Bond Parameters Compatible with the AMBER Package**

covalent bond parameters		
atoms RK	(kcal mol <sup>-1</sup> Å <sup>-2</sup> ) <sup>a</sup>	REQ (Å) <sup>b</sup>
Z1-Z1	120	5.2
Z1-Cg	120	5.2
Os-Z1	120	2.8

<sup>a</sup>Force constant. <sup>b</sup>Equilibrium bond length.

**Table 4. Z1 Pseudoatom Angle Parameters Compatible with the AMBER Package**

angle parameters		
atoms in the angle	TK (kcal mol <sup>-1</sup> rad <sup>-2</sup> ) <sup>a</sup>	TEQ (deg) <sup>b</sup>
Z1-Z1-Z1	100	160
Z1-Z1-Cg	100	160
Z1-Cg-H2	70	108.5
Z1-Cg-Cg	70	108.5
Z1-Cg-Os	60	110
Cg-Os-Z1	100	160
Os-Z1-Z1	100	160

<sup>a</sup>Force constant. <sup>b</sup>Equilibrium angle value.

In the case of dihedral angles (Table 5), periodicity was set to 1 or 3 depending on the number of maxima/minima of the potential per 360°, and the amplitude was obtained as the difference between the global minimum and the highest energetical barrier between global and local minima. In the case of artifacts observed during simulations, particular parameters were manually refined.

**Nonbonded Parameters (Charges, Lennard-Jones Parameters).** The charge of the pseudoatom of the monomeric unit of the HP was set accordingly to the number of sulfate and carboxyl groups, which is  $-1$  per group in the unit. In the case of Lennard-Jones parameters, the RvdW (van der Waals radius) and EDEP (energy well depth) values were empirically assigned to the doubled and equal values obtained for the internal pyranose ring in our previous CG model of GAGs, respectively (Table 6).<sup>45</sup>

**Mixed AA/CG Simulations: CG Elongation of a GAG.** To evaluate our CG model (Figure 1) of the HP, MD

**Table 5. Z1 Pseudoatom Dihedral Angle Parameters Compatible with the AMBER Package**

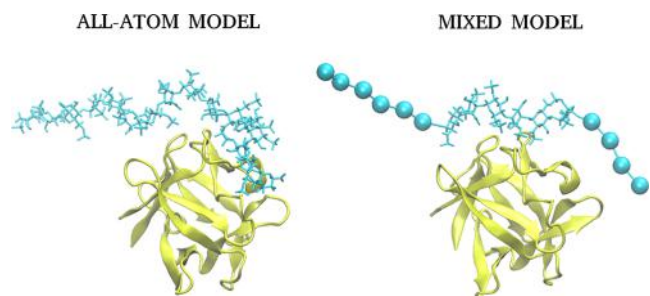
dihedral angle parameters				
atoms in the dihedral angle	IDIVF <sup>a</sup>	PK (kcal mol <sup>-1</sup> ) <sup>b</sup>	phase (deg) <sup>c</sup>	PN <sup>d</sup>
Z1-Z1-Z1-Z1	1	1	0	1
Z1-Z1-Z1-Cg	1	1	0	1
Z1-Z1-Cg-Cg	1	0.16	0	3
Z1-Cg-Cg-H1	1	0.16	0	3
Z1-Cg-Cg-H2	1	0.16	0	3
Z1-Z1-Cg-H2	1	0.16	0	3
Z1-Z1-Cg-Os	1	0.16	0	3
Z1-Cg-Cg-Ng	1	-1.3	0	1
Z1-Cg-Cg-Cg	1	-0.27	0	1
Z1-Cg-Os-Cg	1	-0.27	0	1
Cg-Cg-Os-Z1	1	0.16	0	3
Cg-Os-Z1-Z1	1	0.16	0	3
H1-Cg-Os-Z1	1	0.27	0	3
Z1-Cg-Cg-Os	1	0.16	0	3
Os-Z1-Z1-Z1	1	0.16	0	3

<sup>a</sup>Factor by which the torsional barrier is divided. <sup>b</sup>Barrier height divided by a factor of 2. <sup>c</sup>Phase shift angle in the torsional function. <sup>d</sup>Periodicity of the torsional barrier.

**Table 6. Z1 Pseudoatom Lennard-Jones Parameters Compatible with the AMBER Package**

basic information		Lennard-Jones parameters	
CG pseudoatom	mass (au)	RvdW <sup>a</sup> (Å)	EDEP <sup>b</sup> (kcal mol <sup>-1</sup> )
Z1	225	4	3.4

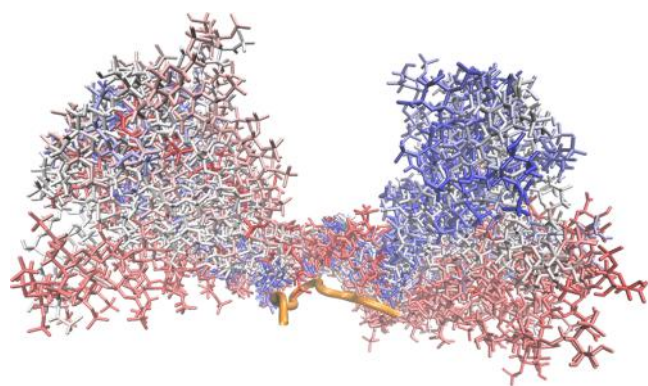
<sup>a</sup>van der Waals radius. <sup>b</sup>Energy well depth.



**Figure 1.** Graphical representation of all-atom (left) and mixed (right) model of dp16 heparin in complex with FGF-2. Protein is in cartoon representation (yellow); all-atom and CG GAGs are in licorice and van der Waals sphere representation, respectively (cyan).

simulations with CG atoms were performed and compared to all-atom MD simulations. In AA runs, we observed that the core of GAG—the part that is especially the closest to the binding side of the protein/peptide—is in the closest proximity of the protein and barely moved. In contrast, it is the lateral parts of the GAGs that tend to move freely (Figure 2). It suggests that interactions between those parts and the protein are even less specific and thus almost purely electrostatics-driven. Therefore, we believe that replacing lateral parts of the GAGs with CG model units should not substantially affect the nature of the interactions established between the analyzed molecules.

First, we compared the convergence of MD simulations for the AA and CG approaches in terms of the structural flexibility and energetics (Figures S2 and S3, respectively). In most of the



**Figure 2.** Graphical representation of the MD run of complex of APRIL peptide (orange cartoon) with HP dp16 (licorice). The color scheme from red to blue indicates heparin conformations ranging from the beginning to the end of the MD simulation.

cases, the convergence in terms of RMSD was observed already after 20 ns. Clearly, the flexibility of the AA GAGs is significantly higher than in the mixed AA/CG model. For MM/GBSA binding free energy, the converge is already reached after 10 ns of the simulation, and there are slightly higher variations of the energy observed for the AA simulation, while there are no differences in the time needed for the convergence. The trends of the convergence observed here should not be expected to be the same for other protein–GAG or peptide–GAG complexes. Indeed, in other systems, MD simulations may take longer or shorter to converge. Nevertheless, the goal of the MD simulations performed in this study is not to reach a convergence but to show that the transition from AA to CG representation of the GAG part does not substantially affect the results of the free energy calculations in the same system.

Starting positions of the molecules from all-atom simulations were taken. Original dp6 part of the HP was not modified, and only atoms that were manually added to build dp16 were replaced with CG pseudoatoms for HP rings. Additionally, the user can use the script ([Supporting Information](#)) for automatic addition of pseudoatoms. Then, MD simulations with a GAG represented as AA in the binding core and as CG in its lateral parts were performed, and the results obtained from MM/GBSA energy analysis from mixed model simulations of dp16 HP are listed in [Table 7](#). Average difference obtained from energy analysis of mixed CG/AA model compared to the AA model was 5.6%. Compared to the difference that is a consequence of using shorter GAGs, which is on average 24% (dp6) and 39% (dp4) underestimation of the value, it is a substantial improvement. In the case of the mixed model, most of the values were also underestimated (compared to the AA model): 7% for the N-terminal fragment of APRIL, 3% for the FGF-2 and OutLYS peptide, and 1% for the InLYS peptide. However, the binding free energy calculations showed 14%

overestimation in the case of the FGF-1/HP complex. Additional energy analysis was performed in the form of LIE calculations and is described in the Supporting Information ([Table S3](#)).

During MD runs of both AA and mixed AA/CG models, we observed similar motions of the GAGs molecules with respect to the protein/peptide, which suggests that the used CG model also properly reflects the dynamics of the system ([Figure 3](#)).

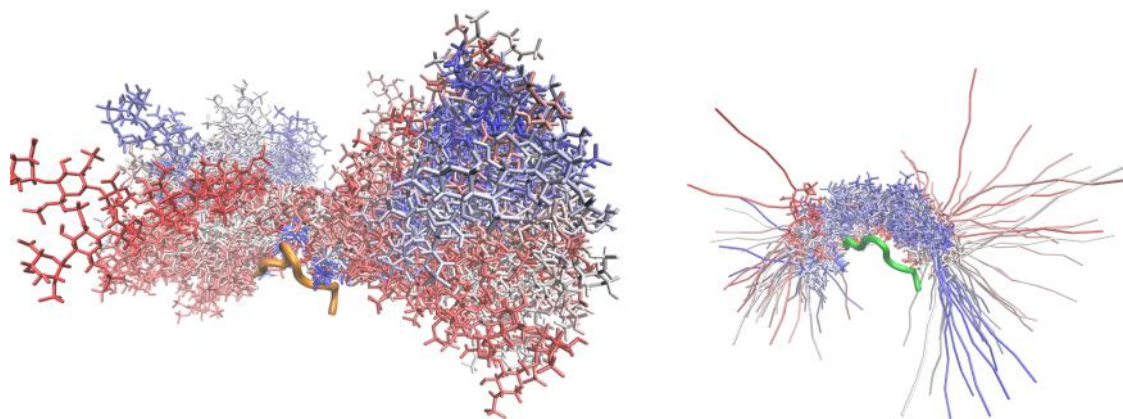
**Mixed All-Atom/Coarse-Grained Simulations Based on Per-Residue Energy Analysis.** The division of the modeled GAG chain into AA and CG parts for the further MD analysis could be done by analyzing the free energy properties of the binding poses instead of using visual inspection of AA MD followed by the manual selection of the residues to substitute. For this, we performed per-residue energy analysis of the complexes from AA MD simulations. This procedure allows us to define the particular contributions of the individual GAG units to binding a protein or a peptide. Then, only the residues with “weak” contributions to the binding energies were selected and further modeled by the CG approach. The threshold was set to  $-0.5 \text{ kcal mol}^{-1}$ , and any residue with energy value less favorable than this value was replaced. The idea behind such a procedure to substitute only the monosaccharide units with less substantial contributions in terms of binding energy is related to our goal to use the CG model for residues that are further away from the binding region and so less affecting the binding. Interestingly, the obtained error was higher (on average 10% of free energy difference compared to the AA simulation) when the residues were picked based on per-residue free energy decomposition than when the elongation was completed independently of such calculations ([Table 7](#)).

**Energy Prediction for GAG Elongation.** Furthermore, we aimed to extrapolate binding energies obtained from the analysis of the dp6 GAG to calculate them for the elongated GAG molecules without performing any further MD simulations. First, we proposed an equation based on Coulomb’s law to calculate the factor (depicted as  $W$  factor) that would allow us to obtain the binding energy of the complex containing GAGs of any length. Such an approach assumes that only electrostatic interactions are substantial for the added GAG part. We also proposed a script (see the [Supporting Information](#)) that would automatically calculate the binding energy of the elongated fragment of the GAG when given two files (pdb file of a bound GAG molecule and a receptor) and predefined  $W$  factor.

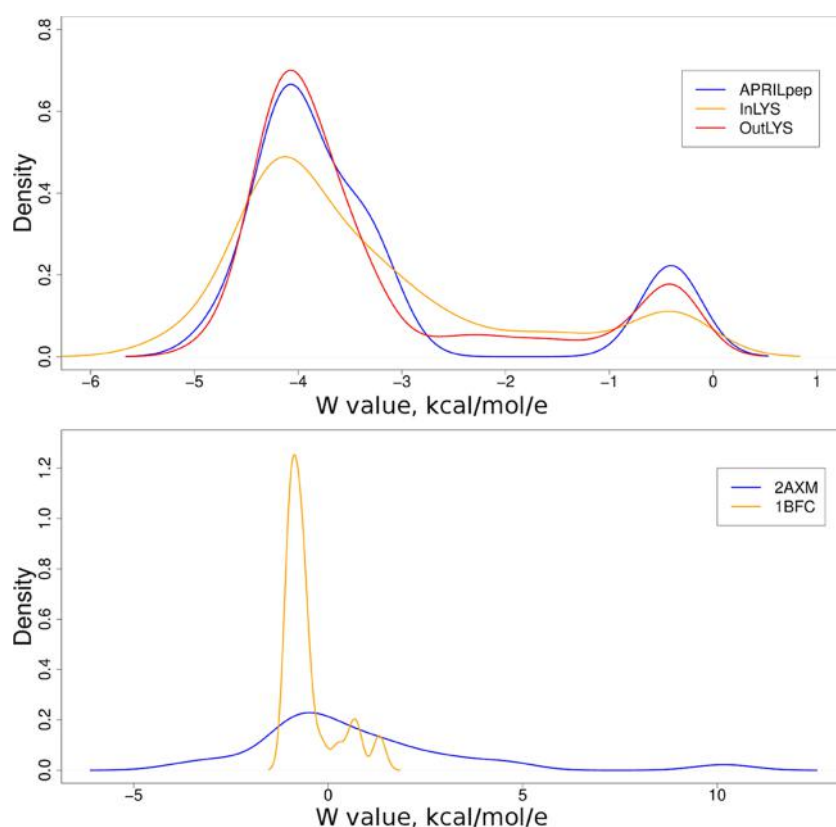
To calculate the  $W$  factor for the particular GAG residue, we use the following equation

**Table 7.** MM/GBSA Energy Analysis from Mixed Model Simulations of dp16 HP

model	description	OutLys (kcal mol <sup>-1</sup> )	InLys (kcal mol <sup>-1</sup> )	APRIL peptide (kcal mol <sup>-1</sup> )	2AXM (kcal mol <sup>-1</sup> )	1BFC (kcal mol <sup>-1</sup> )
AA	AA residues	-39.2	-36.9	-51.4	-86.6	-144.7
	elongated fragments of the GAG replaced with CG residues	-37.9	-36.8	-47.8	-98.3	-140.2
AA/CG	AA residues replaced with CG residues based on decomposed energy values	-46.4	-30.8	-51.0	-90.1	-130.5



**Figure 3.** Graphical representation of the MD run of complexes of APRIL peptide (cartoon) with all-atom (left, orange) and mixed model (right, green) HP dp16 (licorice). The color scheme from red to blue indicates heparin conformations ranging from the beginning to the end of the MD simulation.



**Figure 4.** Plot of  $W$  value probability densities calculated from MD runs (5 MD runs for each individual complex) for HP dp16 and short peptides (top) or proteins (bottom) used in this study.

$$W = \Delta G_{\text{res}} / \left( \sum_i (\text{positively charged residues}) - \sum_j (\text{negatively charged residues}) \right)$$

where  $W$  is the factor,  $\Delta G_{\text{res}}$  is the energy obtained from per-residue energy decomposition from MM/GBSA analysis, and  $\sum_{i/j}$  is the sum of reciprocities of the distances between GAG residues and all of the positively/negatively charged residues of the protein.

Each positive and negative residue is taken into account if it is within the cutoff of nonbonded interactions in the corresponding MD simulation. The  $W$  factor for the whole complex is the mean of the  $W$  factors for each of the GAG residues calculated from the simulations with HP dp16, and its usage for HP dp16 energy prediction would, therefore, yield the same energies as the ones obtained from the MD simulation.

The  $W$  factors and their distribution (Figure 4) for the peptide–GAG complexes were very similar for the peptides:  $-3.35$ ,  $-3.31$ , and  $-3.33$  kcal mol $^{-1}$  e $^{-1}$  for InLys, OutLys, and N-terminal fragment of the APRIL protein, respectively. In contrast, in the case of protein complexes, they differed

substantially in terms of mean of the  $W$  factors (0.65 and  $-0.50 \text{ kcal mol}^{-1} \text{ e}^{-1}$  for FGF-1 and FGF-2, respectively), and their distribution (Figure 4). It indicates that bigger and therefore more complex systems need an individual approach each time they are analyzed. However, in the case of simple and short systems (e.g., small peptide and GAG) individual approach is not necessary and the binding energy could be calculated directly using  $W$  factor of  $-3.33 \text{ kcal mol}^{-1} \text{ e}^{-1}$ . In this case, performing MD simulations and binding energy analysis for longer GAG variants is not needed.

Then, similarly to the previously described procedure, the Debye–Hückel equation ( $\Delta G \sim e^{-\kappa r}/r$ , where  $r$  is the distance and  $\kappa$  is the reversed Debye screening length) was used to calculate the  $W$  factor. In this approach, electrostatics screening in the electrolyte solution is taken into account. Physiological value of the ionic strength (0.15 M) was used in the calculations. The obtained data also suggested that  $W$  is very similar for all three peptides: 86.40, 89.13, and 85.50  $\text{kcal mol}^{-1} \text{ e}^{-1}$  for APRILpep, OutLYS, and InLYS, respectively. The calculated values for the protein–GAG systems were essentially different for the two systems and compared to the peptides:  $-0.83$  and  $20.00 \text{ kcal mol}^{-1} \text{ e}^{-1}$  for 2AXM and 1BFC, respectively.

Therefore, the energies could be, in principle, predicted for HP using a specific  $W$  factor for each system (in the case of three peptides,  $W$  factors are essentially the same), and such predictions applied for longer GAGs with this particular  $W$  factor would yield similar values to those in the MD simulations. However, for proteins, it is not possible to make such predictions *a priori* without performing MD simulations that are needed to define the  $W$  factor.

Based on these results, we believe that the difference in  $W$  profiles for two proteins obtained by calculations based on two dissimilar physics-based models is originated in the different charge distribution topology, protein surface geometry, and thus resulting electrostatic screening effects that do not allow us to find the same uniform factor for distinct protein receptors.

**Single Pseudoatom as an Extension of the GAG Molecule.** Furthermore, we aimed to design a model where only a single pseudoatom would function as an elongated lateral part of the bound GAG. Unfortunately, among the different parameters that were used, none yielded promising results in terms of reliably obtaining binding energies for the complexes compared to the ones from AA simulations, both when compared energies from MM/GBSA and LIE analysis (Table S3). Some artifacts were also observed when pseudoatom had a high negative charge ( $-5$  or lower) causing the interruption of the MD simulation. We believe that this approach does not have broad applicability. It is rather unlikely to propose parameters for a pseudoatom that would work consistently for the complexes with different electrostatic properties and geometry topologies. Additionally, one would need to propose a complete library of parameters for pseudoatoms distinct for every different length of an elongated GAG part that pseudoatom is replacing. The possible reason for this could be that an attempt to approximate an elongated molecule with a spherical particle could probably be physically inappropriate in terms of molecular symmetry.

## CONCLUSIONS

While docking long GAG molecules may require additional laborious technical work than docking shorter (dp4/6) GAG

oligomers, it is definitely worth the effort. In our approach, we use Autodock3 to find the best starting poses for the dp6 GAGs<sup>34,41</sup> that can be used for further GAG elongation. At the same time, it is important to mention that our approach is not limited to any special docking software. We expect that carbohydrate- and GAG-specific docking programs as Vina-Carb<sup>56</sup> or GlycoTorch Vina,<sup>57</sup> respectively, which also belong to the family of Autodock programs, would perform similarly or even outperform Autodock3 for obtaining the initial structures of protein/peptide complexes with short GAGs that are to be further elongated using the procedure proposed in this manuscript. In this procedure, we elongate a docked GAG using the CG model for the monosaccharide units and use it in conventional MD simulations. In this study, it was proven that elongating GAGs substantially increases the binding energy of the complex. While it is not a linear increase of binding strength, it is still substantial when dp16 is compared to dp4 or dp6. We consider that GAG elongation using a CG model for the monosaccharide units provides nearly equivalent outcome as the AA elongation, resulting only in a 5.6% difference in assessed binding energies, without introducing excessive technical complications. This suggests that a straightforward description of electrostatic interactions of the GAG parts not establishing direct contacts with their protein target is sufficient to describe the energetics of the system accurately enough. Binding energies obtained when using our script that elongates a GAG molecule (Supporting Information) and the CG model that are provided in this work are more accurate than using shorter GAGs with a standard AA approach. This method can be utilized by any user of AMBER and standard docking software like Autodock3 in a straightforward manner. It is a great advantage that with this approach, a user can specify the length of the extended lateral part of GAG to properly satisfy his needs. We also believe that this method with minor modifications could be implemented to other linear polysaccharides or negatively charged linear polymers like nucleic acids, in general.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jcim.1c00664>.

MM/PBSA free binding energy per residue decomposition analysis; DMD docking in protein–GAG systems: comparison of AA and CG GAG representations; LIE energy calculations; comparison of MM/GBSA and MM/PBSA binding free energies; RMSD of the bound HP dp16 (to investigated proteins/peptides) in AA and AA/CG representation; elongation script; script for automatic calculation of the energy of elongated GAG molecule; and frcmod file (PDF)

## AUTHOR INFORMATION

### Corresponding Author

Sergey A. Samsonov – Faculty of Chemistry, University of Gdańsk, 80-308 Gdańsk, Poland; [orcid.org/0000-0002-5166-4849](https://orcid.org/0000-0002-5166-4849); Email: [sergey.samsonov@ug.edu.pl](mailto:sergey.samsonov@ug.edu.pl)

### Authors

Mateusz Marcisz – Faculty of Chemistry, University of Gdańsk, 80-308 Gdańsk, Poland; Intercollegiate Faculty of Biotechnology of UG and MUG, 80-307 Gdańsk, Poland

Martin Zacharias – Center of Functional Protein Assemblies,  
Technical University of Munich, 85748 Garching, Germany

Complete contact information is available at:  
<https://pubs.acs.org/10.1021/acs.jcim.1c00664>

## Notes

The authors declare no competing financial interest. All of the PDB files were downloaded from the RCSB Protein Data Bank (<https://www.rcsb.org>). AMBER libraries for GAG residues are accessible on the website of GAG Computational Group under “Libraries” (<http://www.comp-gag.org/downloads>). The following free pieces of software were used: Autodock3 for molecular docking ([autodock.scripps.edu](http://autodock.scripps.edu)), Open Babel for formats conversion (<http://openbabel.org>), VMD for molecular visualization (<https://www.ks.uiuc.edu/Research/vmd>), and scripts for docking analysis (<https://gehrcke.de/code/>). All of these softwares are freely accessible on their websites. The differences between the input scripts used in molecular docking and the default values are listed on the website of GAG Computational Group under “Scripts/protocols” (<http://www.comp-gag.org/downloads>). All MD simulations were performed by the commercial AMBER16 molecular dynamics package (<https://ambermd.org>) purchased by the University of Gdańsk. The XLEAP and CPPTRAJ modules of AMBER16 were used for adding terminal residues to the polysaccharide chains and post-processing of trajectories, respectively. MM/GBSA free binding energy calculations were performed with the MM/GBSA module of AMBER16.

## ACKNOWLEDGMENTS

This research was funded by National Science Centre of Poland, grant number UMO-2018/30/E/ST4/00037 and Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—SFB 863—A10—111166240. The molecular dynamics simulations were performed on the PROMETHEUS cluster provided by Polish Grid Infrastructure (PLGRID, `gpuaprilgags`, `plgAprilgag2`) as well as on the local “piasek” cluster.

## REFERENCES

- (1) Habuchi, H.; Habuchi, O.; Kimata, K. Sulfation Pattern in Glycosaminoglycan: Does It Have a Code? *Glycoconjugate J.* **2004**, *21*, 47–52.
- (2) Varki, A.; Cummings, R. D.; Esko, J. D.; Stanley, P.; Hart, G. W.; Aebi, A.; Darvill, A. G.; Kinoshita, T.; Packer, N. H.; Prestegard, J. H.; Schnaar, R. L.; Seeberger, P. H. *Essentials of Glycobiology*, 3rd ed.; Cold Spring Harbor Laboratory Press, 2015.
- (3) Bu, C.; Jin, L. NMR Characterization of the Interactions Between Glycosaminoglycans and Proteins. *Front. Mol. Biosci.* **2021**, *8*, No. 165.
- (4) Clerc, O.; Mariethoz, J.; Rivet, A.; Lisacek, F.; Perez, S. D.; Ricard-Blum, S. A Pipeline to Translate Glycosaminoglycan Sequences into 3D Models. Application to the Exploration of Glycosaminoglycan Conformational Space. *Glycobiology* **2019**, *29*, 36–44.
- (5) Vallet, S. D.; Clerc, O.; Ricard-Blum, S. Glycosaminoglycan–Protein Interactions: The First Draft of the Glycosaminoglycan Interactome. *J. Histochem. Cytochem.* **2020**, *69*, 93–104.
- (6) Imberty, A.; Lortat-Jacob, H.; Pérez, S. Structural View of Glycosaminoglycan–Protein Interactions. *Carbohydr. Res.* **2007**, *342*, 430–439.
- (7) Uciechowska-Kaczmarzyk, U.; Babik, S.; Zsila, F.; Bojarski, K. K.; Beke-Somfai, T.; Samsonov, S. A. Molecular Dynamics-Based Model

of VEGF-A and Its Heparin Interactions. *J. Mol. Graphics Modell.* **2018**, *82*, 157–166.

(8) Bojarski, K. K.; Sieradzan, A. K.; Samsonov, S. A. Molecular Dynamics Insights into Protein-Glycosaminoglycan Systems from Microsecond-Scale Simulations. *Biopolymers* **2019**, *110*, No. e23252.

(9) Nordsieck, K.; Baumann, L.; Hintze, V.; Pisabarro, M. T.; Schnabelrauch, M.; Beck-Sickinge, A. G.; Samsonov, S. A. The Effect of Interleukin-8 Truncations on Its Interactions with Glycosaminoglycans. *Biopolymers* **2018**, *109*, No. e23103.

(10) Derler, R.; Gesslbauer, B.; Weber, C.; Strutzmann, E.; Miller, I.; Kungl, A. Glycosaminoglycan-Mediated Downstream Signaling of CXCL8 Binding to Endothelial Cells. *Int. J. Mol. Sci.* **2017**, *18*, No. 2605.

(11) Penk, A.; Baumann, L.; Huster, D.; Samsonov, S. A. NMR and Molecular Modeling Reveal Specificity of the Interactions between CXCL14 and Glycosaminoglycans. *Glycobiology* **2019**, *29*, 715–725.

(12) Mason, I. J. The Ins and Outs of Fibroblast Growth Factors. *Cell* **1994**, *78*, 547–552.

(13) Faham, S.; Hileman, R. E.; Fromm, J. R.; Linhardt, R. J.; Rees, D. C. Heparin Structure and Interactions with Basic Fibroblast Growth Factor. *Science* **1996**, *271*, 1116–1120.

(14) Risau, W. Mechanisms of Angiogenesis. *Nature* **1997**, *386*, 671–674.

(15) McCaffrey, T. A.; Falcone, D. J.; Du, B. Transforming Growth Factor- $\beta$ 1 Is a Heparin-binding Protein: Identification of Putative Heparin-binding Regions and Isolation of Heparins with Varying Affinity for TGF- $\beta$ 1. *J. Cell. Physiol.* **1992**, *152*, 430–440.

(16) Lyon, M.; Rushton, G.; Gallagher, J. T. The Interaction of the Transforming Growth Factor-Bs with Heparin/Heparan Sulfate Is Isoform-Specific. *J. Biol. Chem.* **1997**, *27*, 266–270.

(17) Moustakas, A.; Souchelnytskyi, S.; Heldin, C. H. Smad Regulation in TGF- $\beta$  Signal Transduction. *J. Cell Sci.* **2001**, *114*, 4359–4369.

(18) van der Smissen, A.; Samsonov, S.; Hintze, V.; Scharnweber, D.; Moeller, S.; Schnabelrauch, M.; Pisabarro, M. T.; Anderegg, U. Artificial Extracellular Matrix Composed of Collagen {I} and Highly Sulfated Hyaluronan Interferes with TGF $\beta$ (1) Signaling and Prevents TGF $\beta$ (1)-Induced Myofibroblast Differentiation. *Acta Biomater.* **2013**, *9*, 7775–7786.

(19) Koehler, L.; Samsonov, S.; Rother, S.; Vogel, S.; Köhling, S.; Moeller, S.; Schnabelrauch, M.; Rademann, J.; Hempel, U.; Teresa Pisabarro, M.; Scharnweber, D.; Hintze, V. Sulfated Hyaluronan Derivatives Modulate TGF-B1:Receptor Complex Formation: Possible Consequences for TGF-B1 Signaling. *Sci. Rep.* **2017**, *7*, No. 1210.

(20) Crijns, H.; Vanheule, V.; Proost, P. Targeting Chemokine—Glycosaminoglycan Interactions to Inhibit Inflammation. *Front. Immunol.* **2020**, *11*, No. 483.

(21) Luster, A. D. Chemokines—Chemotactic Cytokines That Mediate Inflammation. *N. Engl. J. Med.* **1998**, *338*, 436–445.

(22) Pichert, A.; Samsonov, S. A.; Theisgen, S.; Thomas, L.; Baumann, L.; Schiller, J.; Beck-Sickinge, A. G.; Huster, D.; Pisabarro, M. T. Characterization of the Interaction of Interleukin-8 with Hyaluronan, Chondroitin Sulfate, Dermatan Sulfate and Their Sulfated Derivatives by Spectroscopy and Molecular Modeling. *Glycobiology* **2012**, *22*, 134–145.

(23) Schlorke, D.; Thomas, L.; Samsonov, S. A.; Huster, D.; Arnhold, J.; Pichert, A. The Influence of Glycosaminoglycans on IL-8-Mediated Functions of Neutrophils. *Carbohydr. Res.* **2012**, *356*, 196–203.

(24) Joseph, P. R. B.; Mosier, P. D.; Desai, U. R.; 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.

(25) Künze, G.; Gehrcke, J. P.; Pisabarro, M. T.; Huster, D. NMR Characterization of the Binding Properties and Conformation of Glycosaminoglycans Interacting with Interleukin-10. *Glycobiology* **2014**, *24*, 1036–1049.

- (26) Künze, G.; Köhling, S.; Vogel, A.; Rademann, J.; Huster, D. Identification of the Glycosaminoglycan Binding Site of Interleukin-10 by NMR Spectroscopy. *J. Biol. Chem.* **2016**, *291*, 3100–3113.
- (27) Tanegashima, K.; Suzuki, K.; Nakayama, Y.; Hara, T. Antibody-Assisted Enhancement of Biological Activities of CXCL14 in Human Monocytic Leukemia-Derived THP-1 Cells and High Fat Diet-Induced Obese Mice. *Exp. Cell Res.* **2010**, *316*, 1263–1270.
- (28) Gandhi, N. S.; Mancera, R. L. Molecular Dynamics Simulations of CXCL-8 and Its Interactions with a Receptor Peptide, Heparin Fragments, and Sulfated Linked Cyclitols. *J. Chem. Inf. Model.* **2011**, *51*, 335–358.
- (29) Marcisz, M.; Huard, B.; Lipska, A. G.; Samsonov, S. A. Further Analyses of APRIL/APRIL-Receptor/Glycosaminoglycan Interactions by Biochemical Assays Linked to Computational Studies. *Glycobiology* **2021**, *31*, 772–786.
- (30) Gehrcke, J. P.; Pisabarro, M. T. Identification and Characterization of a Glycosaminoglycan Binding Site on Interleukin-10 via Molecular Simulation Methods. *J. Mol. Graphics Modell.* **2015**, *62*, 97–104.
- (31) Panitz, N.; Theisgen, S.; Samsonov, S. A.; Gehrcke, J. P.; Baumann, L.; Bellmann-Sickert, K.; Köhling, S.; Teresa Pisabarro, M.; Rademann, J.; Huster, D.; Beck-Sicking, A. G. The Structural Investigation of Glycosaminoglycan Binding to CXCL12 Displays Distinct Interaction Sites. *Glycobiology* **2016**, *26*, 1209–1221.
- (32) Bojarski, K. K.; Sieradzian, A. K.; Samsonov, S. A. Molecular Dynamics Insights into Protein-Glycosaminoglycan Systems from Microsecond-Scale Simulations. *Biopolymers* **2019**, *110*, No. e23252.
- (33) Kogut, M. M.; Maszota-Zieleniak, M.; Marcisz, M.; Samsonov, S. A. Computational Insights into the Role of Calcium Ions in Protein-Glycosaminoglycan Systems. *Phys. Chem. Chem. Phys.* **2021**, *23*, 3519–3530.
- (34) Uciechowska-Kaczmarzyk, U.; de Beauchene, I.; Samsonov, S. A. Docking Software Performance in Protein-Glycosaminoglycan Systems. *J. Mol. Graphics Modell.* **2019**, *90*, 42–50.
- (35) Samsonov, S. A.; Zacharias, M.; Beauchene, I. C. de. Modeling Large Protein–Glycosaminoglycan Complexes Using a Fragment-Based Approach. *J. Comput. Chem.* **2019**, *40*, 1429–1439.
- (36) Siebenmorgen, T.; Engelhard, M.; Zacharias, M. Prediction of Protein–Protein Complexes Using Replica Exchange with Repulsive Scaling. *J. Comput. Chem.* **2020**, *41*, 1436–1447.
- (37) Maszota-Zieleniak, M.; Marcisz, M.; Kogut, M. M.; Siebenmorgen, T.; Zacharias, M.; Samsonov, S. A. Evaluation of Replica Exchange with Repulsive Scaling Approach for Docking Glycosaminoglycans. *J. Comput. Chem.* **2021**, *42*, 1040–1053.
- (38) Gutterres, H.; Im, W. Improving Protein-Ligand Docking Results with High-Throughput Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2020**, *60*, 2189–2198.
- (39) Ahmadian, N.; Mehrnejad, F.; Amininasab, M. Molecular Insight into the Interaction between Camptothecin and Acyclic Cucurbit[4]Urils as Efficient Nanocontainers in Comparison with Cucurbit[7]Uril: Molecular Docking and Molecular Dynamics Simulation. *J. Chem. Inf. Model.* **2020**, *60*, 1791–1803.
- (40) Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA Methods to Estimate Ligand-Binding Affinities. *Expert Opin. Drug Discovery* **2015**, *10*, 449–461.
- (41) Samsonov, S. A.; Pisabarro, M. T. Computational Analysis of Interactions in Structurally Available Protein–Glycosaminoglycan Complexes. *Glycobiology* **2016**, *26*, 850–861.
- (42) Hintze, V.; Samsonov, S. A.; Anselmi, M.; Moeller, S.; Becher, J.; Schnabelrauch, M.; Scharnweber, D.; Pisabarro, M. T. Sulfated Glycosaminoglycans Exploit the Conformational Plasticity of Bone Morphogenetic Protein-2 (BMP-2) and Alter the Interaction Profile with Its Receptor. *Biomacromolecules* **2014**, *15*, 3083–3092.
- (43) Nordsieck, K.; Pichert, A.; Samsonov, S. A.; Thomas, L.; Berger, C.; Pisabarro, M. T.; Huster, D.; Beck-Sicking, A. G. Residue 75 of Interleukin-8 Is Crucial for Its Interactions with Glycosaminoglycans. *ChemBioChem* **2012**, *13*, 2558–2566.
- (44) Möbius, K.; Nordsieck, K.; Pichert, A.; Samsonov, S. A.; Thomas, L.; Schiller, J.; Kalkhof, S.; Teresa Pisabarro, M.; Beck-Sicking, A. G.; Huster, D. Investigation of Lysine Side Chain Interactions of Interleukin-8 with Heparin and Other Glycosaminoglycans Studied by a Methylation-NMR Approach. *Glycobiology* **2013**, *23*, 1260–1269.
- (45) Samsonov, S. A.; Bichmann, L.; Pisabarro, M. T. Coarse-Grained Model of Glycosaminoglycans. *J. Chem. Inf. Model.* **2015**, *55*, 114–124.
- (46) Samsonov, S. A.; Pisabarro, M. T. Computational Analysis of Interactions in Structurally Available Protein-Glycosaminoglycan Complexes. *Glycobiology* **2016**, *26*, 850–861.
- (47) Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E., III; Cruzeiro, V.W.D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Greene, D.; Harris, R.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D. J.; Merz, K. M.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Smith, J.; Salomon-Ferrer, R.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A. *AMBER16*; University of California: San Francisco, 2018.
- (48) Samsonov, S. A.; Gehrcke, J. P.; Pisabarro, M. T. Flexibility and Explicit Solvent in Molecular-Dynamics-Based Docking of Protein-Glycosaminoglycan Systems. *J. Chem. Inf. Model.* **2014**, *54*, 582–592.
- (49) DiGabriele, A. D.; Lax, I.; Chen, D. I.; Svahn, C. M.; Jaye, M.; Schlessinger, J.; Hendrickson, W. A. Structure of a Heparin-Linked Biologically Active Dimer of Fibroblast Growth Factor. *Nature* **1998**, *393*, 812–817.
- (50) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. GLYCAM06: A Generalizable Biomolecular Force Field. *Carbohydrates. J. Comput. Chem.* **2008**, *29*, 622–655.
- (51) Huige, C. J. M.; Altona, C. 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.
- (52) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function. *J. Comput. Chem.* **1998**, DOI: 10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B.
- (53) Ester, M.; Kriegel, H.-P.; Sandervina, J.; Xu, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise; AAAI, 1996; pp 226–231.
- (54) Onufriev, A.; Case, D. A.; Bashford, D. Effective Born Radii in the Generalized Born Approximation: The Importance of Being Perfect. *J. Comput. Chem.* **2002**, *23*, No. 1297.
- (55) Tschlp, W.; Krcmer, K.; Baloulis, J.; Bürger, T.; Halm, O. Simulation of Polymer Melts. I. Coarse-Graining Procedure for Polycarbonates. *Acta Polym.* **1998**, *49*, 61–74.
- (56) Nivedha, A. K.; Thieker, D. F.; Makeneni, S.; Hu, H.; Woods, R. J. Vina-Carb: Improving Glycosidic Angles during Carbohydrate Docking. *J. Chem. Theory Comput.* **2016**, *12*, 892–901.
- (57) Boittier, E. D.; Burns, J. M.; Gandhi, N. S.; Ferro, V. GlycoTorch Vina: Docking Designed and Tested for Glycosaminoglycans. *J. Chem. Inf. Model.* **2020**, *60*, 6328–6343.

## Publication D4

Further analyses of APRIL/APRIL-receptor/glycosaminoglycan interactions by biochemical assays linked to computational studies

## Computational Biology

# Further analyses of APRIL/APRIL-receptor/ glycosaminoglycan interactions by biochemical assays linked to computational studies

Mateusz Marcisz<sup>2,3</sup>, Bertrand Huard<sup>4</sup>, Agnieszka G Lipska<sup>2</sup> and Sergey A Samsonov<sup>1,2</sup>

<sup>2</sup>Faculty of Chemistry, University of Gdańsk, ul. Wita Stwosza 63, Gdańsk 80-308, Poland, <sup>3</sup>Intercollegiate Faculty of Biotechnology of UG and MUG, ul. Abrahama 58, Gdańsk 80-307, Poland, and <sup>4</sup>TIMC-IMAG, Université Grenoble-Alpes, CNRS UMR, La Tronche 5525, France

<sup>1</sup>To whom correspondence should be addressed: Tel: +48-58-523-51-66; Fax: +48-58-523-50-12; e-mail: sergey.samsonov@ug.edu.pl

Received 4 January 2021; Revised 23 February 2021; Editorial Decision 23 February 2021; Accepted 23 February 2021

## Abstract

A proliferation-inducing ligand (APRIL) is a member of the tumor necrosis factor superfamily. APRIL is quite unique in this superfamily for at least for two reasons: (i) it binds to glycosaminoglycans (GAGs) via its positively charged N-terminus; (ii) one of its signaling receptors, the transmembrane activator and CAML interactor (TACI), was also reported to bind GAGs. Here, as provided by biochemical evidences with the use of an APRIL deletion mutant linked to computational studies, APRIL–GAG interaction involved other regions than the APRIL N-terminus. Preferential interaction of APRIL with heparin followed by chondroitin sulfate E was confirmed by *in silico* analysis. Both computational and experimental approaches did not reveal the heparan sulfate binding to TACI. Together, computational results corroborated experiments contributing with atomistic details to the knowledge on this biologically relevant trimolecular system. Additionally, a rigorous high-throughput analysis of the free energy calculations data was performed to critically evaluate the applied computational methodologies.

**Key words:** APRIL, ELISA, flow cytometry, glycosaminoglycans, MM/GBSA

## Introduction

Heparin (HP), chondroitin sulfate-C (CSc) and -E (CSe) belong to glycosaminoglycans (GAGs). GAGs are long, linear, anionic and periodic polysaccharides playing a crucial role via interactions with a variety of proteins in the extracellular matrix processes. They are made of repeating disaccharide unit consisting of an amino sugar and a uronic acid or galactose (Varki *et al.* 2015). Those saccharide units may manifest different sulfation patterns that influence the polysaccharide conformational and binding properties (Habuchi *et al.* 2004). In many cases, protein–GAG interactions are considered as nonspecific and mostly electrostatic-driven due to the high negative charge of the polysaccharides and positive charge of the protein-binding sites

(Imberty *et al.* 2007). Some of the proteins that interact with GAGs belong to the group of growth factors (Uciechowska-Kaczmarzyk *et al.* 2018; Bojarski *et al.* 2019) and chemokines (Derler *et al.* 2017; Nordsieck *et al.* 2018; Penk *et al.* 2019). In case of fibroblast growth factors (FGF), GAGs may form complexes with FGF1 (Digabriele *et al.* 1998) and FGF2 (Faham *et al.* 1996). GAGs can enhance the activity of the growth factors by either changing their conformation or by binding multiple FGFs and thus facilitating oligomerization of FGFR receptors, which plays a role in cell signaling (Mason 1994; Faham *et al.* 1996). Our recent computational study implementing a microsecond-scale molecular dynamics (MD) simulations (Bojarski *et al.* 2019) showed few novel insights on the



HP-FGF1 interactions and helped to deepen the knowledge on the topic of protein-GAG molecular systems. In particular, it was shown that the length of the simulation could be crucial for the calculation of the protein-GAG molecular system dynamic and energetic parameters, while conformational selection mechanism of binding as well as recognition specificity determinant were proposed from these long simulations for this complex. Another recent computational study on HP (Uciechowska-Kaczmarzyk et al. 2018) shows its effect on the dynamics of vascular endothelial growth factor (VEGF), a key player in the angiogenesis and regenerative processes, arthritis and cancer (Risau 1997). In this work, it was demonstrated that GAG binding could induce global conformational changes of a protein target, rendering its capability to bind its receptor on the cell membrane (Uciechowska-Kaczmarzyk et al. 2018). One more growth factor that is known to bind HP and heparan sulfate is transforming growth factor  $\beta$ 1 (TGF- $\beta$ 1) (McCaffrey et al. 1992; Lyon et al. 1997) TGF- $\beta$ 1's role is to regulate proliferation, adhesion, cell migration and differentiation (Moustakas et al. 2001). There is evidence that sulfated hyaluronan derivatives are able to bind this growth factor and so can enhance or inhibit its activity depending on GAG sulfation and the order of binding events in its tertiary complex with its receptors (Van Der Smissen et al. 2013; Koehler et al. 2017). Second important group of proteins that interacts with GAGs are chemokines (Derler et al. 2017; Crijns et al. 2020). They are a large group of predominantly proinflammatory cytokines, and they may influence the cell in a variety of different ways—while some of them can promote angiogenesis, tumor growth and metastasis, others can inhibit them (Luster 1998). GAGs are also known to bind these molecules mediating the activation of the leukocytes by affecting the binding of a respective chemokine to a G protein-coupled receptor (Larsen et al. 1989). Many computational and experimental studies successfully investigated GAGs' effects and interactions with chemokines/cytokines, such as CXCL-8 (Gandhi and Mancera 2011; Joseph et al. 2015; Nordsieck et al. 2018), IL-10 (Künze et al. 2014; Gehrcke and Pisabarro 2015; Künze et al. 2016), CXCL-12 (Panitz et al. 2016) and CXCL-14 (Penk et al. 2019). This proved theoretical approaches not only as useful complementation to experimental studies but also as an important and stand-alone work.

A proliferation-inducing ligand (APRIL) is a member of the tumor necrosis factor (TNF) superfamily (Hahne et al. 1998). It is produced first as a transmembrane protein before being processed by furin proteases to act as a soluble factor (López-Fraga et al. 2001). Notably, APRIL also binds to the GAGs of the heparan sulfate (HS) class (Hendriks et al. 2005; Ingold et al. 2005). Such binding is quite unique in the TNF superfamily since only one other member, ectodysplasin A, has been reported to date to bind GAG (Swee et al. 2009). APRIL binding to GAG allows its oligomerization to efficiently signal into target cells (Kimberley et al. 2009). Indeed, unlike other TNF ligands such as the TNF itself, the soluble APRIL trimer formed by noncovalent, mostly hydrophobic, interactions between beta-pleated sheets is not active to signal its two receptors transmembrane activator and CAML interactor (TACI) and B-cell maturation antigen (BCMA) (Bossen et al. 2008). The numerous negatively charged sulfate residues along the heparan chain of GAG create a platform with multiple binding sites for APRIL, hence mediating oligomerization. The GAG-binding region of APRIL has been located in its N-terminus that contains a stretch of positively charged lysine, three and four in the human and mouse molecules, respectively. Addition of a cross-linking anti-flag antibody renders trimeric flag-tagged APRIL signalization active, indicating that APRIL does not need a high order of oligomerization to signal. One of the APRIL signaling receptor, TACI but not BCMA, was further shown to interact with GAGs

(Bischof et al. 2006; Sakurai et al. 2007; Moreaux et al. 2009). Such ternary complex between a ligand, a coreceptor and a receptor may resemble the one described for FGF/FGF-R (Pomin 2016). APRIL main cellular targets are the antibody-producing plasmocytes (Baert et al. 2018). On the surface of these cells, TACI and/or BCMA, depending on their stage of differentiation, are present as the APRIL signaling receptors. They also express ubiquitously a unique GAG, CD138, also known as syndecan-1 (Wijdenes et al. 1996). CD138 has a mixed composition of HS and chondroitin (CS) chains (Kokenyesi and Bernfield 1994). However, only HS on CD138 appears to play a role on APRIL binding (Matthes et al. 2015). Recently, a new target cell for APRIL has been identified in the central nervous system with astrocytes (Baert et al. 2019). Notably, APRIL binds to CS GAG on astrocytes, and selectivity in APRIL binding according to CS types was observed. Here, we further investigate the binding of APRIL to HP and CS GAG in a computational and experimental study.

## Materials and methods

### Structures

**GAG Structures.** All the GAG structures—heparin (HP) tetramer/dp4 and hexamer/dp6 (dp stands for degree of polymerization), chondroitin-4,6-sulfate (CSe) dp4 and dp6, chondroitin-6-sulfate (CSc)  $\delta$ dp4 and dp6—were constructed from the building blocks of the sulfated GAG monomeric units' libraries (Pichert et al. 2012) that are compatible with the AMBER16 package (Case et al. 2018). GLYCAM06 force field (Kirschner et al. 2008) and literature data (Huige and Altona 1995) were the sources of GAGs' charges.

**Protein Structures.** The structure from PDB ID 4ZCH (2.43 Å) (Schuepbach-Mallepell et al. 2015) of the single-chain human APRIL protein from the APRIL-BAFF-BAFF complex was used for the construction of truncated human variant trimer— $H_{115}$ APRIL (136 amino acid residues starting from HIS 115). For this, chimera (Pettersen et al. 2004) and Modeller (Šali and Blundell 1993) were used to obtain the model of the human  $H_{115}$ APRIL trimer based on its murine homolog—PDB ID 1XU1 (1.90 Å) structure (Hymowitz et al. 2005). Later, using AMBER software package, minimization and 10 ns equilibration by MD simulation were performed to obtain the structure used for the further studies (see the details on the MD protocols Molecular dynamics).

In case of the full-length—natural form— $A_{105}$ APRIL protein variant, additional 10 amino acid residues were added to the N-terminus (146 amino acid residues in total, starting from ALA 105) of each monomer. Coarse-grained modeling UNRES (from UNited RESidue) software (Liwo et al. 1997) was used to predict the structure of the trimer (see the protocol Calculation of full-length APRIL protein model in UNRES). Five different models corresponding to the energy minima were obtained. To choose the best one in terms of potential GAG binding, docking of HP dp4 to APRIL trimer and MD simulations were performed. Models with the overall lowest binding free energy were chosen for further analysis. To calculate energies, molecular mechanics generalized Born surface area (MM/GBSA) and linear interaction energy (LIE) analysis methods were used (see the protocols below).

Peptide fragment corresponding to the N-terminus of the full-length human variant and missing in the truncated one (ALA-VAL-LEU-THR-GLN-LYS-GLN-LYS-LYS-GLN) was minimized, and extensive MD simulation of 12  $\mu$ s was performed in order to analyze its structural properties.

## Calculation of full-length APRIL protein model in UNRES

In order to calculate the conformations of the N-terminal fragments of each chain, we applied a coarse-grained multiplexed replica exchange molecular dynamics (MREMD) (Sugita and Okamoto 2000; Young and Pande 2003) approach implemented in UNRES (Adcock and McCammon 2006; Czaplewski et al. 2009). The protocol was similar to that used in our previous work (Vallet et al. 2018; Porthoff et al. 2019). Distance restraints were imposed on protein, except the first 10 amino acids of each chain. MREMD simulation consisted of 40 trajectories run at temperatures from 240 K to 350 K, with two trajectories for every temperature. Each trajectory consisted of  $1.6 \times 10^7$  MD steps with 4.89 fs length. Only conformations from the second part of the simulation were taken into further analysis, with the use of the weighted histogram analysis method (WHAM) (Kumar et al. 1992). The next step was minimum variance cluster analysis (Ling and Späth 1981) of the conformational ensemble at  $T = 300$  K, which enabled us to obtain five clusters, ranked according to summary probabilities of the ensembles and containing the most probable structures to the cluster with the least probable structures. For each cluster, one representative structure, closest to the cluster centroid, was selected as the representative conformation.

## Electrostatic potential calculations

Amber16 PBSA (Poisson–Boltzmann surface area) program was used to calculate and visualize the solvent-mediated electrostatic potentials of different variants of the APRIL protein using a default 1 Å grid spacing step. Later on, it was visualized using VMD program (Humphrey et al. 1996) to assess the regions of potential GAG binding to APRIL protein. Previously, this method proved to be successful in GAG-binding regions' predictions (Samsonov and Pisabarro 2016).

## Molecular docking

Autodock 3 (Morris et al. 1998) was used for docking simulations. This program with this particular version yielded the best results among other docking programs for protein–GAG systems (Samsonov and Pisabarro 2016; Uciechowska-Kaczmarzyk et al. 2019). Maximum gridbox size was used ( $126 \text{ \AA} \times 126 \text{ \AA} \times 126 \text{ \AA}$ ), which covered the entire APRIL region of predicted GAG-binding region with the default grid step of 0.375 Å. Lamarckian genetic algorithm was used for 1000 independent runs. The size of 300 for the initial population and  $10^5$  generations for termination conditions were chosen.  $9995 \times 10^5$  energy evaluations were performed. Clustering was performed using DBSCAN algorithm (Ester et al. 1996) on top 50 docking results. The metric used for clustering accounted for the equivalence of the atoms of the same atomic type, which is more appropriate than the classical RMSD for periodic ligands (Samsonov et al. 2014). One to two clusters of each GAG's docking solutions were chosen for the further analysis. Each time clustering parameter was chosen individually to obtain one to three representative clusters.

## Molecular dynamics

Every all-atom MD simulation of different APRIL variants complexes obtained from molecular docking was performed using AMBER16 software package. Truncated octahedron TIP3P periodic box of 8 Å water layer from the box's border to solute was used to solvate complexes. Charge was neutralized with  $\text{Cl}^-$  counterions.

Cysteines were connected to form appropriate disulfide bridges according to the structure from the PDB (PDB ID 1XU1) (Hymowitz et al. 2005; Schuepbach-Mallepell et al. 2015). Energy minimization was performed preceding the production MD runs: 500 steepest descent cycles and  $10^3$  conjugate gradient cycles with 100 kcal/mol/Å<sup>2</sup> harmonic force restraints, continued with  $3 \times 10^3$  steepest descent cycles and  $3 \times 10^3$  conjugate gradient cycles without any restraints. Following minimization steps, the system was heat up to 300 K for 10 ps with harmonic force restraints of 100 kcal/mol/Å<sup>2</sup>. Then, the system was equilibrated at 300 K and  $10^5$  Pa in the isothermal isobaric ensemble for 500 ps. Afterward, the actual MD run was carried out in the same isothermal isobaric ensemble for either 10 ns or 100 ns (except for the 10 amino acid residues peptide, in which case, 12 μs MD run was performed). 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

To calculate free energy and per-residue energy decomposition, the obtained trajectories from MD simulations were analyzed using AMBER16 by two approaches—MM/GBSA model  $\text{igb} = 2$  (Onufriev et al. 2002) and LIE analysis with dielectric constant of 80, performed by CPPTRAJ scripts. Particular frames taken for this analysis varied for different simulations to be representative in terms of the structural convergence.

## Dynamic molecular docking

Compatibly with the MD simulations described Molecular dynamics, AMBER16 was used for dynamic molecular docking (DMD). This method allows for full flexibility of both the receptor and ligand as well as for taking into account the explicit solvent (Samsonov et al. 2014). First, a ligand was placed at 20–30 Å from the surface of the protein (50–60 Å from the center of the protein), which was significantly more than the cutoff value = 8 Å that was used in the MD to avoid any influence on the dynamics of the ligand at the beginning of the docking run. Then, the truncated octahedron TIP3P periodic box of 4 Å water layer from the box's border to solute was used to solvate the complex, and the charge was neutralized using  $\text{Cl}^-$  counterions. Disulfide bonds between cysteines were created accordingly with the structure from the PDB. Minimization and equilibration runs were performed as described previously in the Molecular dynamics section. After equilibration, the distances from ligand (O4S atom of residue 47Y) to protein (O atom of residue GLY403) were calculated and were assigned to the initial distance in the targeted MD run. Afterward, first 4 ns MD run was performed with the biased potential of 200 kcal/mol/Å<sup>2</sup> applied to the above-mentioned atoms of the receptor and the ligand as described in our previous work (Samsonov et al. 2014), employing Jarzynski procedure (Jarzynski 1997; Park and Schulten 2004) This step was repeated 100 times to obtain 100 different docking poses. Next, structures from final frames from all runs were taken as the starting structures for the unbiased MD simulation; each protein–GAG complex was solvated with 8 Å layer of TIP3P water in a form of truncated octahedron;  $\text{Na}^+$  or  $\text{Cl}^-$  counterions were added and disulfide bonds were created. Minimization and equilibration steps were performed once more under the same conditions as in the first step of the DMD simulation. Finally, 10 ns MD production runs were carried out in the same isothermal isobaric ensemble for each

of the 100 docking poses. The obtained structures were considered as docking poses and were scored used the MM/GBSA protocol for all 10 ns of the MD production run.

### Cells and reagents

L363 and 293 T HEK cells were obtained from the American tissue culture collection. All cells were propagated in RPMI 1640 medium (Gibco BRL, US) containing 10% fetal calf serum (Eurobio, France). Biotinylated anti-flag (clone M2) was from Sigma (US). The anti-TACI (1A1, rat immunoglobulin G2a [IgG2a]) and soluble human Thy-1 (aa 20-130), BCMA (aa 2-54) TACI (aa 2-118) fused to human Fc were obtained from EnzoLife sciences (Switzerland). Expression constructs for Fc-tagged and flag-tagged human  $A_{105}$ APRIL (aa 105-246) and  $H_{115}$ APRIL (aa 115-246) have been described previously (Baert et al. 2018). SDS-PAGE analysis was performed to confirm dimerization of APRIL trimer proteins (Supplementary Figure S1). Fc-tagged and flag-tagged proteins were produced transiently in 293 T HEK cells following polyethylenimine-based transfection in serum-free opti-mem medium (Gibco BRL). Fc-tagged and flag-tagged proteins were purified with Protein-A Sepharose (GE Healthcare, US) and anti-flag-Sepharose (Sigma, US) and acidic elution with 0,1 M glycine, pH: 2.5, respectively. Positive fractions were pooled and dialyzed against PBS before use. The plasmid encoding for human TACI has been described by (Schwaller et al. 2007). Heparin (Liquemin, 5000 IU/mL, Drossapharm, Switzerland) was used at 1/500. Heparan sulfate proteoglycan (HSPG) from mouse basement membrane sarcoma were obtained from Sigma. HSPG were biotinylated with biotin hydrazide and carbodiimide as previously described (Ahmed and Huard 2021).

### Enzyme-linked immunosorbent assay

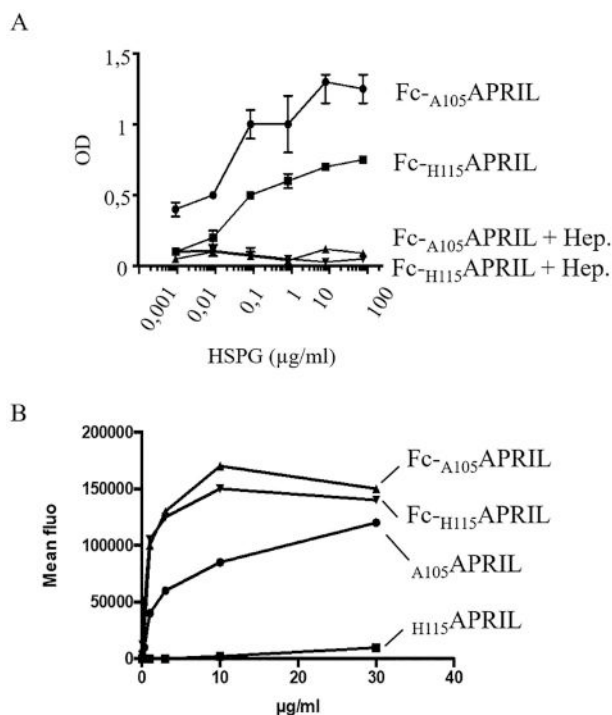
Proteins coating was performed overnight at 10  $\mu$ g/mL in 50  $\mu$ L of PBS at 4°C. Plates were blocked in PBS with 1% BSA for 1 h at RT. Incubation with ligands were performed for 1 h at RT. Washing buffer was PBS with 1% BSA, 0.05% Tween 20. Streptavidin conjugated to horseradish peroxidase (R&D Systems, US) and 3,3',5,5'-tetramethylbenzidine (Sigma) were used to reveal binding of biotinylated HSPG. Reaction was stopped with  $H_2SO_4$  2 N. Optical density was read at 450 nm on a VICTOR multilabel plate reader (Perkin Elmer, UK).

### Flow cytometry

$0,5 \times 10^6$  cells were stained at 4°C in PBS with 1% BSA for 30 min. Secondary reagents included goat antirat (anti-TACI, 10  $\mu$ g/mL) and goat antihuman immunoglobulin G (IgG) (Fc-APRIL, TACI-Fc) conjugated to Alexa 488. Streptavidin conjugated to Alexa 488 (BD Biosciences) was used to detect HSPG and anti-flag binding. Washes were performed with PBS. Fluorescence was acquired on a BD Accuri C6 flow cytometer.

## Results and discussion

In this work, APRIL-GAG interactions were comprehensively analyzed. For all the experiments, trimers of APRIL were used (with the addition of the hexameric variants that were used for biological experiments) to reproduce the conditions in the cell. It was shown in enzyme-linked immunosorbent assay (ELISA) and flow cytometry experiments that HSPG binds to both wild-type and truncated variants of APRIL protein. Potential region of GAG binding on the



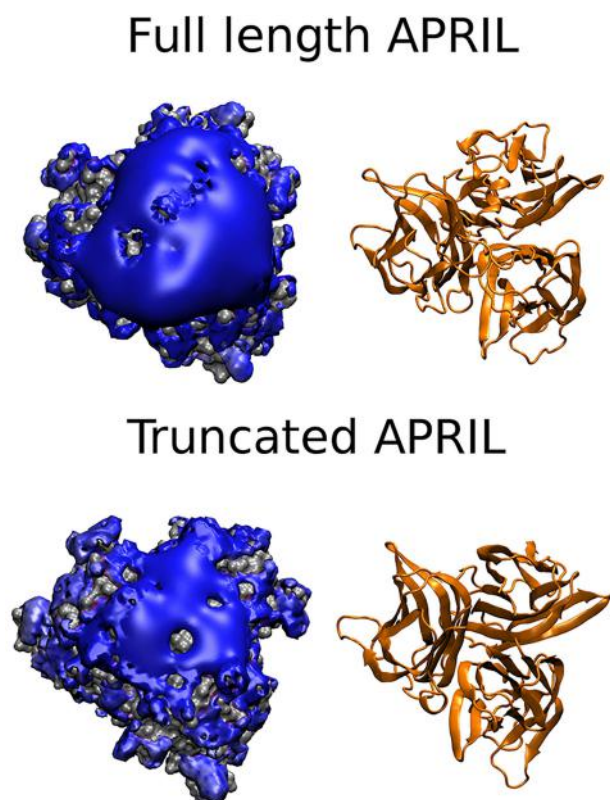
**Fig. 1.** Binding of oligomerized  $H_{115}$ APRIL to HSPG. **(A)** The binding of increasing concentrations of biotinylated HSPG to the indicated forms of coated APRIL was tested by ELISA. In some conditions, heparin was used as a competitor. Results are shown with the mean and min/max values of duplicates. Three independent experiments were performed. **(B)** Binding of trimeric APRIL and hexameric Fc-APRIL on HSPG expressed at the surface of HEK cells was assessed by flow cytometry. The results are representative of at least five for **(A)** and two for **(B)** independent experiments.

surface of the APRIL was inspected with the PBSA method. Then, using computational approach, the results from ELISA/flow cytometry were confirmed. Furthermore, interactions of APRIL protein with different lengths of HP, CSe and CSc were examined. It was shown which residues promote GAG binding and which obstruct it. Then, the structural properties of N-terminal fragment that truncated APRIL was lacking were analyzed. Afterward, experimental and computational investigations of APRIL's receptors—TACI and BCMA—and their potential binding to HP were carried out. At the end, an in-depth analysis of the performance and prediction power of tools used in this study was carried out.

### Binding of $H_{115}$ APRIL/ $A_{105}$ APRIL with HSPG

We assessed the binding of  $H_{115}$ APRIL to HSPG by ELISA. We could detect a binding to soluble HSPG when  $H_{115}$ APRIL was oligomerized upon coating onto the plastic surface (Figure 1A). This binding was to a lesser extent than for coated  $A_{105}$ APRIL. We also detected by flow cytometry that a binding of  $H_{115}$ APRIL to HSPG expressed at the surface of HEK cells, but again, only when it was oligomerized by dimerization with an Fc fusion partner (Figure 1B). In this latter experiment, trimeric  $A_{105}$ APRIL could bind to the HEK cells. Taken together, these experiments show that  $H_{115}$ APRIL lacking its N-terminal tail could still bind to HSPG but with an overall lower affinity since requiring oligomerization.

After the experimental confirmation of binding of APRIL protein variants to HSPG, we performed a series of computational experiments to gain atomistic insights into the APRIL-GAG interactions.



**Fig. 2.** Positive electrostatic potential isosurfaces (colored in blue; 4 kcal/mol $\cdot$ e $^{-1}$  in case of truncated APRIL and 5 kcal/mol $\cdot$ e $^{-1}$  in case of full-length APRIL) obtained by PBSA approach show potential capability of N- and C-terminal regions of the APRIL protein to bind GAGs.

### PBSA-based prediction of GAG-binding region

To predict GAG-binding regions on the APRIL protein surface, PBSA method from AMBER suite was used. It was previously proven that, due to the highly charged nature of these ligands, such methodology is successful in predicting GAG-binding regions on the protein surface (Samsonov and Pisabarro 2016). The results obtained with the PBSA method showed positive electrostatic potential in the N-terminal region of each monomer that is spatially close in the trimeric APRIL (Figure 2). This region with the three lysines, LYS 110, LYS 112 and LYS 113, was previously demonstrated in binding assays with HS-positive cells to be responsible for binding GAGs (Hendriks et al. 2005; Ingold et al. 2005). Nevertheless, we found out that the C-termini may contribute to the positive potential suggesting that APRIL–GAG interactions may be more complex than previously thought.

### Computational investigation of APRIL–GAG interactions

HP dp4, HP dp6, CSe dp4, CSe dp6, CSc dp4 and CSc dp6 were analyzed for their binding properties to full-length soluble APRIL ( $A_{105}$ APRIL) and a N-term truncated variant ( $H_{115}$ APRIL) using MM/GBSA and LIE methods, with protocols calibrated for this system as described at the end of the Results section. To obtain the  $A_{105}$ APRIL protein structure, we applied coarse-grained approach to calculate theoretical models for the protein as there is no experimental structure available yet. In order to achieve this, we used  $H_{115}$ APRIL variant PDB structure and modeled additional 10 amino

acid residues using UNRES software as described in the Materials and methods section. This full-length protein model was used for further APRIL–GAG interactions analysis along with truncated variant for molecular docking and binding free energy assessment. Both MM/GBSA and LIE were in agreement that HP binds the strongest out of the compared GAGs. The weakest binding was found for CSc, while the CSe bound stronger than CSc but weaker than HP, suggesting not only net electrostatic effect on binding but also the specific role of sulfation pattern for CS.

All the binding energies obtained by LIE and MM/GBSA for both  $A_{105}$ APRIL and  $H_{115}$ APRIL variants are listed in the Table I. Overall, GAGs bind much stronger to  $A_{105}$ APRIL than to  $H_{115}$ APRIL. This finding is in agreement with our PBSA electrostatic potential analysis that suggested much higher positive potential in the region of N-terminus. It is also clear that dp6 GAGs are bound stronger than dp4 counterparts, especially for CSe and HP GAGs. The only exception is CSc dp4 in the case of  $A_{105}$ APRIL that shows insignificantly lower MM/GBSA and LIE values than for CSc dp6. However, the overall differences in energies are lower for CSc than for other GAGs, which suggests very weak or no binding for this GAG. The differences in the results obtained by MM/GBSA and LIE methods may appear due to few factors. First of all, LIE is a less complex method and may not account for as many energy components, and in particular, in terms of solvent treatment, as the MM/GBSA does. There are several differences in free energy calculation in both methods (Genheden and Ryde 2011): LIE takes into account the Van der Waals (VdW) component and the electrostatic component (ELE) in vacuo scaled by a dielectric constant to consider the interactions between the ligand with the receptor and solvent environments; MM/GBSA calculates free energy using VdW and electrostatics in vacuo energies, polar solvation energy partially including the entropic component of the solvent, nonpolar solvation free energy. Second, LIE is used in our work with standard parameters (dielectric constant of 80), which should be calibrated for each particular molecular system type using experimental data if available.

### Significant residues in the APRIL–GAG-binding site

All of the mentioned GAGs were docked with Autodock3 software to the N-/C-terminus region of the APRIL protein (Figure 3). Per-residue free energy decomposition analysis allowed to propose the residues defining the putative binding site. The most important residues for  $H_{115}$ APRIL were: HIS 115, ARG 143, ARG 144, GLY 145, ARG 146 and LYS 249 (Supplementary Figures S2 and S3), while the following residues provided the most unfavorable contribution to the binding: ASP 159, GLU 185, GLU 191, ASP 223 and LEU 250 (Supplementary Figure S4). For  $A_{105}$ APRIL, the same residues were found to be disruptive (Figure 4). However, when analyzing the most important residues favorable for binding, there was a significant difference because of the additional LYS 110, GLN 111, LYS 112 and 113 that were present in the N-terminus (Figure 5 and Supplementary Figure S2). This confirmed that  $A_{105}$ APRIL is more prone to binding GAGs.

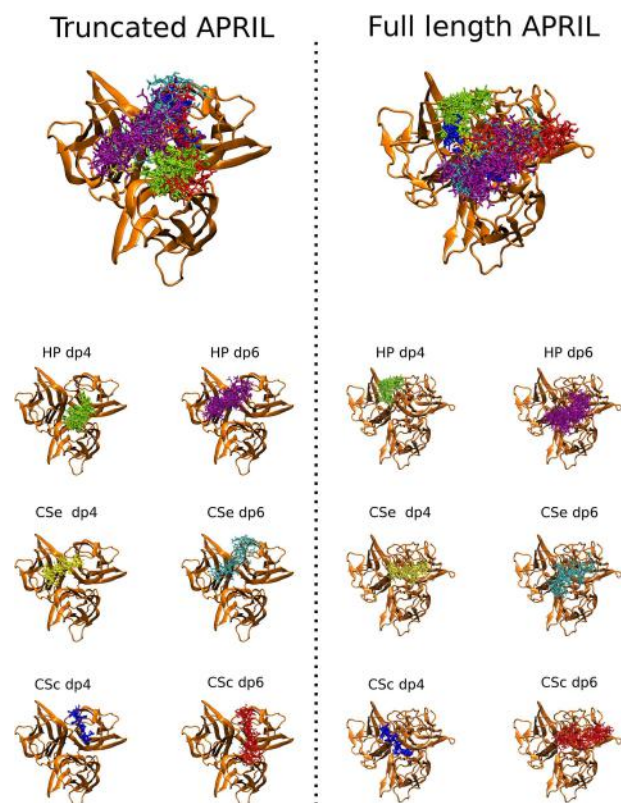
### Heparin binding to the LYS-rich N-term peptide of APRIL

Additional computational experiments were performed to analyze more comprehensively APRIL–GAGs interactions. We took the first 10 amino-acid residues from the N-terminus of the APRIL protein that are missing in the truncated form and simulated (as described in the Methods and materials section) them for over 12  $\mu$ s to assure

**Table 1.** MM/GBSA and LIE binding free energy analysis for the truncated and full-length APRIL in complex with HP dp4 and CSc dp4

	H115APRIL, $\Delta G$ (kcal/mol)		H115APRIL, $\Delta G$ (kcal/mol)	
	MM/GBSA	LIE	MM/GBSA	LIE
HP dp4	$-56.4 \pm 8.3$	$-38.1 \pm 5.8$	$-119.9 \pm 20.8$	$-61.9 \pm 9.2$
HP dp6	$-88.4 \pm 16.7$	$-50.3 \pm 9.5$	$-141.1 \pm 23.5$	$-69.3 \pm 11.7$
CSe dp4	$-34.1 \pm 0.28$	$-31.9 \pm 3.0$	$-52.1 \pm 9.8$	$-42.6 \pm 8.5$
CSe dp6	$-47.9 \pm n/a$	$-40.7 \pm n/a$	$-90.0 \pm 18.0$	$-62.8 \pm 9.9$
CSc dp4	$-33.7 \pm 8.1$	$-29.3 \pm 5.9$	$-63.7 \pm 20.3$	$-43.8 \pm 11.0$
CSc dp6	$-36.5 \pm 1.0$	$-30.8 \pm 12.8$	$-50.9 \pm 13.2$	$-40.3 \pm 9.4$

n/a: not available; only two values were obtained for this system.



**Fig. 3.** The representative clusters of GAG solutions (in sticks) obtained for H115APRIL/A105APRIL (in orange cartoon) by Autodock3 docking. On the top of the panel, all clusters are shown together, while on the bottom side of the panel they are shown separately per each GAG.

proper folding. Then, clustering (using DBSCAN algorithm) with a set of different parameters was used to obtain the representative structures (Figure 6). We have chosen the cluster that appeared in 10% of the frames of the 12  $\mu$ s MD simulation. The peptide was analyzed in terms of its structural parameters. DSSP algorithm (Kabsch and Sander 1983) was used to assess the secondary structure of the peptide and N-terminus of A105APRIL to compare their properties (Supplementary Figure S5). The most common elements in case of the peptide and two out of three N-termini from A105APRIL were the bend and the turn, while in case of the third N-terminus, a high number of 3<sub>10</sub>-helix was found. Afterward, using Autodock3, HP dp6 was docked to this 10 amino acid residue peptide. It was followed by 10 ns MD runs that were repeated 50 times (each individual run

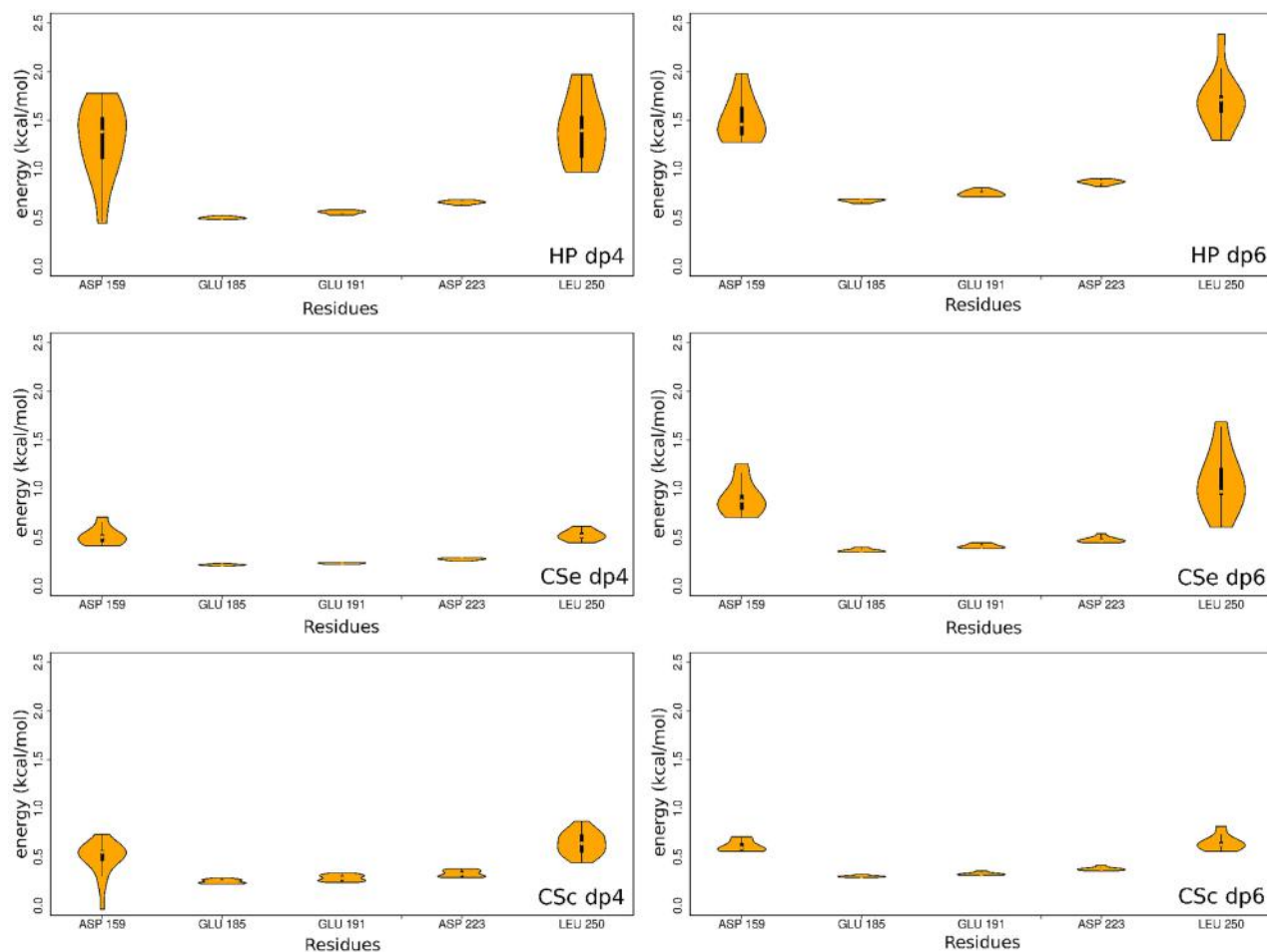
for each of the 50 best docking poses). Those runs were analyzed in terms of binding free energy, using both MM/GBSA and LIE approaches. Mean value of  $-29.5$  kcal/mol and  $-32.9$  kcal/mol for the 50 runs were obtained, respectively. Experiment was repeated using DMD method (as described in Methods and materials section) and was analyzed with MM/GBSA. Obtained binding free energy values were very close to those obtained after rigid docking to the most probable peptide conformer with Autodock3 ( $-30.0$  vs.  $29.5$  kcal/mol). When compared to the values of  $-141.1$  kcal/mol (MM/GBSA) and  $-69.3$  kcal/mol (LIE) in case of binding to APRIL, it is clear that the strength of the binding to the peptide representing N-terminus is far from that of binding to the protein. In fact, in one of the cases, we even observed start of the dissociation of HP from the peptide during the MD run. This may indicate that the N-terminus itself could not be sufficient for proper GAG binding and other peripheral/C-terminus residues in the protein (such as ARG 143, ARG 144, GLY 145, ARG 146 and LYS 249) are essential.

### No evidence for an interaction of TACI with HSPG in biological assays

We next assessed the binding of HSPG to TACI in ELISA by coating TACI-Fc. No binding of biotinylated HSPG ranging from 1 ng/mL up to 100  $\mu$ g/mL was detected (Figure 7A). In this experiment, coated TACI was biochemically active since it could bind to a complex of Fc-A105APRIL/HSPG. In the latter condition, presence of A105APRIL did not potentiate the interaction of TACI with HSPG since the binding recorded was not superior to the conditions with coated BCMA. When APRIL was coated, we still did not observe any potentiation of the TACI/HSPG interaction (Figure 7B). We next assessed the interaction with TACI expressed at the surface of the transfected cells. We again did not detect the binding of biotinylated HSPG at any concentrations tested (Figure 7C). Multiple myeloma cells such as the L363 cell line express the HSPG, syndecan 1. On these cells, A105APRIL could bind in an HSPG-dependent manner once inhibited by heparin (Figure 7D). We could not detect the binding of up to 30  $\mu$ g/mL of TACI-Fc to L363 cells.

### APRIL's receptor TACI and BCMA—their contribution to GAG binding

It was proposed that APRIL's receptor TACI was able to bind to HSPG while BCMA could not (Bischof et al. 2006). Those findings are in contrary to our experiments and to the data published recently by Kowalczyk-Quintas et al. (2019). To further analyze those contra-



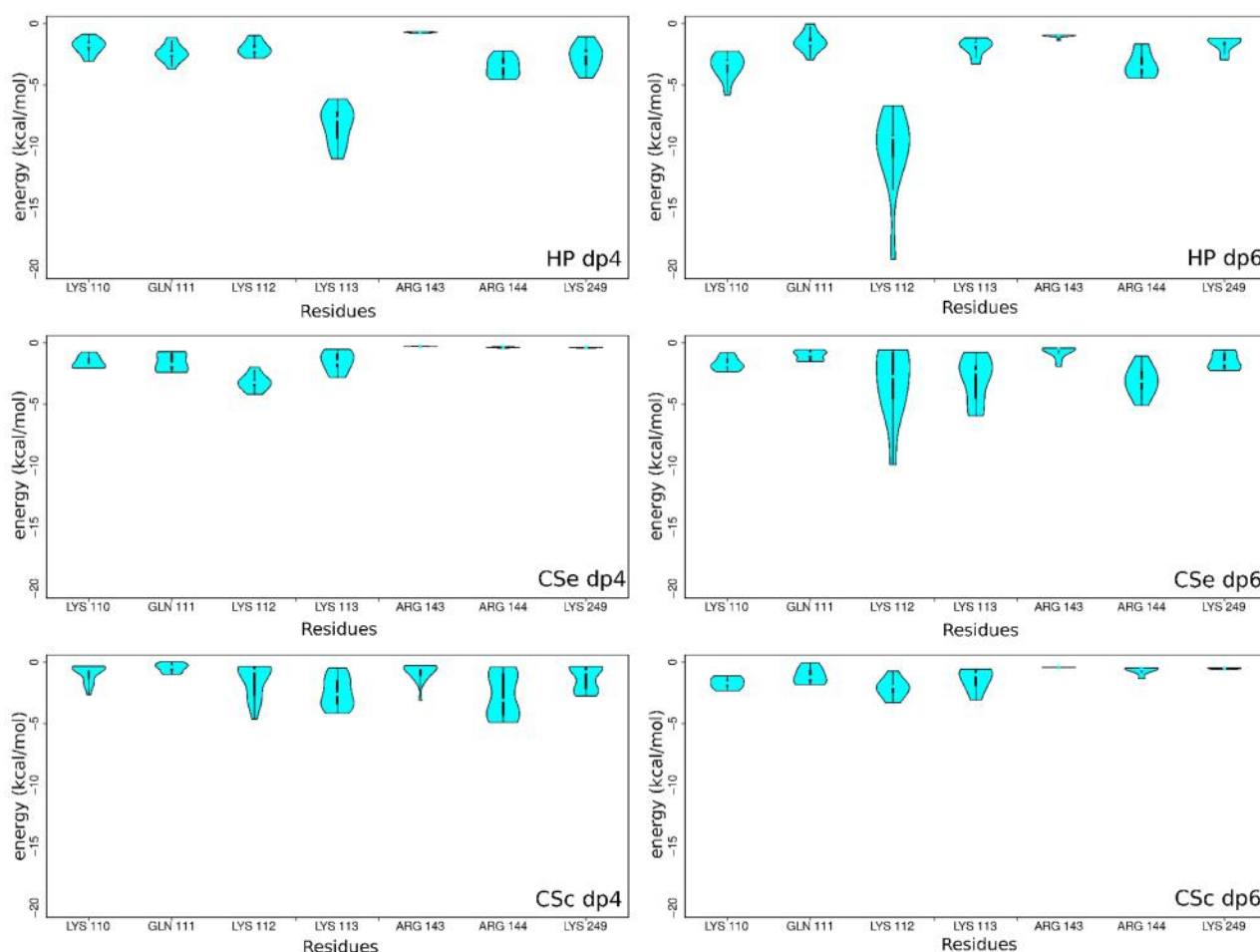
**Fig. 4.** Per-residue free energy decomposition of the residues that show the most unfavorable GAG-binding impacts in case of the  $A_{105}$ APRIL variant. Violin plots represent the energy distribution for particular residues binding different GAGs from all MD simulations with the starting structures obtained by Autodock3 docking.

dictory findings and to understand TACI–GAGs interactions at the molecular level, we first performed electrostatic potential analysis using the PBSA method from AMBER suite. APRIL protein with soluble fragments of TACI or BCMA in complexes (PDB IDs 1XU1 and 1XU2, respectively) was used for this analysis. The obtained data indicate more favorable electrostatic potential in case of complex of APRIL–TACI than in APRIL–BCMA (Supplementary Figure S6). Then, we docked HP dp6 to both TACI- and BCMA-soluble fragments using Autodock3. Obtained binding poses were clustered (Methods and materials section), and poses from the best clusters for TACI (eight poses) and BCMA (10 poses) were used in the MD simulations. Afterward, the MM/GBSA method was applied to assess free binding energies in TACI–HP dp6 and BCMA–HP dp6 complexes. MM/GBSA analysis showed free binding energy values of  $-35.4$  kcal/mol (standard deviation [SD] 7.9) and  $-15.5$  kcal/mol (SD 8.7) for TACI and BCMA complexes, respectively. Those values clearly indicate the distinguishable strength of potential GAG binding to two different APRIL's receptors. Free binding energy of  $-15.5$  kcal/mol suggests that HP binding to BCMA is unlikely.  $-35.4$  kcal/mol value from the MM/GBSA analysis in case of the TACI–HP complex could suggest potential weak binding. LIE analysis has shown the same mean value of  $-35.4$  kcal/mol (SD 6.8). However, the value of  $-35.4$  kcal/mol indicates some uncertainty,

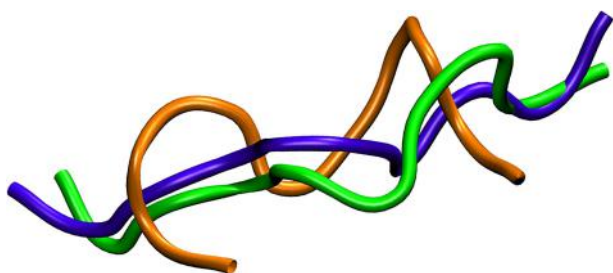
and we are unable to clearly and confidently claim for this particular complex whether there is a binding or not. It is also worth mentioning that HP is the most sulfated form of HS and thus the most charged variant of HS, which consists of mixture of GAGs with different sulfation patterns. Taking this into account, the HP may show slightly better binding properties in this particular complex than the rest of the HS family. Therefore, the borderline free binding energy values for the HP binding to TACI could suggest that potentially weaker binding of HS to TACI would not be sufficient enough for an effective binding that could be observed in the experiment.

#### DMD in GAG-binding site prediction

In order to further analyze and understand the predictive power of our docking results from Autodock3 and to provide more details on GAGs binding to  $A_{105}$ APRIL, we performed DMD for HP dp4 and 6. DMD should allow for more flexibility for the molecules in the docking process, which could suggest that DMD values may be more reliable. Another advantage of DMD is that it allows for the inclusion of explicit solvent in local molecular docking. Previously, it was shown that in comparison to AD3, the DMD method was capable to reliably identify the receptor residues contributing most to binding and it had higher complex structure prediction performance,



**Fig. 5.** Per-residue free energy decomposition of the residues that show the most favorable GAG-binding impacts in case of  $A_{105}$ APRIL variant. Violin plots represent the energy distribution for particular residues binding different GAGs from all MD simulations with the starting structures obtained by Autodock3 docking.



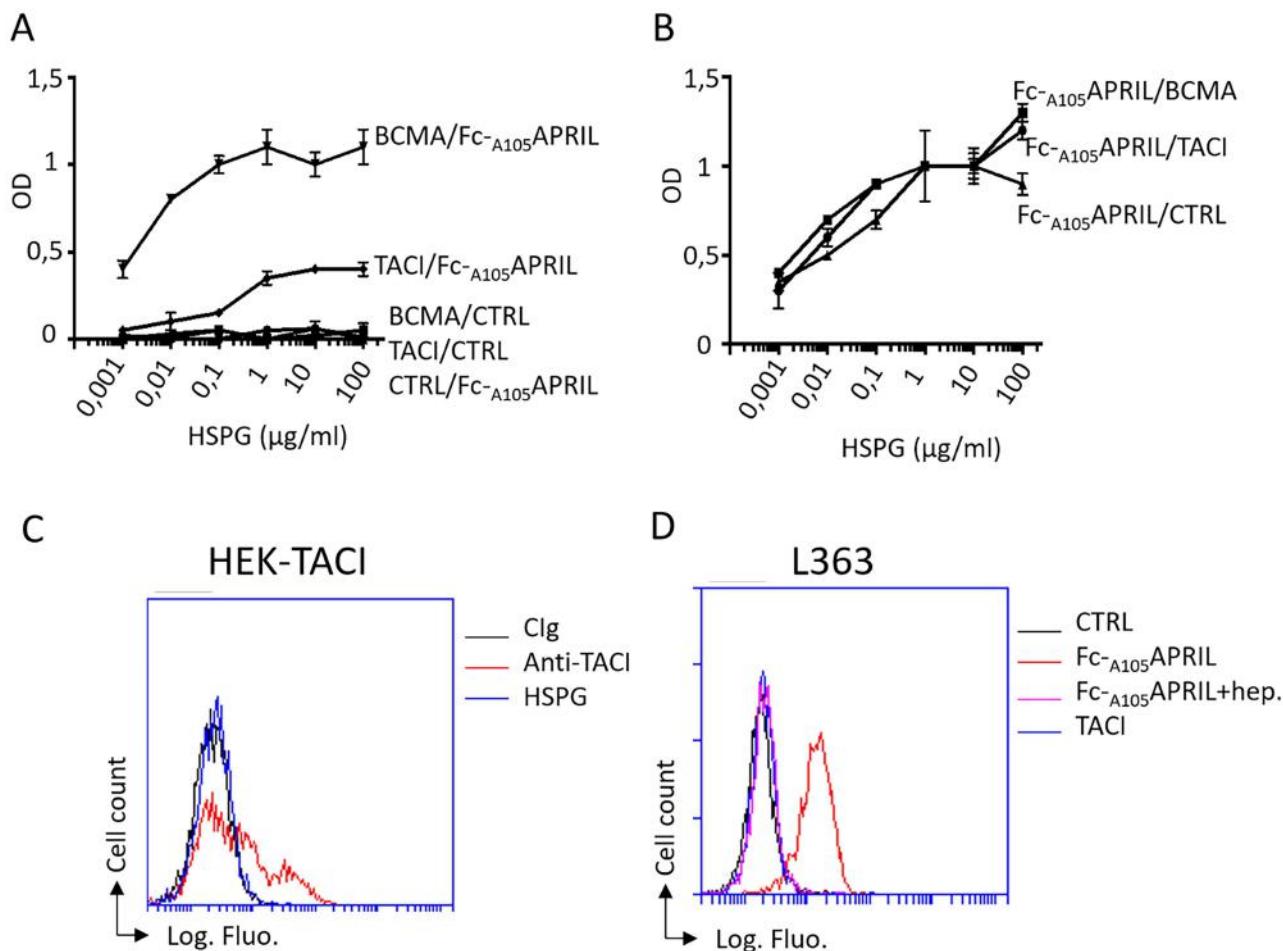
**Fig. 6.** Three representative structures of the N-terminal fragment models (aa 105–114) from the  $A_{105}$ APRIL protein obtained from the 12  $\mu$ s MD simulation. The DBSCAN algorithm was used for clustering with following parameters: orange—epsilon = 2 and minpoints = 3, green—epsilon = 2 and minpoints = 5, violet (chosen for the further analysis)—epsilon = 3 and minpoints = 4.

especially for systems with highly flexible and negatively charged ligands like GAGs (Gehrcke and Pisabarro 2015; Salbach-Hirsch et al. 2015; Babik et al. 2017) 100 independent docking runs were obtained for each GAG. MM/GBSA binding energy analysis showed an average energy of  $-84.8$  kJ/mol for the HP dp4 and  $-93.1$  kJ/mol for HP dp6. Therefore, it is shown once more that dp6 GAGs bind

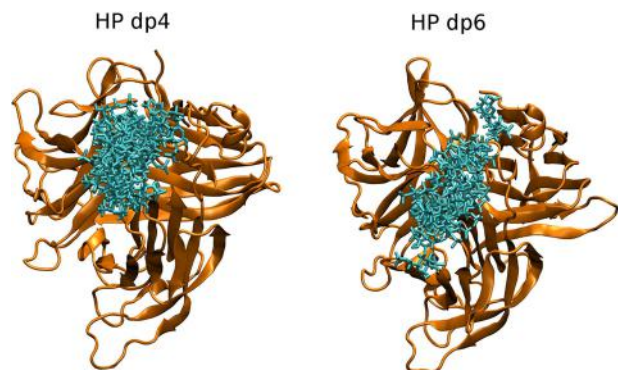
stronger to APRIL protein than dp4 GAGs. Obtained docking poses are similar to those docked with Autodock3 (Figure 8). However, MM/GBSA applied to DMD docking poses yields lower binding energy values than the ones obtained from MD simulations, starting from the binding poses produced by Autodock3. This points out the importance of taking into account the flexibility of the receptor for such calculations. The potentially unstructured nature of  $A_{105}$ APRIL N-termini corroborates the fact that up to now no structural data on the N-terminus of the protein were obtained by X-ray/NMR/CryoEM experiments. In our simulations, we observed high mobility of the N-termini of the  $A_{105}$ APRIL trimer both at coarse-grained and in all-atom levels.

### Multiple GAG binding to APRIL protein

DMD method was adopted to check whether binding of multiple GAG fragments of different types is possible. To our knowledge, this is a novel approach, and no one has docked multiple GAG fragments to any protein yet. We used DMD to dock CSe dp6 to the full-length APRIL with already harbored HP dp6 (Supplementary Figure S7). The free binding energy analysis shows much weaker CSe binding to APRIL in the presence of HP in comparison to its absence. The mean value of 100 analyzed runs with MM/GBSA approach was



**Fig. 7.** No evidence for a direct interaction of TACI with HSPG. The binding of increasing concentrations of biotinylated HSPG to TACI was tested in ELISA. (A) APRIL receptors TACI and BCMA were coated, and the interaction was tested in the presence or absence of 1  $\mu\text{g}/\text{mL}$  of soluble  $\text{Fc}_{\text{A105}}\text{APRIL}$ . (B)  $\text{Fc}_{\text{A105}}\text{APRIL}$  was coated and the binding was tested in the presence or absence of 1  $\mu\text{g}/\text{mL}$  of the soluble receptors, TACI and BCMA. Results are presented as in Figure 1A. Three independent experiments were performed. (C) Binding of biotinylated HSPG ranging from 1 ng/mL to 10  $\mu\text{g}/\text{mL}$  was assessed on HEK cells transiently transfected with full length human TACI. Results for only the highest HSPG concentration is shown. Reactivity with an anti-TACI is also shown (Clg = isotype-matched control immunoglobulin). (D) Binding of TACI-Fc (from 1 ng/mL to 30  $\mu\text{g}/\text{mL}$ ) was tested on L363 cells. Results from only the highest concentration of TACI are shown. HSPG expression on L363 cells is shown by the binding of  $\text{Fc}_{\text{A105}}\text{APRIL}$  (1  $\mu\text{g}/\text{mL}$ ) inhibited by heparin (hep.). Overlaid histograms plots are representative of at least two independent experiments. Thy-1-Fc was used as control (CTRL) Fc-fused molecule.



**Fig. 8.** Cluster representing the best docked structures for the HP dp4 and dp6 (in cyan sticks) after DMD to  $\text{A}_{105}\text{APRIL}$  variant (in orange cartoon).

–27.2 kcal/mol compared to –90 kcal/mol found in the regular MD run analysis with MM/GBSA. LIE calculations also showed drop in the binding strength represented by –29.7 kcal/mol compared to

–62.8 kcal/mol to what we saw in the absence of the prebound HP. When visually analyzing the runs, we observed the dissociation of several CSe monomeric units in two of the runs (out of 100). In those two cases, the MM/GBSA binding analysis yielded values of –5.4 kcal/mol and –5.6 kcal/mol, while the LIE approach claimed –7.5 kcal/mol and –9.2 kcal/mol, respectively. Therefore, it is clear that CSe dp6 binding is much weaker in the presence of HP. When such a multiple GAG binding may not be favorable, it is still feasible. We performed energy decomposition to further analyze those relations and to check for the residues that hinder the interaction of CSe dp6 and the preformed complex with HP the most. It was not surprising that among the first six residues that repelled CSe the most belonged to HP. The other residues that negatively affected CSe–complex binding were the same that were shown in the absence of HP: ASP 159, GLU 185, GLU 191, ASP 223 and LEU 250. Interaction-favoring residues were found to be similar to those from the earlier analysis without HP. However, we found two new additional amino acids that are essential for the strengthening of the interactions, VAL 106 and LEU 107, while LYS 112, ARG 146 and LYS 249 have been shown to be still important.



**Table II.** MM/GBSA and LIE binding free energy analysis for the truncated APRIL variant in complex with HP dp4 and CSc dp4

	HP dp4, $\Delta G$ (kcal/mol)		CSc dp4, $\Delta G$ (kcal/mol)	
	MM/GBSA	LIE	MM/GBSA	LIE
First ns	$-66.7 \pm 12.6$	$-45.8 \pm 5.6$	$-43.9 \pm 15.5$	$-36.9 \pm 13.6$
Last ns	$-57.1 \pm 12.5$	$-37.5 \pm 5.7$	$-31.4 \pm 10.1$	$-28.5 \pm 9.5$
First 10 ns	$-58.4 \pm 13.1$	$-41.0 \pm 7.2$	$-36.4 \pm 12.1$	$-31.5 \pm 6.3$
Last 10 ns	$-56.6 \pm 13.4$	$-36.9 \pm 8.1$	$-32.7 \pm 8.8$	$-28.4 \pm 6.9$
First 50 ns	$-56.5 \pm 9.4$	$-38.5 \pm 6.2$	$-33.8 \pm 9.4$	$-29.7 \pm 5.9$
Last 50 ns	$-56.4 \pm 9.7$	$-37.6 \pm 6.6$	$-33.3 \pm 8.3$	$-28.8 \pm 6.9$
Whole simulation	$-56.4 \pm 8.3$	$-38.1 \pm 5.8$	$-33.7 \pm 8.1$	$-29.3 \pm 5.9$

### Methodological aspects of binding free energy analysis by MM/GBSA and LIE

For the verification of our results and for the evaluation of the methodology used in our study, we performed extensive investigation regarding the statistical relevance of the data from free energy calculations. Beside such a verification, this analysis was aimed to understand our particular molecular system more comprehensively but also to gain insights into the nature of protein–GAGs interactions in general.

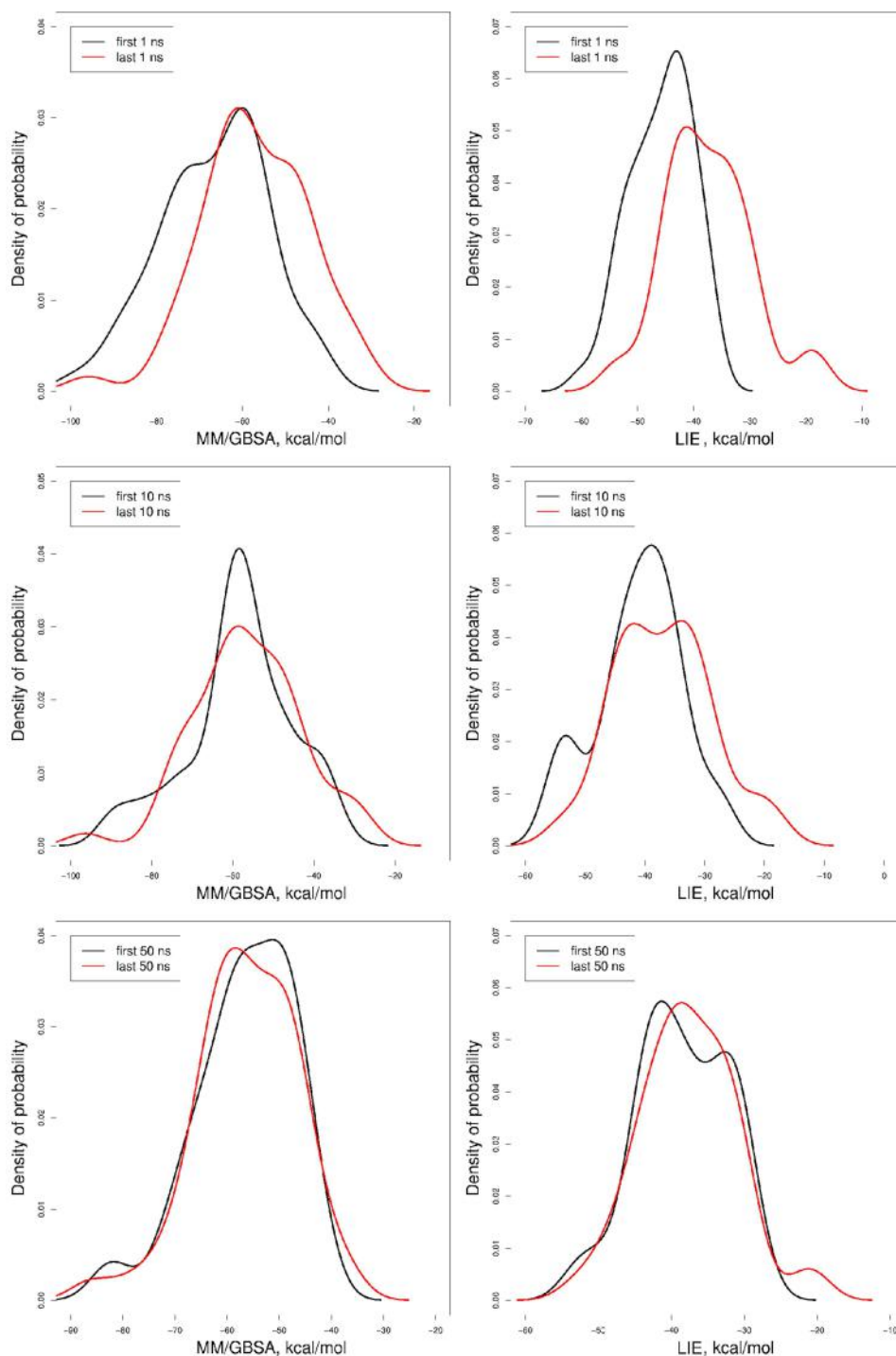
Binding free energy calculations is one of the most important and still challenging parts of computational analysis of protein–ligand interactions (Kollman et al. 2000; Genheden and Ryde 2015). There are several ways to assess binding free energy. The most commonly used ones are molecular mechanics Poisson–Boltzmann surface area (MM/PBSA) and MM/GBSA, where the evaluation of the free energy is described as a sum of in vacuo molecular mechanics energy terms and a solvation free energy term in implicit solvent (Kollman et al. 2000). The obtained MM/PBSA (MM/GBSA) binding energies should be rather understood as enthalpies with the entropy of the solvent implicitly accounted for than the full free binding energies (Genheden and Ryde 2015). Another approach for free energy calculations is LIE, which is in general less accurate but computationally less expensive and is based on VdW and electrostatic energies linear combination obtained directly from each frame of the MD trajectory. In LIE, electrostatics are calculated in vacuo and are simply scaled by a dielectric constant. However, it is hard to evaluate how precise and reliable are the mentioned methods when applied on molecular docking simulations output, in particular, in case of the highly charged systems as ours. In this study, we used both methods in more high-throughput manner to investigate our model comprehensively.

For this methodological part of the study, we have chosen H115APRIL—a truncated version of the protein—as it lacks the first 10 amino acid residues that seem to be unstructured. Those residues could negatively affect our methodological analysis and contribute to the noise in our data, which in such a high-throughput and technique-oriented analysis, we would avoid at all cost. 50 top-scored poses obtained for each of CSc4 and HP dp4 GAGs docked to the truncated variant of APRIL protein with Autodock3 program were analyzed in terms of the binding energy obtained from 100 ns MD simulations by both the MM/GBSA and LIE approaches. First, the differences between the first and the last 1, 10 and 50 ns were compared for MM/GBSA (Table II, Figure 9). For most of the CSc4 dp4 binding poses, energies were higher at the end of the MD runs than at the beginning (Table II). This is probably due to the fact that molecular docking predicts interactions between ligand and protein without taking into account the solvent explicitly which is supposed

to be very important for the protein–GAG interactions (Teyra et al. 2006; Samsonov et al. 2011; Samsonov et al. 2014).

In the course of MD simulations, the interface between the receptor and the ligand is filled with the molecules of water, which in general, weakens the interactions in the complex. This result may look counterintuitive since it is rather expected that the MD simulation would correct the initial structure and, therefore, decrease the total energy of binding. However, the solvent impact is not considered explicitly in the free energy calculations, while the penetration of the water molecules into the protein–GAG interface leads to the generally less favorable MM/GBSA energies of binding in comparison to the case when no water molecules are in the complex interface. In case of CSc dp4, the density of probability maxima of the binding energy spectrum at the last part of the runs were sharper, which suggests tending toward the same energy values in the course of the MD run due to the convergence of the various MD simulations starting from different initial conformations (Supplementary Figure S8). It shows that the longer the run, the more reliable results are, which is, in general, expected. However, it is worth mentioning that extending the length of the MD runs over a certain point is highly cost-inefficient and yields from insignificant to no improvements. Our data clearly show how the uncertainties of the simulation result in such an electrostatically driven system change with the elongation of the simulation. In HP dp4 spectra, there were no differences observed in terms of the width of the peaks (Figure 9). Similarly to CSc dp4, for the HP dp4, there was an increase in binding energy during simulation. Almost no differences in the case of binding energy comparison after the first and second half of the runs for both CSc dp4 and HP dp4 were observed (Figure 9, Table II and Supplementary Figure S8). Moreover, no significant differences were found when the binding energies were compared for the first and the last 10 ns of the run for both CSc and HP. On the other hand, differences were significant and rather high when the first and the last ns of the run were compared. This suggests that the length of the simulation of 10 ns could be enough for converging results from the 50 starting poses obtained by docking for this system. It was also shown that the binding energies of the 50 docking poses differed essentially (Table II and Figure 10). Fortunately, our clustering protocol and the parameter choice for DBSCAN algorithm yielded poses spread normally across the range of energy binding values (Figure 10). This supports the idea that the obtained clusters of structures from 50 AD3 top-scored binding poses represent a structural ensemble properly.

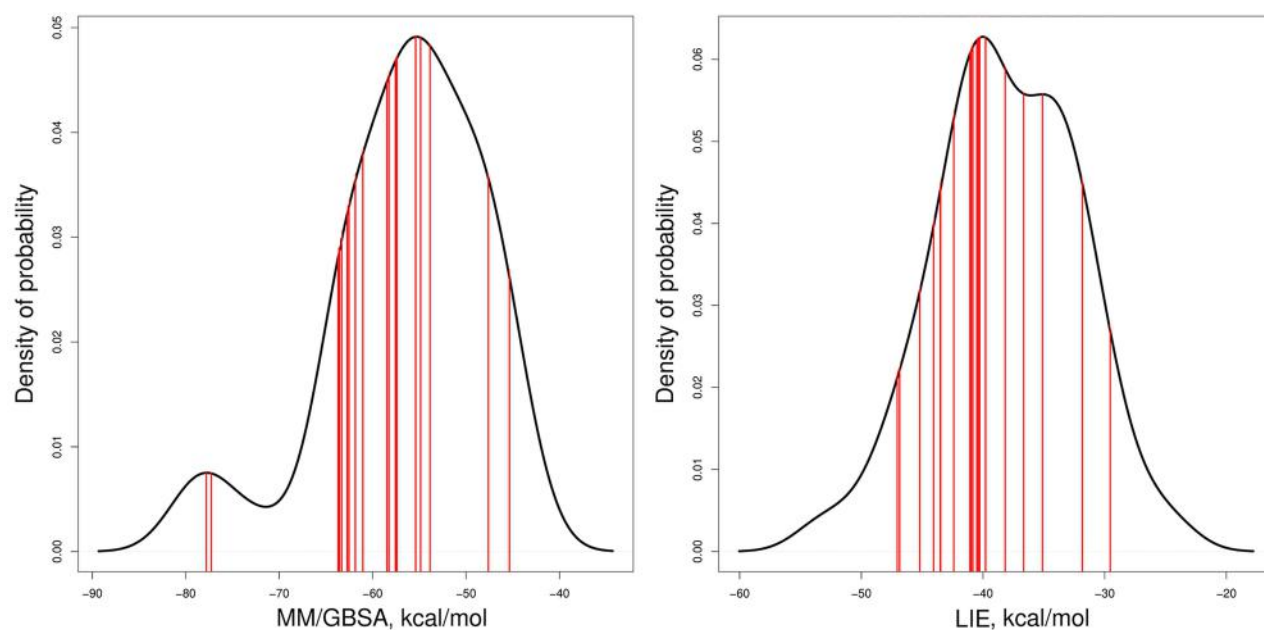
EGB and ESURF are the components of the MM/GBSA free binding energies describing the impact of the solvent in the molecular interactions. We analyzed their effect on the final outcome of the free energy calculations. When comparing the first and the second



**Fig. 9.** MM/GBSA and LIE binding free energy analysis of HP dp4 to H<sub>115</sub>APRIL variant showing differences in total free energies between the first and the last ns (top), the first and the last 10 ns (middle), and the first and the second halves of the MD run (bottom).

halves of the runs—for both CSc dp4 and HP dp4—we detected almost no difference in the shapes of the EGB values distribution (Supplementary Figures S9 and S10). The same applied to the ESURF density of the probability curves (Supplementary Figures S9 and S10). There was a slight curve shift toward the lower values in the last 10 ns when compared to the first 10 ns for CSc dp4 and HP dp4 (Supplementary Figures S9 and S10). For the ESURF component, the

curves were slightly moved toward higher values—again for both CSc dp4 and HP dp4. Looking at the first and the last ns of the runs, trends remained the same in terms of the difference of the curve's shift. There was a much higher curves' shift for the first and the last ns of the runs for CSc dp4 and HP dp4 (Supplementary Figures S9 and S10). Those findings are in agreement with the results for other MM/GBSA components, showing that 1 ns of MD run is not sufficient to yield



**Fig. 10.** Total binding free energy distribution analysis for the complex between  $H_{115}$ APRIL variant and HP dp4 from 50 independent MD runs. Red vertical lines correspond to binding free energies from particular MD simulations representing the structures belonging to the obtained clusters.

reliable results in terms of free energy of binding by the MM/GBSA approach for this particular molecular system.

LIE analysis yielded qualitatively similar results (Figure 9, Supplementary Figure S8 and Table II). Despite in general lower values of the obtained binding energies, they followed the same trend as in the case of the MM/GBSA analysis. Similarly, LIE showed no significant differences between the first and the last 10 or 50 ns of the MD simulation. LIE method resulted in narrower energy distribution across different binding poses, however, the results themselves look less distinctive and show smaller differences between certain types of GAGs (Supplementary Figures S9 and S10). These observations suggest that MM/GBSA could be considered as a more appropriate method to calculate the binding free energies in protein–GAG systems, which is similar to the data obtained for other biomolecular complexes (Tsui and Case 2000; Genheden and Ryde 2011).

Similarly to the MM/GBSA free energy components, we analyzed both ELE and VdW components of LIE in the course of the MD simulations (Supplementary Figures S9 and S10). For both components, when compared, the first and the second halves of the MD simulations curves looked almost identical in the case of both analyzed GAGs. In comparison of the first and the last 10 ns of MD runs, we observed only small shifts in the distribution curves' shapes. During the run, both VdW and ELE components shifted toward higher values with the exception that the ELE component moved only slightly toward higher values. This applied to both CSc dp4 and HP dp4. Likewise, in the MM/GBSA analysis, the differences were more distinctive when looking at the first and the last ns of the MD simulations. The highest shift was observed in the VdW component for both CSc dp4 and HP dp4.

The conclusion drawn from these free energy calculations is that 10 ns simulation in case of APRIL–GAG or similar systems in terms of size and interaction patterns is sufficient for the analysis of MD trajectories, especially when working on the bigger data sets of 50 repeated MD runs, like it was done in this study. Working only on around 10 MD runs with the length of 10 ns may not always be

enough to get a full insight of the protein–ligand interactions. The gains between 10 ns and 50 ns analyses were little, and thus, this would not be computationally effective to elongate the simulation up to 50 ns. Therefore, we suggest that it is better to enlarge the number of MD runs rather than elongate the runs. Similar statements were previously reported in other study (Genheden and Ryde 2015). The improvements were even more negligible after 50th ns of the MD runs when considering free energy analysis, and further elongation of the simulation is computationally expensive but would not improve the accuracy of the results. It is worth to notice that those insights cannot be extrapolated for all types of molecular systems. Minimal sufficient length of the MD simulation may depend on many variables, such as the system's charge, size of the ligand, flexibility of both ligand and receptor, geometry of the receptor-binding region, dominant interactions' types in the site of binding and most importantly on the goal of the study itself.

## Conclusions

We present novel data which indicate that GAGs may also bind to the truncated version of the APRIL, but this binding is definitely less potent than the one to the full-length APRIL. In this study, it was shown that heparin binds the strongest to both variants of APRIL. This binding was confirmed both by computational and wet lab (ELISA/flow cytometry) experiments. CS-4,6-sulfate manifest weaker binding than heparin, while chondroitin-6-sulfate shows very weak-to-no binding properties. All of the mentioned GAGs when bound were located near the N-termini of the monomers that are spatially close in the APRIL trimer and form one common binding site. This region was analyzed with PBSA method and positive electrostatic potential was revealed near the N- and C-termini. Not surprisingly, in energy decomposition analysis, we also found most of the residues that significantly affected binding of the GAGs in this region. However, our study on heparin binding to peptide consisting of the first 10 amino acid residues from

the N-terminus of the APRIL protein indicates that GAG binding to peptide itself is not as strong as the binding to the full protein. The latter represents another argument in favor of a contribution of other peripheral and C-terminus residues in GAG binding. Structurally, the peptide lacked essential propensities for secondary structure features, and in this term, was similar to two out of three N-termini from the model of A<sub>105</sub>APRIL. In this study, we further assessed the multiple GAG binding to single APRIL. Our findings show that it could be possible, but the strength of the binding is much weaker than in case for a single GAG molecule.

The biological assays performed in the present study showed no binding of HSPG to TACI. This is in agreement with the recent findings of Kowalczyk-Quintas et al. (2019) who claimed no binding of TACI to HP, which is the most sulfated compound in the HS family, unless oligomerized. Our computational analysis yielded borderline energy values for TACI–HP binding that suggests no binding or very weak one. Energies obtained for the BCMA–HP complex were not favorable enough and thus indicate no binding in this case.

To analyze our data, we used two methods of free binding energy assessment: MM/GBSA and LIE. While the MM/GBSA yielded more statistically reliable results, the LIE method is much faster and requires very little computational resources. It is also worth mentioning that for LIE to be used effectively, it is best that it would be optimized in advance using experimental data rather than default parameters. We also applied DMD and confirmed our results from docking with Autodock3. This novel GAG-specific docking approach allows for the inclusion of explicit solvent in local molecular docking and for molecular flexibility during docking process. To sum up, we rigorously evaluated the currently used MD-based methodology for protein–GAG complexes' theoretical description. We believe that our APRIL–GAG interactions analysis will be useful for further studies on the APRIL-related molecular mechanisms and the insights we shared in this work will help to facilitate drug development for autoimmune diseases and B-cell malignancies.

## Supplementary data

Supplementary data for this article are available online at <http://glycob.oxfordjournals.org/>.

## Acknowledgements

We would like to thank Pascal Schneider (University of Lausanne) for providing APRIL constructs. Computational resources were provided by the Polish Grid Infrastructure (PL-GRID, grants gagstr1, gpuaprilgags, aprilgagsinteractions, plgaprilgag2), ZIH at TU Dresden (grant p\_gag) and our cluster at the Faculty of Chemistry, University of Gdansk.

## Funding

National Science Centre of Poland (UMO-2018/30/E/ST4/00037).

## Conflict of interest statement

None declared.

## Abbreviations

APRIL, A proliferation-inducing ligand; BCMA, B-cell maturation antigen; CS, Chondroitin sulfate; DMD, Dynamic molecular docking; ELE, Electrostatic component of MM/GBSA, MM/PBSA, LIE free

binding energies; ELISA, Enzyme-linked immunosorbent assay; FGF, Fibroblast growth factor; GAG, Glycosaminoglycan; IgG, Immunoglobulin G; IgG2a, Immunoglobulin G2a; HP, Heparin; HS, Heparan sulfate; HSPG, Heparan sulfate proteoglycan; LIE, Linear interaction energy; MD, Molecular dynamics; MM/GBSA, Molecular mechanics generalized Born surface area; MM/PBSA, Molecular mechanics Poisson–Boltzmann surface area; PBSA, Poisson–Boltzmann Surface Area; TACI, Transmembrane activator and CAML interactor; TGF- $\beta$ 1, Transforming growth factor beta one; TNF, Tumor necrosis factor; VdW, Van der Waals; VEGF, Vascular endothelial growth factor; WHAM, Weighted histogram analysis method.

## References

- Adcock SA, McCammon JA. 2006. Molecular dynamics: Survey of methods for simulating the activity of proteins. *Chem Rev.* 106:1589–1615.
- Ahmed MC, Huard B. 2021. Inhibition of chondroitin sulfate proteoglycans by APRIL. In: *Methods in Molecular Biology*. Springer, Berlin. Vol. 2248. p. 43–61.
- Babik S, Samsonov SA, Pisabarro MT. 2017. Computational drill down on FGF1-heparin interactions through methodological evaluation. *Glycoconj J.* Springer, Berlin. 34:427–440.
- Baert L, Manfroi B, Casez O, Sturm N, Huard B. 2018. The role of APRIL—A proliferation inducing ligand—in autoimmune diseases and expectations from its targeting. *J Autoimmun.* 95:179–190.
- Baert L, Benkhoucha M, Popa N, Ahmed MC, Manfroi B, Boutonnat J, Sturm N, Raguenez G, Tessier M, Casez O et al. 2019. A proliferation-inducing ligand-mediated anti-inflammatory response of astrocytes in multiple sclerosis. *Ann Neurol.* 85:406–420.
- Bischof D, Elsawa SF, Mantchev G, Yoon J, Michels GE, Nilson A, Sutor SL, Platt JL, Ansell SM, von Bulow et al. 2006. Selective activation of TACI by syndecan-2. *Blood.* 107:3235–3242.
- Bojarski KK, Sieradzian AK, Samsonov SA. 2019. Molecular dynamics insights into protein-glycosaminoglycan systems from microsecond-scale simulations. *Biopolymers.* 110:23252.
- Bossen C, Cachero TG, Tardivel A, Ingold K, Willen L, Dobles M, Scott ML, Maquelin A, Belnoue E, Siegrist CA et al. 2008. TACI, unlike BAFF-R, is solely activated by oligomeric BAFF and APRIL to support survival of activated B cells and plasmablasts. *Blood.* 111:1004–1012.
- Crijns H, Vanheule V, Proost P. 2020. Targeting chemokine—Glycosaminoglycan interactions to inhibit inflammation. *Front Immunol.* 11:483.
- Czaplewski C, Kalinowski S, Liwo A, Scheraga HA. 2009. Application of multiplexed replica exchange molecular dynamics to the unres force field: Tests with  $\alpha$  and  $\alpha+\beta$  proteins. *J Chem Theory Comput.* 5:627–640.
- Case DA, Ben-Shalom IY, Brozell SR, Cerutti DS, Cheatham III TE, Cruzeiro VDW, Darden TA, Duke RE, Ghoreishi D, Gilson MK et al. 2018. *AMBER16*. San Francisco (CA): University of California.
- Derler R, Gesslbauer B, Weber C, Strutzmann E, Miller I, Kungl A. 2017. Glycosaminoglycan-mediated downstream signaling of CXCL8 binding to endothelial cells. *Int J Mol Sci.* 18:2605.
- Digabriele AD, Lax I, Chen DI, Svahn CM, Jaye M, Schlessinger J, Hendrickson WA. 1998. Structure of a heparin-linked biologically active dimer of fibroblast growth factor. *Nature.* 393:812–817.
- Ester M, Kriegl HP, Sander J, Xu X. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. *AAAI.* 226–231.
- Faham S, Hileman RE, Fromm JR, Linhardt RJ, Rees DC. 1996. Heparin structure and interactions with basic fibroblast growth factor. *Science.* 271:1116–1120.
- Gandhi NS, Mancera RL. 2011. Molecular dynamics simulations of CXCL8 and its interactions with a receptor peptide, heparin fragments, and sulfated linked cyclitols. *J Chem Inf Model.* 51:335–358.
- Gehrcke JP, Pisabarro MT. 2015. Identification and characterization of a glycosaminoglycan binding site on interleukin-10 via molecular simulation methods. *J Mol Graph Model.* 62:97–104.

- Genheden S, Ryde U. 2011. A comparison of different initialization protocols to obtain statistically independent molecular dynamics simulations. *J Comput Chem.* 32:187–195.
- Genheden S, Ryde U. 2015. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin Drug Discovery.* 10:449–461.
- Habuchi H, Habuchi O, Kimata K. 2004. Sulfation pattern in glycosaminoglycan: Does it have a code? *Glycoconj J.* 21:47–52.
- Hahne M, Kataoka T, Schröter M, Hofmann K, Irmeler M, Bodmer JL, Schneider P, Bornand T, Holler N, French LE *et al.* 1998. APRIL, a new ligand of the tumor necrosis factor family, stimulates tumor cell growth. *J Exp Med.* 188:1185–1189.
- Hendriks J, Planelles L, de Jong-Odding J, Hardenberg G, Pals ST, Hahne M, Spaargaren M, Medema JP. 2005. Heparan sulfate proteoglycan binding promotes APRIL-induced tumor cell proliferation. *Cell Death Differ.* 12:637–648.
- Huige CJM, Altona C. 1995. Force field parameters for sulfates 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.
- Hymowitz SG, Patel DR, Wellweber HJA, Runyon S, Yan M, Yin JP, Shriver SK, Gordon NC, Pan B, Skelton NJ *et al.* 2005. Structures of APRIL-receptor complexes: Like BCMA, TACI employs only a single cysteine-rich domain for high affinity ligand binding. *J Biol Chem.* 280:7218–7227.
- Imberty A, Lortat-Jacob H, Pérez S. 2007. Structural view of glycosaminoglycan-protein interactions. *Carbohydr Res.* 342:430–439.
- Ingold K, Zumsteg A, Tardivel A, Huard B, Steiner QG, Cachero TG, Qiang F, Gorelik L, Kalled SL, Acha-Orbea H *et al.* 2005. Identification of proteoglycans as the APRIL-specific binding partners. *J Exp Med.* 201:1375–1383.
- Jarzynski C. 1997. Nonequilibrium equality for free energy differences. *Phys Rev Lett.* 78:2690.
- Joseph PRB, Mosier PD, Desai UR, Rajarathnam K. 2015. Solution NMR characterization of chemokine CXCL8/IL-8 monomer and dimer binding to glycosaminoglycans: Structural plasticity mediates differential binding interactions. *Biochem J.* 47:121–133.
- Kabsch W, Sander C. 1983. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers.* 22:2577–2637.
- Kimberley FC, Van Bostelen L, Cameron K, Hardenberg G, Marquart JA, Hahne M, Medema JP. 2009. The proteoglycan (heparan sulfate proteoglycan) binding domain of APRIL serves as a platform for ligand multimerization and cross-linking. *FASEB J.* 23:1584–1595.
- Kirschner KN, Yongye AB, Tschampel SM, González-Outeiriño J, Daniels CR, Foley BL, Woods RJ. 2008. GLYCAM06: A generalizable biomolecular force field. *Carbohydrates.* *J Comput Chem.* 29:622–655.
- Koehler L, Samsonov S, Rother S, Vogel S, Köhling S, Moeller S, Schnabelrauch M, Rademann J, Hempel U, Pisabarro TM *et al.* 2017. Sulfated hyaluronan derivatives modulate TGF- $\beta$ 1:receptor complex formation: Possible consequences for TGF- $\beta$ 1 signaling. *Sci Rep.* 7:1210.
- Kokenyesi R, Bernfield M. 1994. Core protein structure and sequence determine the site and presence of heparan sulfate and chondroitin sulfate on syndecan-1. *J Biol Chem.* 16:12304–12309.
- Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W *et al.* 2000. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc Chem Res.* 33:889–897.
- Kowalczyk-Quintas C, Willen D, Willen L, Golob M, Schuepbach-Mallepell S, Peter B, Eslami M, Vigolo M, Broly H, Samy E *et al.* 2019. No interactions between heparin and atacept, an antagonist of B cell survival cytokines. *Br J Pharmacol.* 176:4019–4033.
- Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA. 1992. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J Comput Chem.* 13:1011–1021.
- Künze G, Köhling S, Vogel A, Rademann J, Huster D. 2016. Identification of the glycosaminoglycan binding site of interleukin-10 by NMR spectroscopy. *J Biol Chem.* 291:3100–3113.
- Künze G, Gehrcke JP, Pisabarro MT, Huster D. 2014. NMR characterization of the binding properties and conformation of glycosaminoglycans interacting with interleukin-10. *Glycobiology.* 24:1036–1049.
- Larsen CG, Anderson AO, Appella E, Oppenheim JJ, Matsushima K. 1989. The neutrophil-activating protein (NAP-1) is also chemotactic for T lymphocytes. *Science.* 243:1464–1466.
- Ling RF, Späth H. 1981. Cluster analysis algorithms for data reduction and classification of objects. *Dent Tech.* 23:417–418.
- Liwo A, Oldziej S, Pincus MR, Wawak RJ, Rackovsky S, Scheraga HA. 1997. A united-residue force field for off-lattice protein-structure simulations. I. Functional forms and parameters of long-range side-chain interaction potentials from protein crystal data. *J Comput Chem.* 18:849–873.
- López-Fraga M, Fernández R, Albar JP, Hahne M. 2001. Biologically active APRIL is secreted following intracellular processing in the Golgi apparatus by furin convertase. *EMBO Rep.* 2:945–951.
- Luster AD. 1998. Chemokines—Chemotactic cytokines that mediate inflammation. *N Engl J Med.* 338:436–445.
- Lyon M, Rushon G, Gallagher JT. 1997. The interaction of the transforming growth factor- $\beta$ s with heparin/heparan sulfate is isoform-specific. *J Biol Chem.* 272:18000–18006.
- Mason IJ. 1994. The ins and outs of fibroblast growth factors. *Cell.* 4:547–552.
- Matthes T, McKee T, Dunand-Sauthier I, Manfroi B, Park S, Passweg J, Huard B. 2015. Myelopoiesis dysregulation associated to sustained APRIL production in multiple myeloma-infiltrated bone marrow. *Leukemia.* 29:1901–1908.
- McCaffrey TA, Falcone DJ, Du B. 1992. Transforming growth factor- $\beta$ 1 is a heparin-binding protein: Identification of putative heparin-binding regions and isolation of heparins with varying affinity for TGF- $\beta$ 1. *J Cell Physiol.* 152:430–440.
- Moreaux J, Sprynski AC, Dillon SR, Mahtouk K, Jourdan M, Ythier A, Moine P, Robert N, Jourdan E, Rossi JF *et al.* 2009. APRIL and TACI interact with syndecan-1 on the surface of multiple myeloma cells to form an essential survival loop. *Eur J Haematol.* 83:119–129.
- Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ. 1998. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J Comput Chem.* 19:1639–1662.
- Moustakas A, Souchelnyskyi S, Heldin CH. 2001. Smad regulation in TGF- $\beta$  signal transduction. *J Cell Sci.* 114:4359–4369.
- Nordsieck K, Baumann L, Hintze V, Pisabarro MT, Schnabelrauch M, Beck-Sickingler AG, Samsonov SA. 2018. The effect of interleukin-8 truncations on its interactions with glycosaminoglycans. *Biopolymers.* 109:e23103.
- Onufriev A, Case DA, Bashford D. 2002. Effective Born radii in the generalized Born approximation: The importance of being perfect. *J Comput Chem.* 23:1297–1304.
- Panitz N, Theisgen S, Samsonov SA, Gehrcke JP, Baumann L, Bellmann-Sickert K, Köhling S, Pisabarro RJ, Rademann J, Huster D *et al.* 2016. The structural investigation of glycosaminoglycan binding to CXCL12 displays distinct interaction sites. *Glycobiology.* 26:1209–1221.
- Park S, Schulten K. 2004. Calculating potentials of mean force from steered molecular dynamics simulations. *J Chem Phys.* 120:5946.
- Penk A, Baumann L, Huster D, Samsonov SA. 2019. NMR and molecular modeling reveal specificity of the interactions between CXCL14 and glycosaminoglycans. *Glycobiology.* 10:715–725.
- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF chimera—A visualization system for exploratory research and analysis. *J Comput Chem.* 25:1605–1612.
- Pichert A, Samsonov SA, Theisgen S, Thomas L, Baumann L, Schiller J, Beck-Sickingler AG, Huster D, Pisabarro MT. 2012. Characterization of the interaction of interleukin-8 with hyaluronan, chondroitin sulfate, dermatan sulfate and their sulfated derivatives by spectroscopy and molecular modeling. *Glycobiology.* 22:134–145.
- Pomin VH. 2016. Paradigms in the structural biology of the mitogenic ternary complex FGF:FGFR:heparin. *Biochimie.* 127:214–226.
- Potthoff J, Bojarski KK, Kohut G, Lipska AG, Liwo A, Kessler E, Ricard-Blum S, Samsonov SA. 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.* 20:5021.

- Risau W. 1997. Mechanisms of angiogenesis. *Nature*. 386:671–674.
- Sakurai D, Hase H, Kanno Y, Kojima H, Okumura K, Kobata T. 2007. TACI regulates IgA production by APRIL in collaboration with HSPG. *Blood*. 109:2961–2967.
- Salbach-Hirsch J, Samsonov SA, Hintze V, Hofbauer C, Picke AK, Rauner M, Gehrcke JP, Moeller S, Schnabelrauch M, Scharnweber D *et al.* 2015. Structural and functional insights into sclerostin-glycosaminoglycan interactions in bone. *Biomaterials*. 67:335–345.
- Šali A, Blundell TL. 1993. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol*. 234:779–815.
- Samsonov SA, Gehrcke JP, Pisabarro MT. 2014. Flexibility and explicit solvent in molecular-dynamics-based docking of protein-glycosaminoglycan systems. *J Chem Inf Model*. 54:582–592.
- Samsonov SA, Pisabarro MT. 2016. Computational analysis of interactions in structurally available protein–glycosaminoglycan complexes. *Glycobiology*. 26:850–861.
- Samsonov SA, Teyra J, Pisabarro MT. 2011. Docking glycosaminoglycans to proteins: Analysis of solvent inclusion. *J Comput Aided Mol Des*. 25:477–489.
- Schuepbach-Mallepell S, Das D, Willen L, Vigolo M, Tardivel A, Lebon L, Kowalczyk-Quintas C, Nys J, Smulski C, Zheng TS *et al.* 2015. Stoichiometry of heteromeric BAFF and APRIL cytokines dictates their receptor binding and signaling properties. *J Biol Chem*. 290:16330–16342.
- Schwaller J, Schneider P, Mhawech-Fauceglia P, McKee T, Myit S, Matthes T, Tschopp J, Donze O, Le Gal FA, Huard B. 2007. Neutrophil-derived APRIL concentrated in tumor lesions by proteoglycans correlates with human B-cell lymphoma aggressiveness. *Blood*. 109:331–338.
- Van Der Smissen A, Samsonov SA, Hintze V, Scharnweber D, Moeller S, Schnabelrauch M, Pisabarro MT, Anderegg U. 2013. Artificial extracellular matrix composed of collagen I and highly sulfated hyaluronan interferes with TGF $\beta$ 1 signaling and prevents TGF $\beta$ 1-induced myofibroblast differentiation. *Acta Biomater*. 9:7775–7786.
- Sugita Y, Okamoto Y. 2000. Replica-exchange multicanonical algorithm and multicanonical replica-exchange method for simulating systems with rough energy landscape. *Chem Phys Lett*. 329:261–270.
- Swee LK, Ingold-Salamin K, Tardivel A, Willen L, Gaide O, Favre M, Demotz S, Mikkola M, Schneider P. 2009. Biological activity of ectodysplasin A is conditioned by its collagen and heparan sulfate proteoglycan-binding domains. *J Biol Chem*. 284:27567–27576.
- Teyra J, Doms A, Schroeder M, Pisabarro MT. 2006. SCOWLP: A web-based database for detailed characterization and visualization of protein interfaces. *BMC Bioinform*. 7:104.
- Tsui V, Case DA. 2000. Theory and applications of the generalized born solvation model in macromolecular simulations. *Biopolymers*. 56:275–291.
- Uciechowska-Kaczmarzyk U, Babik S, Zsila F, Bojarski KK, Beke-Somfai T, Samsonov SA. 2018. Molecular dynamics-based model of VEGF-A and its heparin interactions. *J Mol Graph Model*. 82:157–166.
- Uciechowska-Kaczmarzyk U, de Beauchene I, Samsonov SA. 2019. Docking software performance in protein-glycosaminoglycan systems. *J Mol Graph Model*. 90:42–50.
- Vallet SD, Miele AE, Uciechowska-Kaczmarzyk U, Liwo A, Duclos B, Samsonov SA, Ricard-Blum S. 2018. Insights into the structure and dynamics of lysyl oxidase propeptide, a flexible protein with numerous partners. *Sci Rep*. 8:11768.
- Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, Darvill AG, Kinoshita T, Packer NH, Prestegard JH *et al.* 2015. *Essentials of Glycobiology*. 3rd ed ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press. p. 2015–2017.
- Wijdenes J, Vooijs WC, Clément C, Post J, Morard F, Vita N, Laurent P, Sun RX, Klein B, Dore JM. 1996. A plasmacyte selective monoclonal antibody (B-B4) recognizes syndecan-1. *Br J Haematol*. 94:318–323.
- Young MR, Pande VS. 2003. Multiplexed-replica exchange molecular dynamics method for protein folding simulation. *Biophys J*. 84:775–786.


# Publication D5

Advanced Molecular Dynamics Approaches  
to Model a Tertiary Complex

APRIL/TACI with Long  
Glycosaminoglycans

## Article

# Advanced Molecular Dynamics Approaches to Model a Tertiary Complex APRIL/TACI with Long Glycosaminoglycans

Mateusz Marcisz<sup>1,2,†</sup>, Martyna Maszota-Zieleniak<sup>1,†</sup>, Bertrand Huard<sup>3</sup> and Sergey A. Samsonov<sup>1,\*</sup> 

<sup>1</sup> Faculty of Chemistry, University of Gdańsk, ul. Wita Stwosza 63, 80-308 Gdańsk, Poland; mateusz.marcisz@gmail.com (M.M.); m.maszota.zieleniak@gmail.com (M.M.-Z.)

<sup>2</sup> Intercollegiate Faculty of Biotechnology of UG and MUG, ul. Abrahamowa 58, 80-307 Gdańsk, Poland

<sup>3</sup> Laboratory TIMC-IMAG, University Grenoble-Alpes, CNRS UMR 5525, 38700 La Tronche, France; bertrand.huard@univ-grenoble-alpes.fr

\* Correspondence: sergey.samsonov@ug.edu.pl; Tel.: +48-58-523-51-66

† These Authors contributed equally to this work.

**Abstract:** Glycosaminoglycans (GAGs) are linear anionic periodic polysaccharides participating in a number of biologically relevant processes in the extracellular matrix via interactions with their protein targets. Due to their periodicity, conformational flexibility, pseudo-symmetry of the sulfation pattern, and the key role of electrostatics, these molecules are challenging for both experimental and theoretical approaches. In particular, conventional molecular docking applied for GAGs longer than 10-mer experiences severe difficulties. In this work, for the first time, 24- and 48-meric GAGs were docked using all-atomic repulsive-scaling Hamiltonian replica exchange molecular dynamics (RS-REMD), a novel methodology based on replicas with van der Waals radii of interacting molecules being scaled. This approach performed well for proteins complexed with oligomeric GAGs and is independent of their length, which distinguishes it from other molecular docking approaches. We built a model of long GAGs in complex with a proliferation-inducing ligand (APRIL) prebound to its receptors, the B cell maturation antigen and the transmembrane activator and calcium modulator and cyclophilin ligand interactor (TACI). Furthermore, the prediction power of the RS-REMD for this tertiary complex was evaluated. We conclude that the TACI–GAG interaction could be potentially amplified by TACI’s binding to APRIL. RS-REMD outperformed Autodock3, the docking program previously proven the best for short GAGs.

**Keywords:** van der Waals replica exchange molecular dynamics; long glycosaminoglycans; APRIL; APRIL receptors; MM/GBSA



**Citation:** Marcisz, M.; Maszota-Zieleniak, M.; Huard, B.; Samsonov, S.A. Advanced Molecular Dynamics Approaches to Model a Tertiary Complex APRIL/TACI with Long Glycosaminoglycans. *Biomolecules* **2021**, *11*, 1349. <https://doi.org/10.3390/biom11091349>

Academic Editors: Thomas R. Caulfield and Alexander T. Baker

Received: 9 July 2021

Accepted: 10 September 2021

Published: 12 September 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/>).

## 1. Introduction

Despite the recent advances in molecular docking of glycosaminoglycan (GAG) oligosaccharides, it still remains a challenge to dock longer GAGs [1]. The main reason for this is the physical–chemical nature of GAGs. They are long, periodic, and linear polysaccharides. They are negatively charged and manifest different binding and conformational properties based on their sulfation pattern and negative charge distribution [2]. GAGs are built of disaccharide units consisting of amino sugars and uronic acid or galactose [3]. Depending on their arrangement and sulfation pattern, those units may display 408 [4] variants, of which 202 are found in mammals [5,6]. Although GAG’s certain binding specificity has been observed in several biologically relevant systems [7–9], protein–GAG interactions are often predominantly electrostatics-driven, and their binding energies correlate with the GAG net charge [10–13]. Despite the fact that computational studies of GAGs persist as a general challenge due to the required conformational sampling and their periodicity, there are numerous successful studies on proteins complexes with shorter GAG oligosaccharides (of length up to octasaccharides) [1,14–18]. On the other hand, there are very few studies that focus on longer GAG molecules in complexes with proteins. The



reason for that is that there is no appropriate tool for the docking long GAGs to properly account for their conformational space and periodicity. In the case of Autodock3, which has been shown the most accurate tool for docking GAGs [19], there is a limitation of 32 torsional degrees of freedom for the docked molecule, which makes a docking for bigger GAG molecules practically rigid. Dynamic molecular docking allows for flexible docking of GAGs of arbitrary length, implementing targeted MD toward the *a priori* known binding region on the protein surface [20]. However, in case there are several potential binding regions or a binding region is too extensive and/or the ligand is particularly long, this approach is computationally too expensive. The next approach designed to overcome this issue of GAG length is a fragment-based method, in which trimeric GAGs are docked with Autodock3 and further assembled into longer chains [21]. This method also has a potential flaw since it may fail to properly dock when the GAG binding domain has some negatively charged residues that restrict the search of favorable binding patches; thus, short GAG probes cannot be docked near those residues, rendering the length of the assembled long GAG chain limited. The RS-REMD (replica exchange molecular dynamics with repulsive scaling) method [22,23] is not restricted by any of the above-mentioned limits. Moreover, it has been proven that this method is appropriate to dock GAGs [24]. In RS-REMD, effective pairwise radii are increased in different Hamiltonian replicas. In GAG–protein complexes, very often, electrostatic interactions play a main role by establishing strong charge–charge interactions and, therefore, limiting dissociation or any dramatic conformational changes allowing for avoiding binding in a local minimum. Increasing pairwise van der Waals radii as it is done in RS-REMD (while not affecting other types of interactions for the system) can be helpful to overcome this challenge. In this way, the mentioned method allows for a robust and extensive search for the proper binding poses on the complete protein surface, allowing, at the same time, for full flexibility of the docked molecule and the receptor side chains. In this study, a 24-mer and a 48-mer of heparin (HP) were docked to two complexes of a proliferation-inducing ligand (APRIL) protein and its receptors—the transmembrane activator and calcium modulator and cyclophilin ligand interactor (TACI) and the B cell maturation antigen (BCMA). APRIL is a member of the TNF superfamily [25] that was shown to bind GAGs (chondroitin sulfate and heparan sulfate) [13,26–28]. Such binding is thought to mediate APRIL’s oligomerization and, therefore, enable its role in cell signaling [29]. The GAG binding region on APRIL’s surface is located near the N-terminus of the protein alongside a stretch of positively charged lysine residues [13]. Additionally, it was reported that the C-terminus spatially close to the N-terminus, together with several arginine residues on the side of the protein, also contribute to GAG binding [13]. While this binding of GAGs to APRIL is believed to be a facilitation agent for its binding to receptors, BCMA and TACI [30,31], it has been shown that BCMA does not bind to HSPG (heparan sulfate proteoglycan) [32,33]. On the other hand, GAG binding to TACI is a little controversial. Few studies report that TACI interacts with proteoglycan [32–34]. However, some studies claim no binding of TACI to HSPG [35] or find it unlikely [13].

The docking of such long GAGs in a biologically relevant system has been performed for the first time to our knowledge. In this work, we analyzed the data obtained with RS-REMD for APRIL–BCMA/TACI–HP complexes and compared the docking performance and predictive power of this method to the ones of the conventional molecular docking method (Autodock3). Our results contribute to the general knowledge about GAG-specific computational approaches.

## 2. Materials and Methods

### 2.1. Structures

**Protein structures.** Following X-ray experimental structures from PDB were used in this work: 1XU2 (the crystal structure of APRIL bound to BCMA) and 1XU1 (the crystal structure of APRIL bound to TACI) [36].

**GAG structures.** HP dp24 and dp48—dp stands for the degree of polymerization—were constructed from building blocks of the sulfated GAG monomeric units’ libraries [37]

compatible with AMBER16 package based on the experimental structure of HP (PDB ID: 1HPN). The GLYCAM06 force field [38] and the literature data for the sulfate groups [39] were the sources of GAGs' charges.

## 2.2. Molecular Docking

Autodock3 was used as a standard docking tool, as it has been previously described to yield the best results for GAG–protein complexes [19,40]. Entire protein was covered using a maximum gridbox size ( $126 \text{ \AA} \times 126 \text{ \AA} \times 126 \text{ \AA}$ ) with a  $0.853 \text{ \AA}$  grid step. The size of 300 for the initial population and  $10^5$  generations for termination conditions were chosen. One thousand independent runs with the Lamarckian genetic algorithm were used. In total,  $9995 \times 10^5$  energy evaluations were performed. DBSCAN algorithm [41] was used for clustering. RMSatd metric was used for clustering, which accounts for the equivalence of the atoms of the same atomic type. This metric was reported to be more appropriate for GAG docking than classical RMSD for periodic ligands [20].

RS-REMD (replica exchange with repulsive scaling) [22] was used as an effective docking alternative to Autodock3 [24]. The ff14SBonlysc force field parameters [42] for protein and the GLYCAM06 [38] for GAGs were used, respectively. Every step of the docking simulation was followed, as described in detail in the work of Maszota et al. [24], and all the parameters were used as described there.

## 2.3. Molecular Dynamics

All the MD simulations of the complexes obtained by Autodock3 docking were performed in AMBER16 package [43]. TIP3P truncated octahedron water box with a distance of  $8 \text{ \AA}$  from the solute to the box's border was used to solvate complexes.  $\text{Na}^+$  counterions were used to neutralize the charge of the system. Energy minimization was performed preceding the production MD runs. 500 steepest descent cycles and  $10^3$  conjugate gradient cycles with  $100 \text{ kcal/mol/\AA}^2$  harmonic force restraint on solute atoms were performed. It was followed by  $3 \times 10^3$  steepest descent cycles and  $3 \times 10^3$  conjugate gradient cycles without any restraints and continued with heating up the system to 300 K for 10 ps with harmonic force restraints of  $100 \text{ kcal/mol/\AA}^2$  on solute atoms. Then, the system was equilibrated at 300 K and  $10^5 \text{ Pa}$  in an isothermal, isobaric ensemble for 500 ps. The actual MD runs were carried out using the same isothermal, isobaric ensemble for 100 ns. The timestep of 2 fs, the cut-off of  $8 \text{ \AA}$  for electrostatics were used. Particle mesh Ewald method for treating electrostatics [44] and SHAKE algorithm for all the covalent bonds containing hydrogen atoms [45] were implemented in the MD simulations. Cpptraj module of AMBER was used for the analysis of the trajectories. In particular, native contacts command with default parameters was used for the analysis of the contacts between protein and GAG molecules established in the course of the simulation.

## 2.4. Binding Free Energy Calculations

For the free energy and per residue energy decomposition calculations, the MM/GBSA (molecular mechanics generalized born surface area) model  $\text{igb} = 2$  [46] from AMBER16 was used on trajectories obtained from MD simulations. These energy values should be rather understood as the enthalpy of binding rather than strictly defined binding free energy and partially include the entropic contribution of the solvent. It was previously shown that for the MM/PBSA (MM/GBSA) approach, entropy calculations would rather increase the uncertainty of the calculated free energy values, in general [47], and for protein–GAG systems, in particular [48]. LIE analysis with a dielectric constant of 80, performed by CPPTRAJ scripts on the same frames as the MM/GBSA.

## 3. Results and Discussion

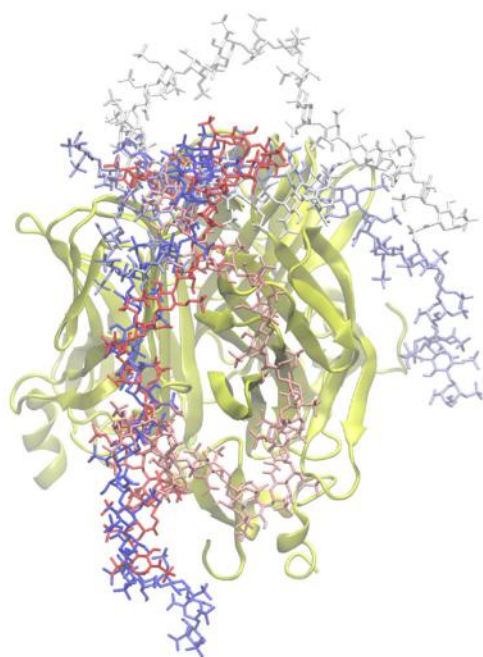
### 3.1. Docking Heparin dp24 to APRIL–BCMA and APRIL–TACI Complexes Using Autodock3

The internal limitation of Autodock3 allowed specification of only 32 torsional degrees of freedom for the docked ligand as free to rotate, rendering flexible docking of longer

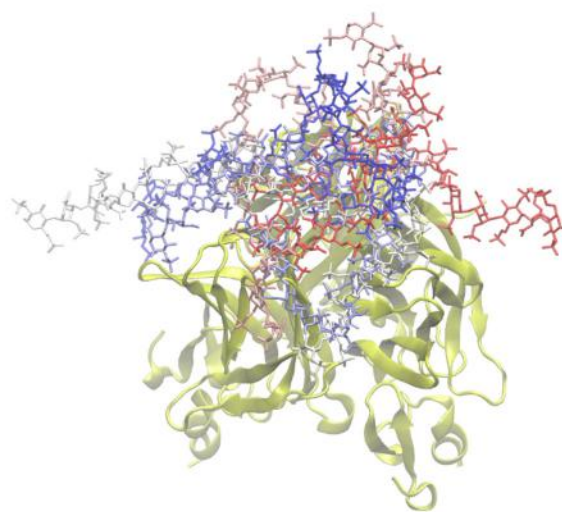
(over dp6 or dp8) GAGs unfeasible. Due to this technical limitation of the software, in the performed docking with Autodock3, the central part of the heparin molecules was kept flexible, while the rest remained rigid. Only central 16 glycosidic linkages, each described by two dihedral angles, were flexible, while other degrees of freedom were not considered in the docking run. For these partially rigid docked molecules, we observed some heparin-binding poses for which the overall oligosaccharide backbone geometry did not look correct. Some of them seemed highly unlikely to occur from an energetic standpoint, considering how spatially close the monosaccharide units with the same charge were placed. After visual analysis of the top 50 docked solutions (in terms of energy scores of Autodock3), we selected three and five ones for APRIL–BCMA and APRIL–TACI, respectively, that looked particularly distorted. (Figures S1 and S2) This suggests that the internal ligand conformations are not properly scored by Autodock3; otherwise, such binding poses might not have been included within the energetically favorable ones. Since it is known that Autodock3 can potentially generate glycosidic linkages with inappropriate geometry [49], we checked if the global distortions of the docked HP chains could be related to the locally distorted geometry of the glycosidic linkages (Figure S3). However, all the glycosidic linkages obtained with Autodock3 were located in the same regions as the glycosidic linkages for the unbound HP molecules from the microsecond-range MD simulations, suggesting that the global distortion of long HP molecules in Autodock3 are independent of the local glycosidic linkage geometry that is correct in the analyzed cases. Therefore, the unexpected overall geometries of long HP ligands are originated from another Autodock3 feature.

The best five poses in terms of energy (as scored by Autodock3) were further analyzed with the MD approach for both complexes. (Figure 1) These solutions were highly heterogeneous, though they all were partially located at the GAG binding site on the APRIL molecule, similar to our previous findings obtained in the absence of BMCA/TACI receptors [13]. This suggests that a long GAG would first bind the APRIL molecule independently of the presence of the receptors.

### APRIL-BCMA with HP dp24



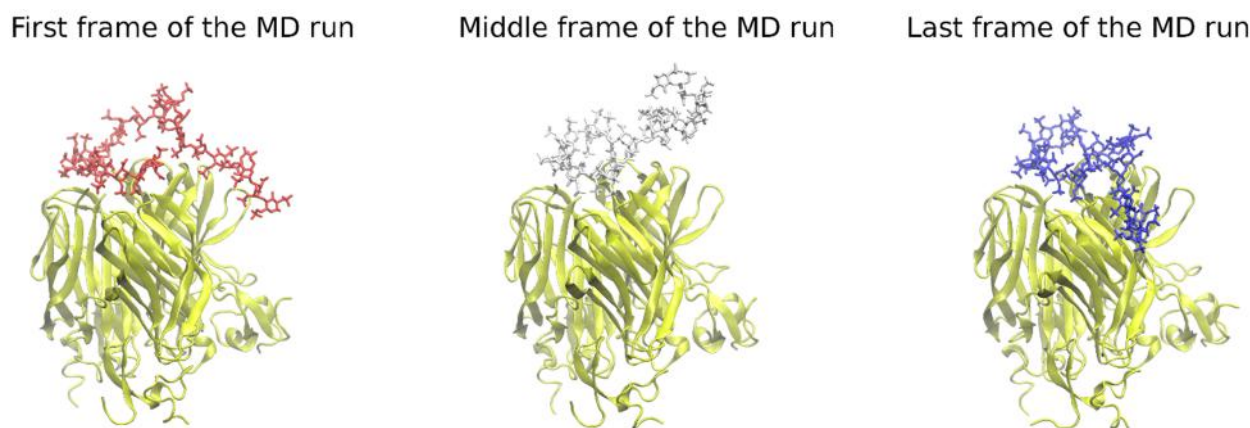
### APRIL-TACI with HP dp24



**Figure 1.** The five most energetically favorable structures of the APRIL–BCMA (left) and APRIL–TACI (right) complexes with heparin dp24, obtained by docking with Autodock3 (heparin in licorice representation, red-white-blue range of color; APRIL, BCMA, TACI in cartoon representation in yellow).

The conformations of the heparin over the course of the MD runs starting from the structures obtained by Autodock3 and expressed by RMSD shows from medium to high values ( $9.1 \pm 4.7 \text{ \AA}$  and  $7.3 \pm 1.9 \text{ \AA}$  for APRIL–BCMA and APRIL–TACI, respectively), indicating that these starting conformations did not correspond to the structures favored by the GLYCAM06 force field used in the MD (Table S1). When also taking into account the movement of the GAG on the protein surface, the RMSD values were  $15.5 \pm 4.7 \text{ \AA}$  and  $11.7 \pm 1.8 \text{ \AA}$  for APRIL–BCMA and APRIL–TACI complexes, respectively. It is worth noting that the part of the HP that seems to be docked properly in the GAG binding region of the APRIL protein remained stable (trapped in a minimum) during the MD runs and, thus, did not significantly contribute to the high RMSD (Figure 2). Therefore, lateral parts of the GAG molecule potentially have been docked wrongly, rendering intra- and intermolecular interactions to be unfavorable and forcing GAG to search for a more energetically convenient pose in the MD simulations. The rigidity of these parts could practically be the reason for such an observation. The evolution of radius of gyration (Rgyr) of the HP dp24 and the whole protein–HP complex, as well as RMSD of the HP molecule (Figures S4 and S5, respectively), show that, except for one of the MD simulations of one BCMA complex, both Rgyr and RMSD converge through the 100 ns of the MD simulation.

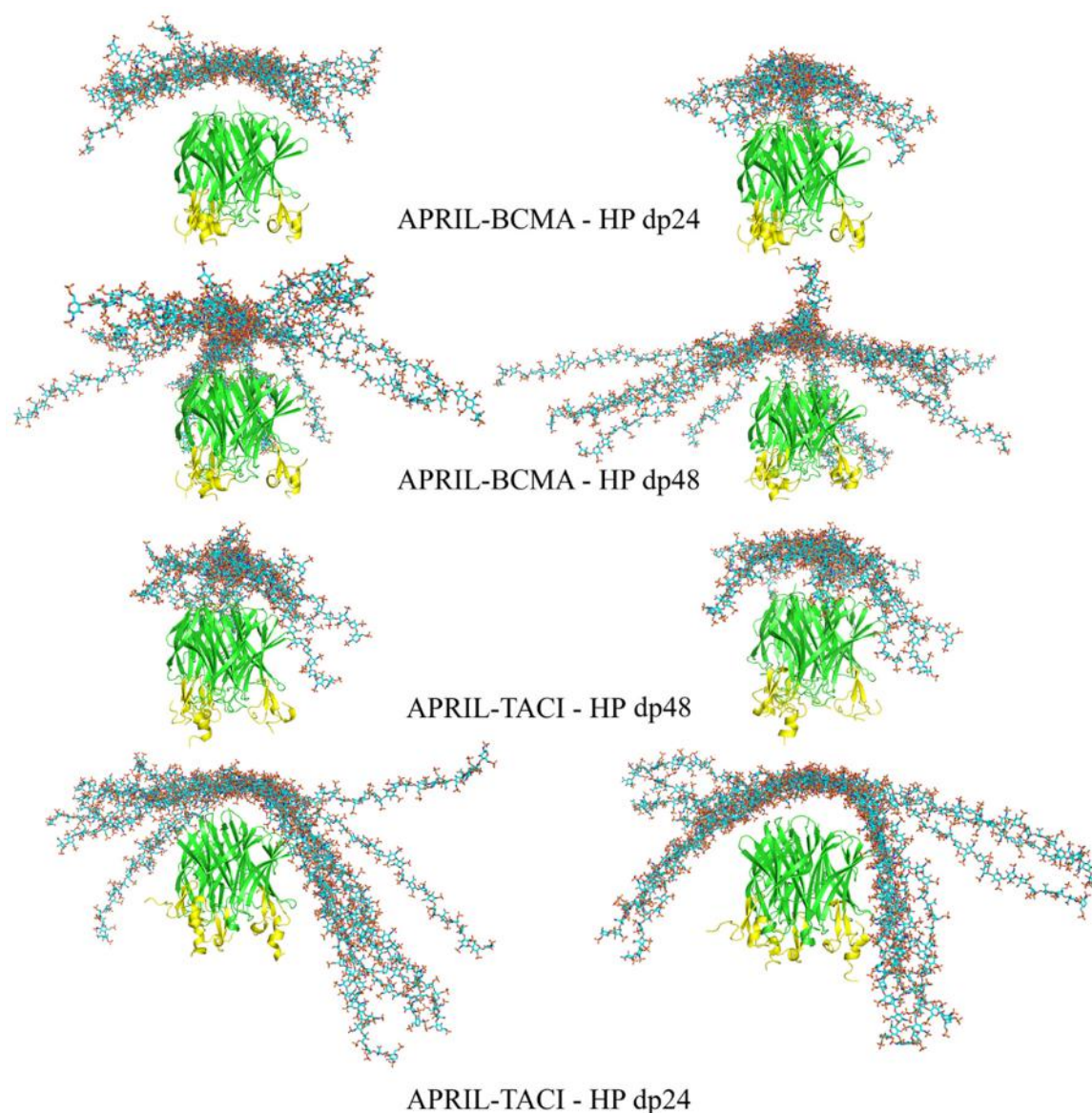
Binding free energy analysis using LIE yielded mean energy values of  $-102.2 \pm 17.8 \text{ kcal/mol}$  and  $-125.8 \pm 13.4 \text{ kcal/mol}$  for the APRIL–BCMA and APRIL–TACI systems, respectively. There is an expected significant difference [13] in the absolute values between the LIE data and the data obtained using MM/GBSA, which yielded more favorable energies:  $-169.7 \pm 34.3$ – $244.9 \pm 15.5 \text{ kcal/mol}$  APRIL–BCMA and for APRIL–TACI, respectively. These differences are expected to be higher in a case when the system is highly charged since, within the MM/GBSA procedure, minimization is performed for all frames, leading to overestimation of the binding strength of the complex. At the same time, the LIE protocol should be optimized in terms of the dielectric constant for each particular molecular system, which requires experimental data. Energies obtained with MM/GBSA for the first and the last 10 ns were compared for both complexes. For the APRIL–BCMA complex, they were  $-155.5 \pm 22.5 \text{ kcal/mol}$  at the start and  $-194.3 \pm 33.9 \text{ kcal/mol}$  at the end of the runs, respectively. The data for the APRIL–TACI complex follows the same trend: the corresponding energies were  $-217.3 \pm 28.9 \text{ kcal/mol}$  at the start and  $-255.8 \pm 38.0 \text{ kcal/mol}$  at the end of the runs, respectively. A decrease in binding energy by such a significant amount suggests that starting conformations were far from being optimal. To sum up, Autodock3, when applied to long GAGs such as HP dp24, yields many artifacts that are partially repaired in the followed MD step. However, even long conventional MD simulations are not able to globally change the conformation of the docked GAG within the practically accessible times.



**Figure 2.** The heparin movement on the protein’s surface in the course of the MD run. Some parts of the HP tend to move substantially during MD runs, searching for an energetically more favorable pose (heparin in licorice representation, red, white, and blue color; APRIL, BCMA, TACI in cartoon representation, yellow color).

### 3.2. Docking Heparin dp24 and dp48 to APRIL–BCMA and APRIL–TACI Complexes Using RS-REMD

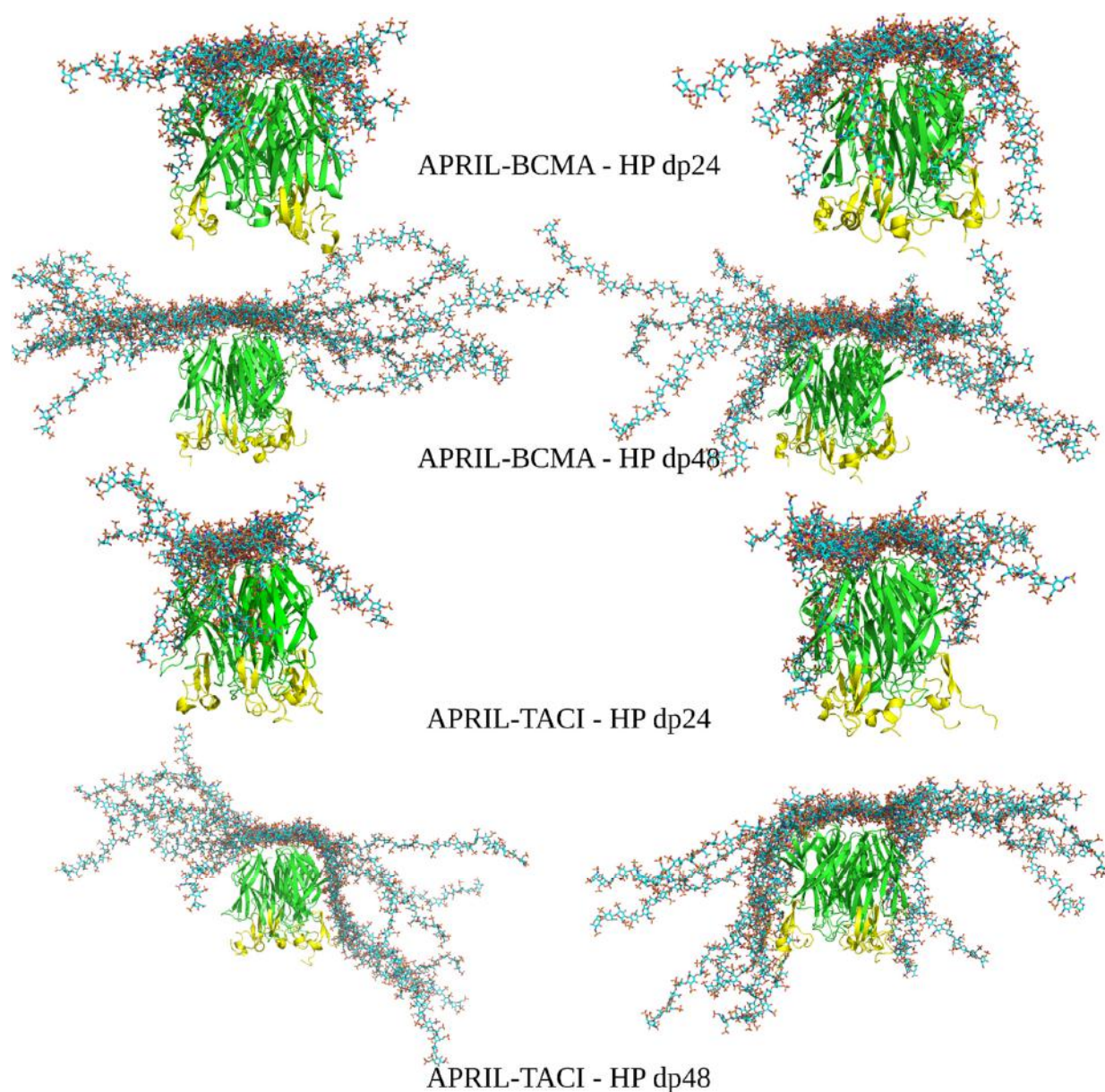
RS-REMD simulations for HP dp24 and dp48 were performed for both APRIL–BCMA and APRIL–TACE complexes, and the results were compared with the ones obtained from Autodock3 described above. Afterward, a refinement procedure was carried out for the docking poses obtained during the RS-REMD simulation, as was described in detail in our previous work [24]. First, MM-GBSA analysis for the whole RS-REMD trajectories was performed, then 10 docking poses with the lowest electrostatic energy were selected for the refinement procedure, as this free energy component proved to perform better for scoring in protein–GAG systems [24]. At the same time, 10 poses were also selected manually based on the visual criteria that the HP lateral parts were directed toward the BCMA/TACI APRIL receptors to determine if such structures could be energetically favorable in comparison to other ones, which potentially could mean that HP molecules favourize the binding between APRIL and its receptors. All starting structures are presented in Figure 3.



**Figure 3.** The most energetically favorable (**left**) and manually chosen (**right**) structures of the APRIL–BCMA and APRIL–TACI complexes with HP dp24 and HP dp48 after RS-REMD simulation (HP in licorice representation, cyan color; APRIL (green color) and BCMA/TACI (yellow color) in cartoon representation).

There are no significant differences between the structures selected for the improvement procedure based on the electrostatic energy value and those selected manually. In the case of the HP dp24 ligand, the differences between the APRIL protein with BCMA and TACI receptors are neither visible. In both cases, the ligand is docked at the top of the protein trimer with its middle part, while the ends point toward the solution. In contrast, for the HP dp48 ligand, it can be observed that it still docks at the top of the protein for both APRIL–BCMA and APRIL–TACI, but its ends, especially in the case of APRIL–TACI, are directed toward the receptors. This means that RS-REMD predicts the potential binding of the long GAGs not only to APRIL, which is by far the strongest, but also to the APRIL receptors.

To refine the contacts between the BCMA/TACI receptors and HP molecules in the explicit solvent, 100 ns and 20 ns of the MD simulations were performed for dp24 and dp48, respectively. The structures of the complexes obtained in the refinement procedure are shown in Figure 4.



**Figure 4.** The most energetically favorable (**left**) and manually chosen (**right**) structures of the APRIL–BCMA and APRIL–TACI complexes with HP dp24 and HP dp48 after refinement procedure (HP in licorice representation, cyan color; APRIL (green color) and BCMA/TACI (yellow color) are in cartoon representation).

It is clearly seen that for all the refined complexes, the distance between the HP ligands and the protein has decreased. The per-frame evolution of HP RMSD and MM/GBSA binding free energy is shown for representative APRIL-BCMA-HP dp24 and APRIL-TACI-HP dp24 refinement MD simulations (Figure S6), where convergence is demonstrated. At the same time, we would like to underline that the work did not aim to achieve proper convergence of the simulations but to obtain an ensemble of docked binding poses using the RS-REMD methodology and to compare its prediction power with the one of Autodock3. The obtained binding poses can be further analyzed by applying longer MD simulations depending on the scientific goal in a particular study, in which an RS-REMD docking scheme is applied as the first step to obtain a starting complex between a protein and a long GAG. For the HP dp24 complexes, no contacts with the BCMA receptor can be observed, while they are visible for the APRIL-TACI-HP dp24 complex. Similarly, the contacts in complexes between APRIL-BCMA/TACI-HP dp48 have been improved. Again, contacts between the HP ligand and the TACI receptor are more frequent than in the case of BCMA.

Contacts evolution for Autodock3-docked and RS-REMD (after the refinement)-docked HP dp24 was compared (Figures S7 and S8). In both procedures, there were more contacts observed in the case of TACI complexes, while the number of contacts does not differ essentially for the Autodock3 and RS-REMD MD simulations. For Autodock3, there are already existing contacts (initial contacts), while due to the nature of the RS-REMD procedure [24], there are barely any initial contacts prior to the refinement.

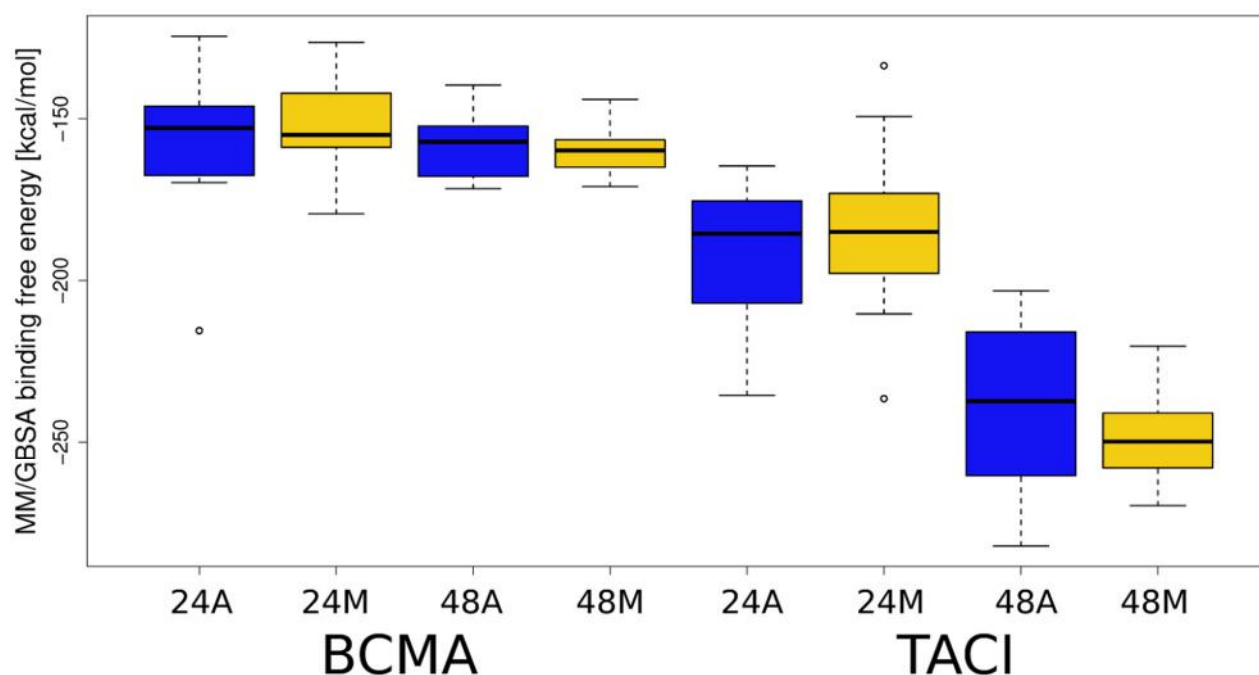
Furthermore, we analyzed the dependence of the calculated binding energies obtained by Autodock3-based and RS-REMD approaches on the angle defined by the terminal (the first) monosaccharide unit of HP-central atom of the HP chain-the middle of the HP molecule-the terminal (the last) monosaccharide unit of HP (Figure S9). For both BCMA and TACI complexes, Autodock3 results show a less pronounced preference of the binding energies for the lower angle values, meaning that the bending of the GAG around the APRIL-receptor complex is less favored than in the case of the RS-REMD procedure.

To discover which of the analyzed complexes have the most favorable energy, as well as which receptor amino acid residues are responsible for binding HP, MM-GBSA binding and per residue decomposition analysis were performed. The results of the MM-GBSA analysis are shown in Figure 5, Table S2. For all analyzed complexes, the free energy of binding is more favorable for HP dp48 than for HP dp24, suggesting that despite the core binding site on the top of the APRIL-receptor complexes, the elongation of a GAG essentially strengthens the binding. Significant differences in  $\Delta G$  value are noticeable for APRIL-BCMA-HP and APRIL-TACI-HP complexes. The mean free energy of binding for the APRIL-BCMA complexes is about 25% less favorable than for the APRIL-TACI-HP complexes. We observed the same trend when GAGs were docked with Autodock3. It is worth noting that for the APRIL-BCMA-HP complexes, there are noticeable differences between the energies of the structures selected on the basis of the RS-REMD electrostatic energy value and those selected manually. However, in the case of APRIL-TACI-HP, the differences are practically imperceptible.

Next, per residue decomposition was performed allowing to identify the most energetically favorable BCMA/TACI amino acid residues involved in the formation of the complexes. The amino acid residues with the decomposed binding free energies below  $-1.5$  kcal/mol are listed in Table S3.

The number of the BCMA/TACI amino acid residues involved in complex formation with HP increased with the increasing ligand length. This is due to the possibility of contacts between HP dp48 and several receptor monomers simultaneously, while HP dp24 is too short to be able to interact with several receptors at once. Free energy values decrease with ligand length for both APRIL-BCMA and APRIL-TACI proteins. Importantly, the free energy values take lower values for APRIL-TACI than for APRIL-BCMA: the lowest and the average values for APRIL-BCMA-HP dp24, APRIL-TACI-HP dp24, APRIL-BCMA-HP dp48, and APRIL-TACI-HP dp48 are  $-2.2$  kcal/mol and  $-1.9 \pm 0.2$  kcal/mol,  $-3.8$  kcal/mol and  $-2.1 \pm 0.6$  kcal/mol,  $-3.1$  kcal/mol and  $-2.9 \pm 0.2$  kcal/mol, and

−11.8 kcal/mol and  $-4.3 \pm 2.0$  kcal/mol, respectively. All these results suggest that in the case when a long GAG molecule is indeed able to bind simultaneously to both APRIL and its receptors, this would strengthen the interactions in an APRIL–receptor complex. Such an effect would be more pronounced in the case of TACI binding to APRIL. This is in agreement with our previous simulations indicating that TACI is more prone to bind GAGs than BCMA [13].



**Figure 5.** MM/GBSA binding free energies obtained for APRIL–BCMA/TACI–HP systems. Boxplot representation for the values provided in Table S2 is plotted: X-axis, BCMA and TACI correspond to the APRIL complexes with BCMA and TACI, respectively; 24 and 48 correspond to HP dp24 and HP dp48, respectively; A and M correspond to the automatically and manually selected structures, respectively.

#### 4. Conclusions

Long (dp24 and dp48) heparin molecules were successfully docked to the binding site of the APRIL–BCMA and APRIL–TACI complexes. In the case of the APRIL–TACI complex with heparin, we observed stronger binding than in the APRIL–BCMA complex. In all cases, the GAG is first bound to the APRIL GAG binding site, while secondary interactions are established with the receptors. It is puzzling, however, how TACI, especially remotely located from a GAG in the case of the HP dp24, could substantially impact GAG binding by the APRIL protein. Probably this is due to the long-range electrostatics effect occurring in this highly charged system. At the same time, multiple studies find no or, rather, unlikely binding of GAGs to TACI [13,35]. We conclude that the TACI–GAG interaction could be amplified by TACI’s binding to APRIL, while this interaction was not detected in the absence of APRIL. In this study, RS-REMD showed clear superiority over Autodock3 in the case of docking quality for such long molecules. First of all, it is almost unfeasible to dock dp48 GAGs properly with Autodock3, while docking dp24 experiences essential challenges. However, we have to admit that despite some geometrically inappropriate docking poses in the case of Autodock3 when applied for HP dp24, most of the docking structures were not distorted. Nonetheless, they showed higher instability of the HP binding poses in the course of the MD simulations than in the case of RS-REMD, suggesting worse sampling during the docking by Autodock3. Considering the fact that RS-REMD yielded better results without the need for more computational resources for a representative protein–GAG dataset [24], we strongly believe that this method is far more superior in the case of docking long GAG molecules.



**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/biom11091349/s1>, Table S1: RMSD values from the performed MD runs. RMS-GAG means RMSD of the GAG molecule in respect to its starting pose, when RMS-GAG-PROTEIN additionally include GAG position in respect to protein, Table S2: MM-GBSA analysis for the APRIL-BCMA/TACI-HP MD simulations, Table S3: MM-GBSA per residue decomposition analysis for the APRIL-BCMA/TACI-HP MD simulations. The residues belonging to BCMA/TACI molecules are only listed. The letter after the residue corresponds to the monomeric unit within the trimer, Figure S1: Five docked poses selected out of top 50 (in terms of energy evaluation with Autodock3) that were distorted in terms of the global heparin geometry for APRIL-BCMA complex (heparin in licorice representation, cyan colour; APRIL and BCMA in cartoon representation, yellow colour), Figure S2: Five docked poses selected out of top 50 (in terms of energy evaluation with Autodock3) that were distorted in terms of the global heparin geometry for APRIL-TACI complex (heparin in licorice representation, cyan colour; APRIL and TACI in cartoon representation, yellow colour), Figure S3: Glycosidic linkage heatmaps for HP  $\varphi$  and  $\psi$  dihedral angles defined as  $O5n+1-C1n+1-O4n-C4n$  and  $C1n+1-O4n-C4n-C4n$  where n stands for sequential number of a sugar monomeric unit, the data are obtained from 5  $\mu$ s MD simulation for the unbound HP dp10. The data for glycosidic linkages of each type (GlcNS(6S)-IdoA(2S) in A and GlcNS(6S)-IdoA(2S) in B) are averaged. The data for glycosidic linkages for one of the distorted poses of HP dp24 docked to APRIL/BCMA obtained with Autodock3 are in magenta points, Figure S4: APRIL-BCMA-HP dp24 (left panel) and APRIL-TACI-HP dp24 (right panel) per frame Rgyr analysis, Figure S5: APRIL-BCMA-HP dp24 (left panel) and APRIL-TACI-HP dp24 (right panel) per frame RMSD analysis, Figure S6: Per frame evolution of HP MM/GBSA binding free energy and RMSD for two representative APRIL-BCMA-HP dp24 (left panel) and APRIL-TACI-HP dp24 (right panel) RS-REMD refinement MD simulations corresponding to the lowest free energy values, Figure S7: AD3-predicted APRIL-BCMA-HP dp24 (left panel) and APRIL-TACI-HP dp24 (right panel) contact number analysis, Figure S8: RS-REMD-predicted APRIL-BCMA-HP dp24 (left panel) and APRIL-TACI-HP dp24 (right panel) contact number analysis, Figure S9: MM/GBSA free binding energy dependence on the angle defined by the terminal (the first) monosaccharide unit of HP-central atom of the HP chain-the terminal (the last) monosaccharide unit of HP obtained by AD3 and RS-REMD approaches for APRIL-BCMA-HP dp24 (left panel) and APRIL-TACI-HP dp24 (right panel) complexes.

**Author Contributions:** M.M. and M.M.-Z. obtained and analyzed the data and adapted the protocols to the studied systems. B.H. and S.A.S. are responsible for the conceptualization of the work. All authors contributed to the writing of the manuscript. S.A.S. supervised M.M. and M.M.-Z. and obtained the funding. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Science Centre of Poland, grant number UMO-2018/30/E/ST4/00037.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The computational resources were provided by the Polish Grid Infrastructure (PL-GRID, grants: plggagstr2gpu for S.A.S., plgaprilgag2, gpuaprilgags for M.M.) and the cluster at the Faculty of Chemistry, University of Gdańsk. The molecular docking calculations were performed on the ZIH cluster at TU Dresden (grant p\_gag for S.A.S.).

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

## References

1. Paiardi, G.; Milanesi, M.; Wade, R.C.; D'ursi, P.; Rusnati, M. A Bittersweet Computational Journey among Glycosaminoglycans. *Biomolecules* **2021**, *11*, 739. [[CrossRef](#)]
2. Habuchi, H.; Habuchi, O.; Kimata, K. Sulfation Pattern in Glycosaminoglycan: Does It Have a Code? *Glycoconj. J.* **2004**, *21*, 47–52. [[CrossRef](#)]
3. Varki, A.; Cummings, R.D.; Esko, J.D.; Stanley, P.; Hart, G.W.; Aebi, A.; Darvill, A.G.; Kinoshita, T.; Packer, N.H.; Prestegard, J.H.; et al. *Essentials of Glycobiology*, 3rd ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, USA, 2015.
4. Bu, C.; Jin, L. NMR Characterization of the Interactions between Glycosaminoglycans and Proteins. *Front. Mol. Biosci.* **2021**, *8*, 165. [[CrossRef](#)]

5. Clerc, O.; Mariethoz, J.; Rivet, A.; Lisacek, F.; Pérez, S.; Ricard-Blum, S. A Pipeline to Translate Glycosaminoglycan Sequences into 3D Models. Application to the Exploration of Glycosaminoglycan Conformational Space. *Glycobiology* **2019**, *29*, 36–44. [[CrossRef](#)]
6. Vallet, S.D.; Clerc, O.; Ricard-Blum, S. Glycosaminoglycan–Protein Interactions: The First Draft of the Glycosaminoglycan Interactome. *J. Histochem. Cytochem.* **2020**, *69*, 93–104. [[CrossRef](#)] [[PubMed](#)]
7. Van Der Smissen, A.; Samsonov, S.; Hintze, V.; Scharnweber, D.; Moeller, S.; Schnabelrauch, M.; Pisabarro, M.T.; Anderegg, U. Artificial Extracellular Matrix Composed of Collagen i and Highly Sulfated Hyaluronan Interferes with TGFβ1 Signaling and Prevents TGFβ1-Induced Myofibroblast Differentiation. *Acta Biomater.* **2013**, *9*, 7775–7786. [[CrossRef](#)]
8. Pichert, A.; Samsonov, S.A.; Theisgen, S.; Thomas, L.; Baumann, L.; Schiller, J.; Beck-Sickinger, A.G.; Huster, D.; Pisabarro, M.T. Characterization of the Interaction of Interleukin-8 with Hyaluronan, Chondroitin Sulfate, Dermatan Sulfate and Their Sulfated Derivatives by Spectroscopy and Molecular Modeling. *Glycobiology* **2012**, *22*, 134–145. [[CrossRef](#)] [[PubMed](#)]
9. Bojarski, K.K.; Samsonov, S.A. Role of Oligosaccharide Chain Polarity in Protein–Glycosaminoglycan Interactions. *J. Chem. Inf. Model.* **2021**, *61*, 455–466. [[CrossRef](#)]
10. Gallagher, J. Fell-Muir Lecture: Heparan Sulphate and the Art of Cell Regulation: A Polymer Chain Conducts the Protein Orchestra. *Int. J. Exp. Pathol.* **2015**, *96*, 203–231. [[CrossRef](#)] [[PubMed](#)]
11. Kjellén, L.; Lindahl, U. Specificity of Glycosaminoglycan–Protein Interactions. *Curr. Opin. Struct. Biol.* **2018**, *50*, 101–108. [[CrossRef](#)]
12. Imberty, A.; Lortat-Jacob, H.; Pérez, S. Structural View of Glycosaminoglycan–Protein Interactions. *Carbohydr. Res.* **2007**, *342*, 430–439. [[CrossRef](#)]
13. Marcisz, M.; Huard, B.; Lipska, A.G.; Samsonov, S.A. Further Analyses of APRIL / APRIL-Receptor / Glycosaminoglycan Interactions by Biochemical Assays Linked to Computational Studies. *Glycobiology* **2021**, *31*, 772–786. [[CrossRef](#)]
14. Uciechowska-Kaczmarzyk, U.; Babik, S.; Zsila, F.; Bojarski, K.K.; Beke-Somfai, T.; Samsonov, S.A. Molecular Dynamics-Based Model of {VEGF}-[A] and Its Heparin Interactions. *J. Mol. Graph. Model.* **2018**, *82*, 157–166. [[CrossRef](#)]
15. Gandhi, N.S.; Mancera, R.L. Molecular Dynamics Simulations of CXCL-8 and Its Interactions with a Receptor Peptide, Heparin Fragments, and Sulfated Linked Cyclitols. *J. Chem. Inf. Model.* **2011**, *51*, 335–358. [[CrossRef](#)]
16. Joseph, P.R.B.; Mosier, P.D.; Desai, U.R.; 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. [[CrossRef](#)] [[PubMed](#)]
17. Clausen, T.M.; Sandoval, D.R.; Spliid, C.B.; Pihl, J.; Perrett, H.R.; Painter, C.D.; Narayanan, A.; Majowicz, S.A.; Kwong, E.M.; McVicar, R.N.; et al. SARS-CoV-2 Infection Depends on Cellular Heparan Sulfate and ACE2. *Cell* **2020**, *183*, 1043–1057.e15. [[CrossRef](#)] [[PubMed](#)]
18. Schuur, Z.P.; Hammond, E.; Elli, S.; Rudd, T.R.; Mycroft-West, C.J.; Lima, M.A.; Skidmore, M.A.; Karlsson, R.; Chen, Y.H.; Bagdonaite, I.; et al. Evidence of a Putative Glycosaminoglycan Binding Site on the Glycosylated SARS-CoV-2 Spike Protein N-Terminal Domain. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 2806–2818. [[CrossRef](#)] [[PubMed](#)]
19. Uciechowska-Kaczmarzyk, U.; de Beauchene, I.; Samsonov, S.A. Docking Software Performance in Protein–Glycosaminoglycan Systems. *J. Mol. Graph. Model.* **2019**, *90*, 42–50. [[CrossRef](#)] [[PubMed](#)]
20. Samsonov, S.A.; Gehrcke, J.P.; Pisabarro, M.T. Flexibility and Explicit Solvent in Molecular-Dynamics-Based Docking of Protein–Glycosaminoglycan Systems. *J. Chem. Inf. Model.* **2014**, *54*, 582–592. [[CrossRef](#)]
21. Samsonov, S.A.; Zacharias, M.; Chauvot de Beauchene, I. Modeling Large Protein–Glycosaminoglycan Complexes Using a Fragment-Based Approach. *J. Comput. Chem.* **2019**, *40*, 1429–1439. [[CrossRef](#)]
22. Siebenmorgen, T.; Engelhard, M.; Zacharias, M. Prediction of Protein–Protein Complexes Using Replica Exchange with Repulsive Scaling. *J. Comput. Chem.* **2020**, *41*, 1436–1447. [[CrossRef](#)]
23. Siebenmorgen, T.; Zacharias, M. Efficient Refinement and Free Energy Scoring of Predicted Protein–Protein Complexes Using Replica Exchange with Repulsive Scaling. *J. Chem. Inf. Model.* **2020**, *60*, 5552–5562. [[CrossRef](#)] [[PubMed](#)]
24. Maszota-Zieleniak, M.; Marcisz, M.; Kogut, M.M.; Siebenmorgen, T.; Zacharias, M.; Samsonov, S.A. Evaluation of Replica Exchange with Repulsive Scaling Approach for Docking Glycosaminoglycans. *J. Comput. Chem.* **2021**, *42*, 1040–1453. [[CrossRef](#)] [[PubMed](#)]
25. Hahne, M.; Kataoka, T.; Schröter, M.; Hofmann, K.; Irmeler, M.; Bodmer, J.L.; Schneider, P.; Bornand, T.; Holler, N.; French, L.E.; et al. APRIL, a New Ligand of the Tumor Necrosis Factor Family, Stimulates Tumor Cell Growth. *J. Exp. Med.* **1998**, *188*, 1185–1890. [[CrossRef](#)]
26. Hendriks, J.; Planelles, L.; de Jong-Odding, J.; Hardenberg, G.; Pals, S.T.; Hahne, M.; Spaargaren, M.; Medema, J.P. Heparan Sulfate Proteoglycan Binding Promotes APRIL-Induced Tumor Cell Proliferation. *Cell Death Differ.* **2005**, *12*, 637–648. [[CrossRef](#)]
27. Ingold, K.; Zumsteg, A.; Tardivel, A.; Huard, B.; Steiner, Q.G.; Cachero, T.G.; Qiang, F.; Gorelik, L.; Kalled, S.L.; Acha-Orbea, H.; et al. Identification of Proteoglycans as the APRIL-Specific Binding Partners. *J. Exp. Med.* **2005**, *201*, 1375–1383. [[CrossRef](#)]
28. Baert, L.; Benkhoucha, M.; Popa, N.; Ahmed, M.C.; Manfroi, B.; Boutonnat, J.; Sturm, N.; Raguenez, G.; Tessier, M.; Casez, O.; et al. A Proliferation-Inducing Ligand–Mediated Anti-Inflammatory Response of Astrocytes in Multiple Sclerosis. *Ann. Neurol.* **2019**, *85*, 406–420. [[CrossRef](#)]
29. Kimberley, F.C.; Medema, J.P.; Hahne, M. APRIL in B-Cell Malignancies and Autoimmunity. *Results Probl. Cell Differ.* **2009**, *49*, 161–182. [[CrossRef](#)]

30. Kimberley, F.C.; Van Bostelen, L.; Cameron, K.; Hardenberg, G.; Marquart, J.A.; Hahne, M.; Medema, J.P. The Proteoglycan (Heparan Sulfate Proteoglycan) Binding Domain of APRIL Serves as a Platform for Ligand Multimerization and Cross-Linking. *FASEB J.* **2009**, *23*, 1584–1595. [[CrossRef](#)] [[PubMed](#)]
31. Huard, B.; McKee, T.; Bosshard, C.; Durual, S.; Matthes, T.; Myit, S.; Donze, O.; Frossard, C.; Chizzolini, C.; Favre, C.; et al. APRIL Secreted by Neutrophils Binds to Heparan Sulfate Proteoglycans to Create Plasma Cell Niches in Human Mucosa. *J. Clin. Investig.* **2008**, *118*, 2887–2895. [[CrossRef](#)] [[PubMed](#)]
32. Bischof, D.; Elsawa, S.F.; Mantchev, G.; Yoon, J.; Michels, G.E.; Nilson, A.; Sutor, S.L.; Platt, J.L.; Ansell, S.M.; Von Bulow, G.; et al. Selective Activation of TACI by Syndecan-2. *Blood* **2006**, *107*, 3235–3242. [[CrossRef](#)]
33. Moreaux, J.; Sprynski, A.C.; Dillon, S.R.; Mahtouk, K.; Jourdan, M.; Ythier, A.; Moine, P.; Robert, N.; Jourdan, E.; Rossi, J.F.; et al. APRIL and TACI Interact with Syndecan-1 on the Surface of Multiple Myeloma Cells to form an Essential Survival Loop. *Eur. J. Haematol.* **2009**, *83*, 119–129. [[CrossRef](#)]
34. Sakurai, D.; Hase, H.; Kanno, Y.; Kojima, H.; Okumura, K.; Kobata, T. TACI Regulates IgA Production by APRIL in Collaboration with HSPG. *Blood* **2007**, *109*, 2961–2967. [[CrossRef](#)]
35. Kowalczyk-Quintas, C.; Willen, D.; Willen, L.; Golob, M.; Schuepbach-Mallepell, S.; Peter, B.; Eslami, M.; Vigolo, M.; Broly, H.; Samy, E.; et al. No Interactions between Heparin and Atacicept, an Antagonist of B Cell Survival Cytokines. *Br. J. Pharmacol.* **2019**, *176*, 4019–4033. [[CrossRef](#)]
36. Hymowitz, S.G.; Patel, D.R.; Wallweber, H.J.A.; Runyon, S.; Yan, M.; Yin, J.P.; Shriver, S.K.; Gordon, N.C.; Pan, B.; Skelton, N.J.; et al. Structures of APRIL-Receptor Complexes: Like BCMA, TACI Employs Only a Single Cysteine-Rich Domain for High Affinity Ligand Binding. *J. Biol. Chem.* **2005**, *280*, 7218–7227. [[CrossRef](#)] [[PubMed](#)]
37. Pichert, A.; Schlorke, D.; Franz, S.; Arnhold, J. Functional Aspects of the Interaction between Interleukin-8 and Sulfated Glycosaminoglycans. *Biomatter* **2012**, *2*, 142–148. [[CrossRef](#)]
38. Kirschner, K.N.; Yongye, A.B.; Tschampel, S.M.; González-Outeiriño, J.; Daniels, C.R.; Foley, B.L.; Woods, R.J. GLYCAM06: A Generalizable Biomolecular Force Field. *Carbohydrates. J. Comput. Chem.* **2008**, *29*, 622–655. [[CrossRef](#)] [[PubMed](#)]
39. Huige, C.J.M.; Altona, C. 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. [[CrossRef](#)]
40. Samsonov, S.A.; Pisabarro, M.T. Computational Analysis of Interactions in Structurally Available Protein–Glycosaminoglycan Complexes. *Glycobiology* **2016**, *26*, 850–861. [[CrossRef](#)]
41. Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. *AAAI* **1996**, *96*, 226–231.
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**, *16*, 528–552. [[CrossRef](#)]
43. Case, D.A.; Ben-Shalom, I.Y.; Brozell, S.R.; Cerutti, D.S.; Cheatham, T.E., III; Cruzeiro, V.W.D.; Darden, T.A.; Duke, R.E.; Ghoreishi, D.; Gilson, M.K.; et al. *AMBER16*; University of California: San Francisco, CA, USA, 2018.
44. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An  $N \cdot \log(N)$  method for Ewald sums in large systems. *J. Phys. Chem.* **1993**, *98*, 10089. [[CrossRef](#)]
45. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, J.C. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of  $n$ -alkanes. *J. Comp. Phys.* **1977**, *23*, 327–341. [[CrossRef](#)]
46. Onufriev, A.; Case, D.A.; Bashford, D. Effective Born Radii in the Generalized Born Approximation: The Importance of Being Perfect. *J. Comput. Chem.* **2002**, *23*, 1297–1304. [[CrossRef](#)] [[PubMed](#)]
47. Homeyer, N.; Gohlke, H. Free Energy Calculations by the Molecular Mechanics Poisson–Boltzmann Surface Area Method. *Mol. Inform.* **2012**, *31*, 114–122. [[CrossRef](#)] [[PubMed](#)]
48. Gadhi, N.S.; Mancera, R.L. Free energy calculations of glycosaminoglycan–protein interactions. *Glycobiology* **2009**, *19*, 1103–1115. [[CrossRef](#)]
49. Nivedha, A.K.; Makeneni, S.; Foley, B.L.; Tessier, M.B.; Woods, R.J. Importance of Ligand Conformational Energies in Carbohydrate Docking: Sorting the Wheat from the Chaff. *J. Comput. Chem.* **2014**, *35*, 526–539. [[CrossRef](#)]

## Publication D6

Solvent models benchmark for molecular  
dynamics of glycosaminoglycans

# Solvent Model Benchmark for Molecular Dynamics of Glycosaminoglycans

Mateusz Marcisz\* and Sergey A. Samsonov\*



Cite This: *J. Chem. Inf. Model.* 2023, 63, 2147–2157



Read Online

ACCESS |



Metrics & More

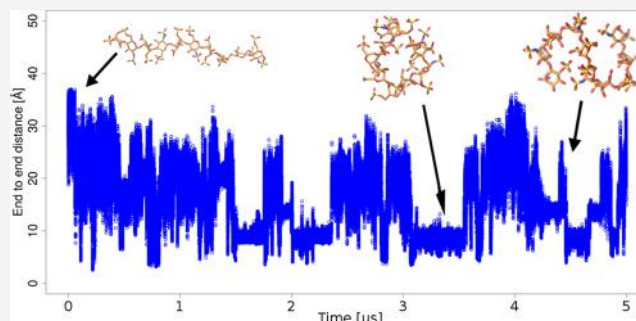


Article Recommendations



Supporting Information

**ABSTRACT:** In computational studies of glycosaminoglycans (GAGs), a group of anionic, periodic linear polysaccharides, so far there has been very little discussion about the role of solvent models in the molecular dynamics simulations of these molecules. Predominantly, the TIP3P water model is commonly used as one of the most popular explicit water models in general. However, there are numerous alternative explicit and implicit water models that are neglected in the computational research of GAGs. Since solvent-mediated interactions are particularly important for GAG dynamic and structural properties, it would be of great interest for the GAG community to establish the solvent model that is suited the best in terms of the quality of theoretically obtained GAG parameters and, at the same time, would be reasonably demanding in terms of computational resources required. In this study, heparin (HP) was simulated using five implicit and six explicit solvent models with the aim to find out how different solvent models influence HP's molecular descriptors in the molecular dynamics simulations. Here, we initiate the search for the most appropriate solvent representation for GAG systems and we hope to encourage other groups to contribute to this highly relevant subject.



## INTRODUCTION

Glycosaminoglycans (GAGs) are a group of anionic polysaccharides that are built of repeating disaccharide units.<sup>1</sup> They are long periodic linear sulfated and highly charged polymers that exhibit different sulfation patterns, which may alter their binding and functional properties as well as conformational characteristics.<sup>2,3</sup> There are six groups of mammalian GAGs: heparan sulfate (HS), heparin (HP), hyaluronic acid (HA), chondroitin sulfate (CS), dermatan sulfate (DS), and keratan sulfate (KS). Depending on the arrangement of the disaccharide units in those molecules, they may display 408 variants,<sup>1</sup> of which 202 can be found in mammals.<sup>4,5</sup> Their diversity contributes to the multifunctional role they play in the extracellular matrix of the cell affecting different types of processes such as cell signaling,<sup>6</sup> cardiovascular diseases,<sup>7</sup> cell maturation,<sup>8</sup> tissue regeneration,<sup>9,10</sup> proliferation,<sup>11</sup> inflammatory response,<sup>12</sup> infection,<sup>6,13</sup> and diseases such as Alzheimer's and Parkinson's diseases<sup>14</sup> or cancer development.<sup>15,16</sup> It has been shown recently that GAGs also play a crucial role in the Sars-Cov-2 viral infection mechanism.<sup>17–20</sup> All of these mentioned processes are affected directly by GAG intermolecular interactions with proteins: chemokines,<sup>21–23</sup> growth factors,<sup>24–26</sup> morphogens,<sup>27</sup> membrane receptors integrins,<sup>28</sup> and lipoproteins.<sup>29</sup> The majority of these interactions are driven by electrostatics and are nonspecific,<sup>30,31</sup> while some of them, in contrast, can be highly specific<sup>32</sup> or selective.<sup>33</sup> The properties of the GAGs make it extremely challenging to study them using computational tools.<sup>34</sup> Additionally, despite some of the recent

advances in the field of GAG docking,<sup>35–40</sup> it still lags behind the abundance of approaches designed for small drug molecules. Although at the moment, there are even several web servers that could be used for GAG docking,<sup>41–43</sup> they do not perform as well as some conventional docking software (e.g., DOCK and Autodock3) for the highly heterogeneous data set of protein–GAG complexes.<sup>44</sup>

Most of the molecular dynamics (MD)-related studies on GAGs are conducted using the TIP3P water model as it is widely accepted in the GAG field and proven to be working well in the protein–GAG systems,<sup>36,45–48</sup> as well as in MD studies of diverse biomolecular systems in general. The basic reason for the wide use of the TIP3P and other three-site water models is their low computational cost compared to four- and five-site solvent models. Implicit water models are used less frequently, especially nowadays when researchers have easier access to high-performance computing facilities than before. However, due to the lower computational costs of implicit solvent, this type of water model is still utilized when computational resources and time are limited or the size of the studied system is particularly

Received: November 22, 2022

Published: March 29, 2023



big. Regarding the impact of the solvent type in computational science, some studies show that explicit solvent can improve the general quality of docking and MD simulations,<sup>36,49–53</sup> while there is an evidence that GB models do not reproduce secondary structures of *de novo* designed peptides.<sup>54</sup> To this day, there has been no exhaustive comparative study conducted on water models that supposes to answer the question which one has advantages to be used in systems involving GAG molecules. Previously, to our knowledge, there was one attempt to compare HP properties in the MD simulations with SPC and SPC/E water models.<sup>55</sup> However, this study was conducted in GROMACS as opposed to AMBER used in this work with GLYCAM06.<sup>56</sup> Additionally, the simulations were only 3 ns long, which is not enough to observe major conformational changes and which is far shorter than the state-of-the-art MD simulation length established standards nowadays. While there has been a general opinion based on common sense and studies not involving GAGs<sup>57,58</sup> that it would be beneficial to use a more advanced water model than three-site TIP3P, e.g., TIP4P or TIP5P, there are no data on the use of different water models and thus no evidence which and if other model should be commonly used. This discussion is of immense importance due to the fact that in the available experimental structures half of the protein–GAG residue contacts are mediated by water and that there are about 10 times more water molecules in the protein–GAG interfaces than in the protein–protein interfaces.<sup>59</sup> Thus, it is necessary to accurately model water-mediated interactions in protein–GAG complexes, especially taking into account the importance of the electrostatic interactions involved there.<sup>49</sup> In fact, it was shown that interactions between proteins and GAGs can be stabilized by solvent molecules functioning as structured water in those complexes.<sup>33,60–63</sup> It was also reported that the effect of the dynamical behavior of the solvent surrounding GAGs on the conformation of the saccharides is of great relevance.<sup>64–66</sup> Taking all of this information into account, it would be beneficial for the computational GAG community to study the influence of the water models on the GAG behavior in the MD simulation.

In this work, we aim to evaluate the properties of the HP in different implicit and explicit water models to find out which of them are best suited for the MD simulation of GAG molecules. For this, 5  $\mu$ s MD simulations were performed involving dp10 (dp stays for the degree of polymerization) HP molecule with the following water models: implicit IGB = 1,<sup>67</sup> 2,<sup>68</sup> 5,<sup>69</sup> 7,<sup>70</sup> and 8,<sup>71</sup> and explicit TIP3P,<sup>72,73</sup> SPC/E,<sup>74</sup> TIP4P,<sup>72</sup> TIP4PEw,<sup>75</sup> OPC,<sup>57</sup> and TIP5P.<sup>76</sup> Additionally, five 200 ns MD simulations were performed for each of the setups as a consistency check for the convergence of the analyzed parameters. Using the obtained trajectories, HP properties such as end-to-end distance (EED), volume, radius of gyration, ring puckering, intramolecular hydrogen bonds, and dihedral angles were analyzed.

## MATERIALS AND METHODS

**Heparin Structure and Parameterization for the MD Simulations.** The initial structure of the HP dp10 used in this study was obtained from the Protein Data Bank,<sup>77</sup> 1HPN structure.<sup>78</sup> Literature data for the sulfate group charges<sup>79</sup> and GLYCAM06<sup>56</sup> force field parameters were used. <sup>1</sup>C<sub>4</sub> conformation for the IdoA2S ring was chosen as it was shown to be the essentially dominant conformation in the microsecond scale simulations performed by Sattelle *et al.* as it is energetically more favorable than the <sup>2</sup>S<sub>0</sub> conformation.<sup>80</sup> Additionally, another study conducted by Bojarski *et al.* claims that the IdoA2S <sup>1</sup>C<sub>4</sub>

pucker conformation reproduces a more probable and extended HP structure than the <sup>2</sup>S<sub>0</sub> conformation.<sup>81</sup> The same study indicates that <sup>1</sup>C<sub>4</sub> is the preferable conformation and that when starting from the <sup>2</sup>S<sub>0</sub> conformation, it changes to the <sup>1</sup>C<sub>4</sub> one during the long MD simulation. Moreover, NMR studies also find that IdoA2S <sup>1</sup>C<sub>4</sub> conformation in HP oligosaccharides is dominant.<sup>82,83</sup> Additionally, it was reported that HP's IdoA2S ring conformation has little impact on other properties (such as radius of gyration and end-to-end distance) of the HP molecule.<sup>84</sup>

**Water Models.** All of the parameters for the water models used in this work are taken from Amber16 and recommended mbondi were used for each particular model. The models used were implicit IGB = 1, 2, 5, 7, 8 and explicit TIP3P, SPC/E, TIP4P, TIP4PEw, OPC, and TIP5P. Graphical representation and the details of the solvent models used in this study are presented in Figure S1.

**MD Simulations.** MD simulations have been carried out in the AMBER package.<sup>85</sup> HP decasaccharide was solvated in an octahedral periodic box with a minimum distance between solute and box edge of 8.0 Å and neutralized with counterions (Na<sup>+</sup>) in the case of the explicit solvent simulations. No “saltcon” option was used in the implicit solvent simulations. Two energy minimization steps were carried out (first  $1.5 \times 10^3$  steepest descent cycles and  $10^3$  conjugate gradient cycles with harmonic force restraints on solute atoms, followed by  $6 \times 10^3$  steepest descent cycles and  $3 \times 10^3$  conjugate gradient cycles without restraints) for the explicit solvent simulations, while only the second minimization step was performed for the implicit solvent simulations. Subsequently, the system was heated up to 300 K for 10 ps with harmonic force restraints of  $100 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$  on solute atoms, and equilibration for 50 ps at 300 K and  $10^5$  Pa in the isothermal isobaric ensemble (NPT) for the explicit solvent simulation. A productive MD run was also carried out in an NPT ensemble in the explicit solvent simulations. The SHAKE algorithm, 2 fs time integration step, 8 Å cutoff for nonbonded interactions, and the particle mesh Ewald method were used.

**Molecular Mechanics Generalized Born Surface Area (MM/GBSA) Analysis.** Molecular mechanics generalized Born surface area (MM/GBSA) model IGB = 2 from AMBER16 was used for free energy calculations on the trajectories obtained from MD simulations in the case of the explicit water models. For the implicit water models, IGB values were chosen according to the model used in MD (e.g., IGB = 1 for IGB = 1 water model).<sup>88</sup>

**Heparin's Properties.** The radius of gyration, end-to-end distance (EED), hydrogen bonds, ring puckers, and dihedral angles were obtained using cpptraj scripts from the AMBER suite. The volume of the HP molecule (similar to the work of Nagarajan *et al.*<sup>89</sup>) was calculated using an in-house script. For this, we extracted the molecule's coordinates from each frame of the trajectory, and afterward, the volume was assessed using mvee\_REX (minimum-volume enclosing ellipsoid) from the OptimalDesign library in the R.<sup>90</sup>

The dihedral angles for glycosidic linkage analysis were defined as  $\text{O5}_{n+1}-\text{C1}_{n+1}-\text{O4}_n-\text{C4}_n$  and  $\text{C1}_{n+1}-\text{O4}_n-\text{C4}_n-\text{C3}_n$ , where  $n$  stands for the sequential number of a sugar monomeric unit. The dihedral angles for the ring puckering were defined as  $\text{C1}_n-\text{C2}_n-\text{C3}_n-\text{C4}_n$  and  $\text{C1}_n-\text{O5}_n-\text{C5}_n-\text{C4}_n$ . AMBER's atom and residue numbering and nomenclature are used for the glycosidic linkage and ring puckering definitions.

Visual representation of the dihedral angles (glycosidic linkages and puckering) is shown in Figure S2.

## RESULTS

In order to observe changes in HP properties depending on the used water model, we performed 5  $\mu$ s MD simulations for all of the investigated models. Then, additionally, we supplemented the data with shorter (five 200 ns for each model) simulations in order to check the consistency of the results and to assure the reproducibility of those simulations. Those relatively short simulations should allow us to compare HP conformational changes between different timescales.

HP behaved drastically differently in implicit and explicit water models. Naturally, there was a significant contrast between particular implicit and explicit water models, but those changes were less significant than when comparing GB implicit models with the explicit models (Tables 1, S1, and S2). The obtained data were compared to the experimental findings of Mulloy *et al.*,<sup>78</sup> Khan *et al.*,<sup>86</sup> and Pavlov *et al.*<sup>87</sup>

Overall end-to-end distances (Figure 1) that were observed during 5  $\mu$ s are about 2 times bigger and follow unimodal distribution in implicit models. This suggests notably more extended structures in implicit water, especially in IGB = 8. On the contrary, the most curved HP structures were observed in the TIP3P model simulation. Additionally, we observed some drastic and sudden conformational changes in the case of the IGB = 7 model where HP changed the conformation from the extended structure to an “O”-shaped structure (Figure S3) for about 1  $\mu$ s and then returned back to the extended conformation. This model is known to be not suited for nucleic acids<sup>85</sup> because Coulomb field approximation (CFA) to define the integral used to calculate effective radii could lead to numerical errors.<sup>70</sup> We suppose that for HP, which is also highly negatively charged, the reason for the obtained dramatic unnatural structural deformation could be the same.

TIP5P and OPC allowed for the most extended HP structure among the explicit solvent models, which was the closest to the experimental data (41 Å) from the 1HPNPDB structure.

Additionally obtained data set was confronted with different experimentally measured EEDs.<sup>86,87</sup> These EED values are gathered in Table S3. They show that higher curvature was observed only in some of the longer (>dp30) HP molecules. However, when normalized per dp10 (as the length of the HP molecule investigated in this work), the lowest EED was 27.6 Å and the next lowest was 32.3 Å. This is still far away from the values obtained in the TIP3P (16.1 ± 6.4) water model and comparable with the values of HP in the OPC (28.4 ± 4.5 Å) and TIP5P (26.1 ± 4.7 Å) models.

As expected, for the more extended HP structure, the radius of gyration (Table 1 and Figure S4) was higher in GB models, especially in IGB = 5 and 8. The lowest radius of gyration was observed in TIP3P (8.8 Å) water, which is also in agreement with end-to-end distance data. Stand-out explicit water models were OPC (12.4 Å) and TIP5P (11.6 Å) where the radius of gyration of HP was closer to the range of values observed in implicit water models rather than in other explicit ones, which is more in agreement with the findings of Mulloy *et al.*<sup>78</sup> (12.8 Å) and Khan *et al.* (15.3 Å when normalized for dp10 – 18.3 Å for dp12).<sup>86</sup>

Next, RMSD to the starting point of the MD, which at the same time is the RMSD to the experimental structure since this structure was used as the starting point of the MD runs, was compared. Much higher RMSD (Table 1) was observed in

Table 1. HP Descriptors Obtained from 5  $\mu$ s MD Simulations

	PDB 1HPN	IGB = 1	IGB = 2	IGB = 5	IGB = 7	IGB = 8	TIP3P	SPC/E	TIP4P	TIP4PEw	OPC	TIP5P
dist <sup>a</sup> (Å)	41.0	40.4 ± 2.9	41.6 ± 2.7	40.4 ± 3.1	38.1 ± 13.3	42.4 ± 2.2	16.1 ± 6.4	20.7 ± 7.6	17.3 ± 7.9	21.2 ± 6.0	28.4 ± 4.5	26.1 ± 4.7
fluct <sup>b</sup> (Å)	N/A <sup>d</sup>	3.5 ± 0.7	3.4 ± 0.6	3.5 ± 0.8	4.8 ± 1.6	3.8 ± 0.4	4.8 ± 1.4	5.3 ± 1.6	4.6 ± 1.9	5.9 ± 1.3	4.5 ± 1.0	4.7 ± 1.2
RMSD (Å)	N/A <sup>d</sup>	4.2 ± 0.7	4.9 ± 0.7	4.0 ± 0.8	5.2 ± 2.9	5.9 ± 0.4	9.0 ± 1.5	7.4 ± 2.0	7.9 ± 2.2	8.7 ± 2.6	4.9 ± 1.4	5.7 ± 1.4
radgyr <sup>c</sup> (Å)	12.8	13.6 ± 0.4	14.1 ± 0.3	14.0 ± 0.4	13.0 ± 3.0	14.0 ± 0.3	8.8 ± 1.0	9.9 ± 1.5	9.0 ± 1.6	9.5 ± 1.6	12.4 ± 1.0	11.6 ± 1.2
volume (MVVE) (Å <sup>3</sup> )	5325	7974 ± 875	8079 ± 841	8311 ± 913	7042 ± 1873	7674 ± 653	6085 ± 787	6551 ± 1032	5853 ± 1140	6902 ± 779	7685 ± 875	7641 ± 890

<sup>a</sup>End-to-end distance. <sup>b</sup>Radius of gyration. <sup>c</sup>Atomic fluctuations and root-mean-square deviation (RMSD) are compared to the 1HPN structure; therefore, the values are 0.

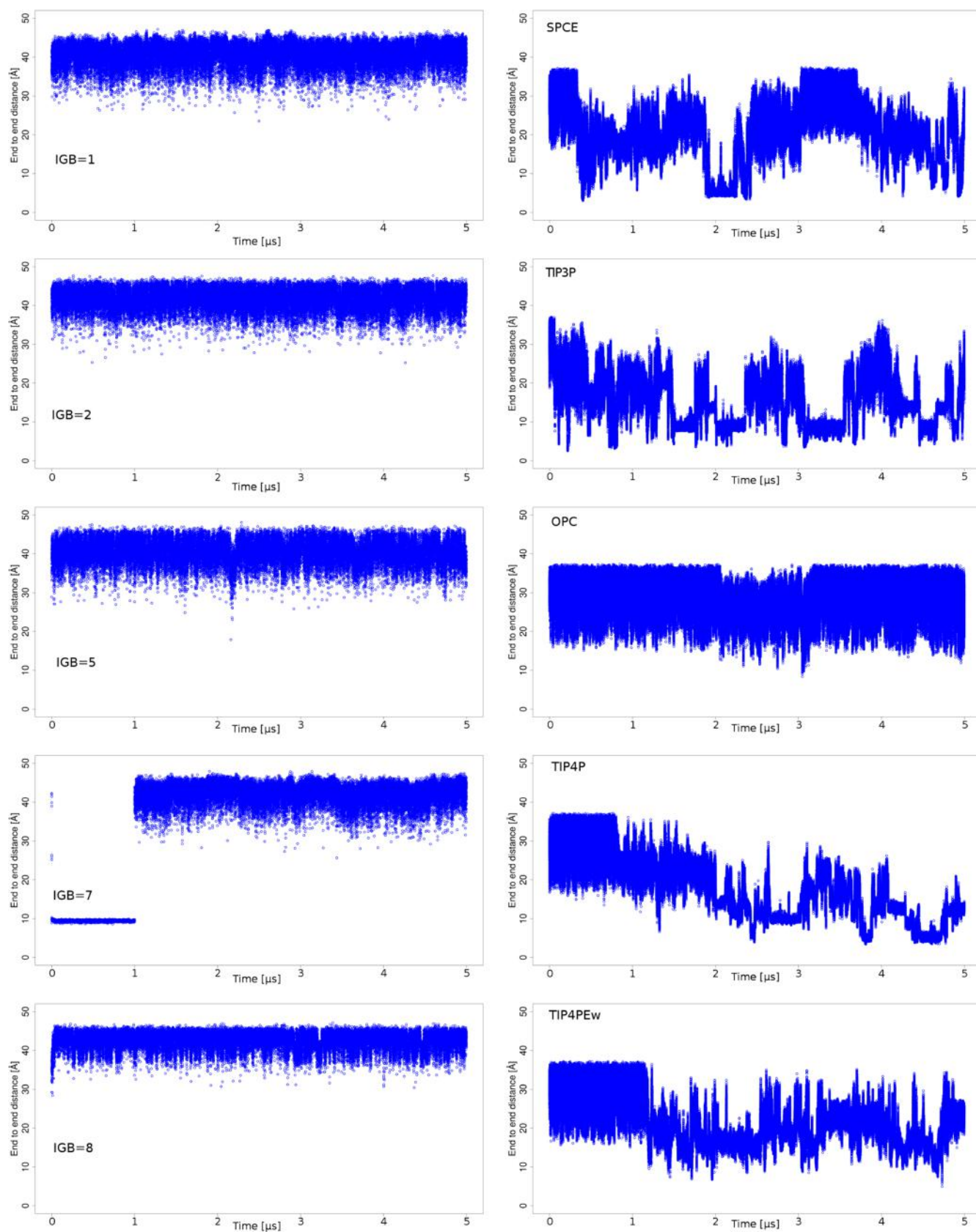
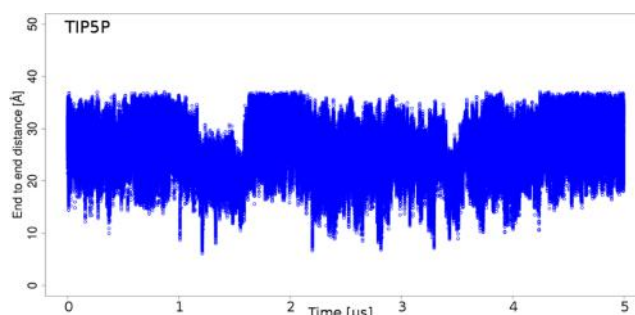
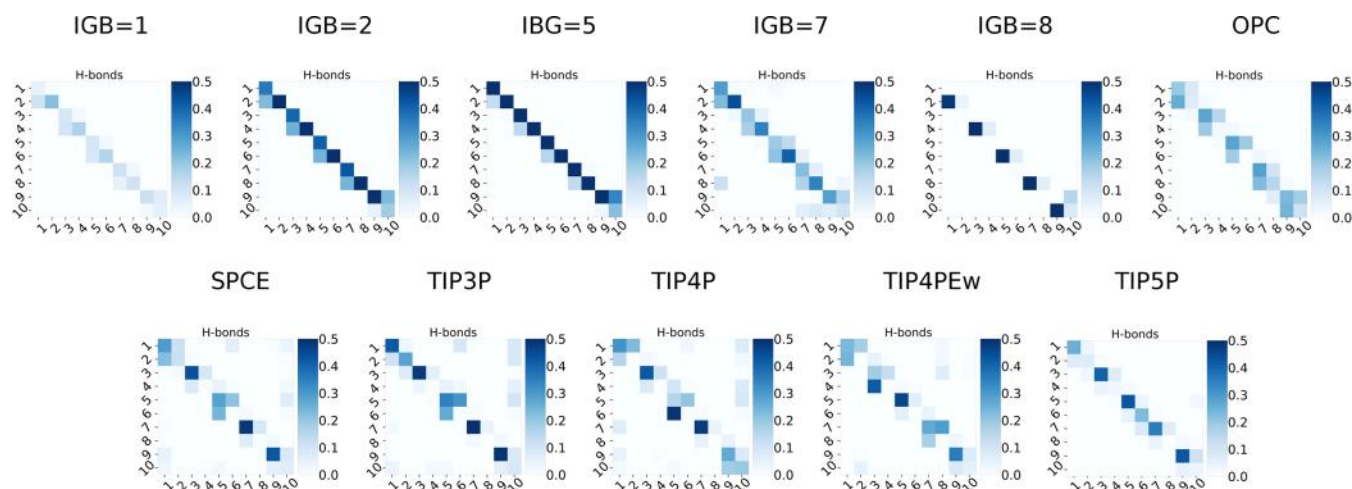


Figure 1. continued





**Figure 1.** End-to-end distances in the course of 5  $\mu\text{s}$  MD simulations showing differences in HP behavior in different water models. More packed structures can be observed in explicit water models.



**Figure 2.** Intramolecular hydrogen bonds established between atoms belonging to the residues in HP (the residue numbers are indicated in both axis labels). More interactions between atoms within the same residue observed in implicit solvent models correspond to the extended conformation. For the explicit solvent models, interactions between different residues can be observed more often suggesting more packed conformations. Dark blue indicates more contacts between residues, light blue shows fewer hydrogen bonds formed, while white space indicates no contacts.

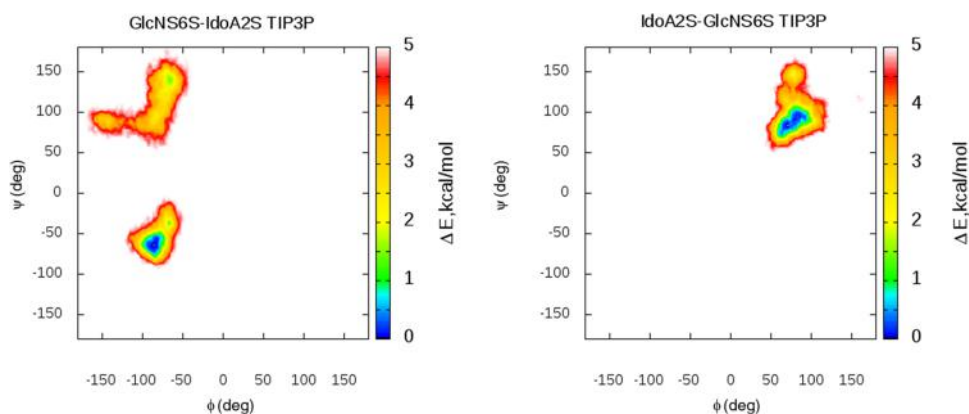
explicit water models than implicit models when compared to the experimental structure (1HPN),<sup>78</sup> suggesting essentially higher conformational diversity of HP in the former ones. The lowest RMSD was observed in IGB = 5 and IGB = 1 while the highest in TIP3P and TIP4PEw. Again, OPC and TIP5P showed values within the range of those of implicit models rather than within other explicit ones. In those two models, the smallest deviation of the HP structure from the crystal structure was observed among explicit solvent models.

Atomic fluctuations (Table 1) followed a similar trend of the RMSD changes. Again, we observed high values in the case of the TIP4PEw model. However, the difference to the RMSD data here is that in the case of the TIP3P model, values were medium within the explicit models. In the case of implicit models, lowest fluctuations were observed in IGB = 1, 2, and 5.

Additionally, we measured the volume of HP (Table 1 and Figure S5). This volume is defined as a space occupied by the molecule's atoms and can be used to detect substantial shape/conformational changes of GAG in the MD simulations.<sup>89</sup> It may be partially dependent on the end-to-end distance and radius of gyration—smaller end-to-end distance and radius of gyration may in some cases mean smaller volume as the molecule in this case is more compact. Details on the way of calculating HP volume are disclosed in the Materials and Methods section. The smallest volume was observed in the case of TIP3P and TIP4P. In the group of explicit water models, HP occupied the largest volume in OPC and TIP5P models and

those values were similar to the values observed in the implicit water models. The highest occupied volume was observed in the IGB = 5 water model. However, none of the water models used in this study managed to reproduce experimental findings observed in the 1HPN structure with high accuracy.

As a next step in the analysis, intramolecular hydrogen bonds were analyzed (Figure 2). In cases when we observed more extended HP structures (GB models and TIP5P), the hydrogen bonds were mostly formed between the atoms belonging to the same residue. However, in the MD runs, when more conformational changes occurred, hydrogen bonds established between atoms in different residues are more often observed. In the case of TIP3P and TIP4P hydrogen bonds between the first and the last GAG residues are formed which indicates a significant amount of the “U”-shaped (Figure S3) HP conformational population during MD simulations, which deviates from the structures observed experimentally. Such conformations are unlikely to occur due to the highly charged nature of the HP molecule, which would rather exclude close proximity of its negatively charged chemical groups in the space. Potentially, such contacts are established in the MD simulation with the TIP3P solvent due to its electrostatic properties allowing for the clustering of counterions that bridge the negatively charged HP groups, which to such an extent was not the case for other, more complex, explicit solvent models (TIP5P and OPC). This phenomenon partially could be explained by similar findings obtained by Pluhařová *et al.*,<sup>91</sup>



**Figure 3.** GlcNS6S-IdoA2S and IdoA2S-GlcNS6S glycosidic linkage free energy heatmaps for the TIP3P case. (Glycosidic linkages of the HP molecule in other water models are shown in Figure S4).

where Li2S04 were clustered and closely packed in the explicit solvent (SPC/E model). The observed artifact was caused by strong electrostatic interactions.

Furthermore,  $\phi$  and  $\psi$  dihedral angles for the glycosidic linkages of HP were analyzed (Figures 3 and S6). The heatmaps were obtained by summing up all GlcNS6S-IdoA2S and IdoA2S-GlcNS6S glycosidic linkage-type populations for each of the water models tested. In the case of the widely used TIP3P water, we observe two minima for the GlcN6S-IdoA2S linkage (first  $\phi \sim -80^\circ$  and  $\psi \sim -50^\circ$  and second  $\phi \sim -80^\circ$  and  $\psi \sim 100^\circ$ ) that match experimental/theoretical data<sup>80</sup> and the computational findings of Bojarski *et al.*<sup>81</sup> where HP was simulated for 10  $\mu$ s in complex with a protein. The same applies for the IdoA2S-GlcN6S linkage where we observe a single minimum ( $\phi \sim 80^\circ$  and  $\psi \sim 100^\circ$ ). For other water models (Figure S6), similar patterns on the heatmaps are observed. However, IGB = 7 data look very concerning as we observe a multiplicity of the various minima that are not present in any of the data known to us (experimental or theoretical). Similar to the work of Holmes *et al.*<sup>92</sup> performed for heparan sulfate oligosaccharides, we could see that the first GlcN6S-IdoA2S glycosidic linkage minimum ( $\phi \sim -80^\circ$  and  $\psi \sim -50^\circ$ ) corresponds to the increased curvature of HP, and it is much more pronounced for the explicit solvent where more curved structures are observed. This minimum is barely visible in the heatmaps of the glycosidic linkages in the implicit solvent. The second minimum ( $\phi \sim -80^\circ$  and  $\psi \sim 100^\circ$ ), which corresponds to the more extended structure, has more favorable energies for the linkages in the implicit solvents. Additionally, for the explicit solvent, this second minimum has more favorable energies in the case of OPC and TIP5P models, and those are the two explicit models where essentially less curvature is observed.

Ring puckering was analyzed to check the effects of a water model that was used on the ring puckering conformations in HP (Tables S4 and S5). In the explicit water environment for the GlcNS6S rings, almost exclusively  ${}^4C_1$  conformation was observed. The exception were the terminal rings of the GlcNS6S HP unit where some population (up to 13% of the frames) of the  ${}^1C_4$  conformation was observed. In case of IdoA2S, the  ${}^1C_4$  conformation was dominant. However, a significant amount of the frames with the  ${}^4C_1$  ring conformation could be observed (up to 74% in case of the first IdoA2S ring in the TIP3P water model) at the ends of the HP molecule (it is a known issue of the used force field that may cause disruption of the ring conformation at termini of the HP molecule, especially for

IdoA derivatives). Very rarely (0–2%),  ${}^2S_0$  conformation was present. In the implicit water model, all HP rings independently of their type almost exclusively adopted  ${}^4C_1$  conformation. The exception are IdoA2S residues at the nonreducing end of the HP where 80–100% of the conformations were  ${}^1C_4$ . Another exception is IGB = 1 where 30–100% of the IdoA2S have proper  ${}^1C_4$  conformation. These data show HP revealing appropriate ring conformations in the explicit solvent (with the exception of the terminal saccharide units of the HP molecule), while in the majority of the implicit water models, they do not reproduce experimental data as all of the IdoA2S ring conformations are disrupted.

End-to-end distance, the radius of gyration, and volume of HP suggest in general more packed structures in the explicit water models. One of the possible reasons for this is the presence of  $\text{Na}^+$  counterions in explicit solvent MD simulations, while these ions are missing in the implicit solvent simulations. In fact, when checked the position of the  $\text{Na}^+$  ions, majority of them were observed in the direct proximity of HP (Figure S3). We suppose that the presence of the ions may favor the curvature of HP due to  $\text{Na}^+$  interactions with sulfate groups. It was previously reported that ions could stabilize/favor curved structure of other GAGs.<sup>92,93</sup>

To look for the potential changes at a shorter timescale of the MD simulation and to check for the consistency of the calculations, five repeats of the 200 ns MD runs for each of the tested water models were performed. Although overall results were similar (Figures S5 and S6 and Table S1), some differences were observed for the analyzed descriptors in the shorter MD runs. In the case of implicit water models, there were almost no differences between those five 200 ns simulations and 5  $\mu$ s MD. One exception was IGB = 7 where in some cases we observed unusual behavior of HP that led to the crash of the MD run. The major differences were found for the explicit models. This is due to the fact that the structural features described—stronger curvature, lower volume, and end-to-end distance—were visible after hundreds of ns or even few  $\mu$ s of the MD run. This is the reason why in the shorter simulations in general more extended HP structures were present in comparison to the long MD. In the past, most of the computational studies regarding GAGs were conducted in a shorter timescale. Nowadays, with the availability of stronger computational resources, more advanced longer simulations become common. Before, changes in the long-scale simulations could have been overlooked as they were very rarely investigated. Therefore, it is necessary to put

more emphasis on the behavior of the HP in a longer timescale. In this case, TIP3P may not be the best solvent model as after hundreds of nanoseconds, some of the HP properties changed dramatically and did not anymore agree with the experiment.

Additionally, MM/GBSA was applied to investigate the energetic properties of the HP molecule in the surrounding of different water solvent types (Table 2). Most of the HP–water complexes showed energies around  $-1000$  kcal/mol. However, there were some outliers: the highest energies were found in TIP3P ( $-940$  kcal/mol) and TIP4P ( $-916$  kcal/mol) environment, while we observed suspiciously low energies with the IGB = 8 model ( $-1581$  kcal/mol). What is worth mentioning is the fact that obtained energies are in agreement with other observations. In most of the GB models, we observe low variation in the molecular descriptors of HP/small conformational changes during MD simulations, while the corresponding energies do not change much with time—they are the same in shorter 200 ns simulations as in long  $5 \mu\text{s}$  simulation for a respective water model. In the case of several explicit water models where the conformational changes were observed the energies are higher in a  $5 \mu\text{s}$  MD than in 200 ns ones where those changes did not yet occur. This is the case with TIP3P ( $-940$  vs  $-1006$  kcal/mol), SPC/E ( $-971$  vs  $-1006$  kcal/mol), TIP4P ( $-916$  vs  $-988$  kcal/mol), and TIP4PEw ( $-954$  vs  $-988$  kcal/mol). This could suggest that the extended conformation is a major one for both implicit and explicit solvent models. What is important in the case of explicit water models is that energy changes between 200 ns and  $5 \mu\text{s}$  MD in the OPC and TIP5P are the smallest. This again suggests that those two models are able to maintain more proper structural conformation in a longer timescale.

To sum up, both the local (dihedral angles, puckering) and the global (volume, end-to-end distance, fluctuations, radius of gyration) structural properties of HP in different water models are properly maintained only by TIP5P and OPC models in a longer timescale. Other explicit solvent models allow for proper modeling of the local parameters; however, they failed to maintain global features, which were reflected in the curved/kinked structure of the HP. On the other hand, implicit solvent models tend to preserve properly global parameters of the HP, but they fail to maintain local structural properties.

**Comparison of Solvent Models in the Analysis of Hyaluronic Acid and Chondroitin Sulfate.** Additionally, we checked the influence of TIP3P, OPC, and TIP5P water models on MD-derived parameters of hyaluronic acid (HA) and chondroitin sulfate 4 (CS-4). HA and CS-4 have lower charge than heparin; thus, the influence of the solvent model could be expected to be less significant. Considering that the unusual behavior of HP in three-site models could be caused by the fact that electrostatics, which is the driving force in GAGs' interactions with other molecules, between GAG molecule and ions may not be properly represented. In fact, the differences observed for TIP3P and OPC/TIP5P models in the case of GAGs with lower net charge are not significant (Table 3).

**Future Perspective for the Use of Water Models in GAG Studies.** Although TIP5P and OPC performed the best in this study when analyzing different HP properties, we urge to treat these results with caution. This study only investigated free HP in the solution and not in the presence of proteins, which potentially could affect the drawn conclusions. In our previous studies involving MD of protein–GAG complexes, we did not observe any unusual behavior of the GAG molecules in the TIP3P solvent. Moreover, TIP5P and OPC seem to only

Table 2. Energies Obtained from the MM/GBSA Analysis<sup>a</sup>

	5 $\mu\text{s}$ simulations										200 ns simulations												
	IGB = 1	IGB = 2	IGB = 5	IGB = 7	IGB = 8	TIP3P	SPC/E	TIP4P	TIP4PEw	OPC	TIP5P	IGB = 1	IGB = 2	IGB = 5	IGB = 7	IGB = 8	TIP3P	SPC/E	TIP4P	TIP4PEw	OPC	TIP5P	
1	$-1015 \pm 12$	$-1038 \pm 12$	$-989 \pm 12$	$-1130 \pm 141$	$-1581 \pm 12$	$-940 \pm 40$	$-971 \pm 32$	$-916 \pm 53$	$-954 \pm 38$	$-1007 \pm 15$	$-1005 \pm 18$	$-1015 \pm 12$	$-1038 \pm 12$	$-989 \pm 12$	$-1130 \pm 141$	$-1579 \pm 12$	$-995 \pm 20$	$-996 \pm 18$	$-981 \pm 26$	$-990 \pm 18$	$-1009 \pm 14$	$-1007 \pm 14$	$-1007 \pm 14$
2	$-1014 \pm 12$	$-1038 \pm 12$	$-988 \pm 12$	$-1444 \pm 43$	$-1580 \pm 12$	$-1009 \pm 15$	$-999 \pm 18$	$-889 \pm 17$	$-976 \pm 31$	$-1007 \pm 14$	$-1018 \pm 13$	$-1014 \pm 12$	$-1038 \pm 12$	$-988 \pm 12$	$-1505 \pm 38$	$-1581 \pm 11$	$-1008 \pm 15$	$-991 \pm 19$	$-972 \pm 32$	$-1009 \pm 14$	$-1009 \pm 14$	$-1016 \pm 13$	$-1016 \pm 13$
3	$-1016 \pm 11$	$-1039 \pm 12$	$-989 \pm 11$	$-1505 \pm 38$	$-1581 \pm 12$	$-1008 \pm 15$	$-1008 \pm 14$	$-984 \pm 23$	$-1001 \pm 16$	$-1008 \pm 14$	$-1016 \pm 13$	$-1016 \pm 11$	$-1039 \pm 12$	$-989 \pm 11$	$-1581 \pm 12$	$-1581 \pm 12$	$-1012 \pm 14$	$-981 \pm 26$	$-1002 \pm 16$	$-1008 \pm 14$	$-1008 \pm 14$	$-1016 \pm 13$	$-1016 \pm 13$
4	$-1016 \pm 12$	$-1038 \pm 12$	$-988 \pm 12$	$-1513 \pm 34$	$-1581 \pm 12$	$-1012 \pm 14$	$-1008 \pm 14$	$-981 \pm 26$	$-1002 \pm 16$	$-1009 \pm 14$	$-1016 \pm 13$	$-1016 \pm 12$	$-1038 \pm 12$	$-988 \pm 12$	$-1579 \pm 12$	$-1579 \pm 12$	$-995 \pm 20$	$-996 \pm 18$	$-990 \pm 18$	$-1009 \pm 14$	$-1009 \pm 14$	$-1007 \pm 14$	$-1007 \pm 14$
5	$-1016 \pm 12$	$-1038 \pm 12$	$-988 \pm 12$	$-1487$	$-1580$	$-1006$	$-1004 \pm 15$	$-988$	$-988$	$-1008$	$-1015$	$-1015$	$-1038$	$-988$	$-1487$	$-1580$	$-1006$	$-988$	$-988$	$-1008$	$-1008$	$-1015$	$-1015$
average	$-1015$	$-1038$	$-988$		$-1580$	$-1006$	$-1004$	$-988$	$-988$	$-1008$	$-1015$	$-1015$	$-1038$	$-988$		$-1580$	$-1006$	$-988$	$-988$	$-1008$	$-1008$	$-1015$	$-1015$

<sup>a</sup>Values are shown in kcal/mol. <sup>b</sup>No data gathered due to crash of the MD simulation.

Table 3. HA and CS-4 Descriptors Obtained from 5  $\mu$ s MD Simulations

	HA				CS-4			
	PDB	TIP3P	OPC	TIP5P	PDB	TIP3P	OPC	TIP5P
dist <sup>a</sup> (Å)	35.9	22.4 ± 5.7	22.5 ± 5.5	24.3 ± 4.9	35.9	23.0 ± 5.2	23.0 ± 5.0	23.2 ± 5.1
fluct <sup>b</sup> (Å)	N/A <sup>d</sup>	3.5 ± 1.0	3.1 ± 0.8	4.2 ± 1.3	N/A <sup>d</sup>	3.3 ± 1.1	3.1 ± 1.1	3.3 ± 1.1
RMSD (Å)	N/A <sup>d</sup>	4.0 ± 1.4	4.6 ± 0.7	4.7 ± 1.4	N/A <sup>d</sup>	4.0 ± 1.0	3.6 ± 1.0	4.1 ± 1.3
radgyr <sup>c</sup> (Å)	11.9	14.0 ± 1.0	14.1 ± 0.7	12.8 ± 1.1	11.9	13.8 ± 0.7	13.8 ± 0.7	13.4 ± 0.8
volume (MVVE) (Å <sup>3</sup> )	7426	7331 ± 1288	7353 ± 1158	7769 ± 1145	7375	8696 ± 1192	8620 ± 1211	9096 ± 1178

<sup>a</sup>End-to-end distance. <sup>b</sup>Atomic fluctuation. <sup>c</sup>Radius of gyration. <sup>d</sup>Atomic fluctuations and RMSD are compared between the experimental structures of HA and CS-4 (PDB IDs: 1HYA and 1C4S, respectively) and the analyzed structures, therefore, the values are 0.

provide improvement in the case of highly charged GAG molecules. In the case of GAGs with a lower negative charge, no significant differences were observed in their behavior in TIP3P and TIP5P or OPC. Therefore, more testing studies in this regard are required. Additionally, we do not find any strong thermodynamic rationale for justification of TIP5P/OPC improvement on the HP structural behavior in the available literature. It also has to be noted that GLYCAM06 was initially tested with TIP3P as a primary water model, which theoretically should provide optimal or at least satisfactory results in the simulations. Taking all of this into account, we recommend careful consideration of used water models in the GAG computational studies. We are certain that further studies are required to fully unravel the impact of the water models on the behavior and structural properties of the GAG molecules.

## CONCLUSIONS

The observed values of MD-based molecular descriptors for HP differed significantly between the groups of implicit and explicit water models used in the simulations. Some interesting features were observed, e.g., HP in OPC and TIP5P models behave similarly to the implicit water models maintaining global properties in agreement with the experimental data. The IGB = 7 model caused sometimes abnormal behavior of HP, and improper glycosidic linkage populations were observed. In general, the implicit solvent models do not even qualitatively agree with the experiment in terms of reproducing the HP ring puckering. Therefore, we do not suggest using implicit models for simulating GAGs unless the restraints are applied to maintain the rings in particular conformations reproducing specific experimental data. The most distinguishable model among the explicit ones was the TIP3P water model, one of the most commonly used models in biomolecular systems in general, in which HP displayed completely different properties than in the implicit models. In the case of TIP3P (and other explicit models, but to lesser extent) we observed characteristic curvature of HP as it was often “U”-shaped. It might be due to the fact that in the case of the explicit water model, there are counterions and thus heavily charged HP willingly interacts with those ions and trap them inside. The reason why we observe this notably less in the case of OPC and TIP5P models might be the fact that those models are much more complex and thus they can provide more appropriate charge distribution leading to a more physically relevant electrostatic description. It is of immense importance as TIP3P is a widely used water model in the studies of GAGs, but here, evidences were presented that in a longer (microsecond range) timescale, TIP3P may fail to maintain proper structural conformational ensemble of the HP molecule. When using all explicit solvent models, HP rings kept the proper conformations (ring puckering) in agreement with the experimental data, while in majority of the implicit models, IdoA2S rings revealed

improper conformations. In our previous research, we did not observe any unusual behavior of HP (or other GAG molecules) as extensive curvature, packed conformation, wrong glycosidic dihedral or pucker angles, etc. However, in all of our previous studies (and vast majority of the studies that are conducted by other groups), GAG molecules are simulated in complex with the proteins, or the MD simulations are significantly shorter than 5  $\mu$ s. In the former case, HP interacts with the charged protein and is unable to freely cluster the ions, which could potentially induce the bend. This could prevent the unusual behavior of the GAG in which case ions are localized inside the curved GAG structure. Therefore, we strongly believe that the effects of water models on HP properties need to be extensively studied together in the presence of protein molecules in a long MD simulation. Similarly, in the work of Neamtu *et al.*,<sup>94</sup> four- and five-site water models performed better than TIP3P in reproducing several structural features of chondroitin sulfate. However, in this study, the first layer of hydration was found to be better represented by the TIP3P model, and it was also found that TIP3P favors the direct intramolecular hydrogen bonding of chondroitin sulfate while the four-site and five-site models disfavor it. Additionally, when using the TIP5P model, a better-defined long-range hydration layer order around the hydroxyl groups of the sugar ring was observed compared with the other studied models. Within the nowadays standards, GAG molecules are almost exclusively simulated in the TIP3P model and it still remains necessary to confirm if it is the appropriate model for the GAG–protein simulations or to select a better one. In particular, this question is justified by considering the fact that in 5  $\mu$ s simulations, TIP3P failed to maintain proper global structural features of the HP, and the HP conformations deviated essentially from the one observed in the experiment. The only two solvent models that allowed for reproduction of both proper local and global structural features of the HP in the MD simulation were TIP5P and OPC. What also needs to be emphasized is the fact that the preliminary investigation of GAGs with lower charge than by HP did not reveal any significant differences in the GAG properties depending on the use of different water models (either TIP3P or TIP5P/OPC). Considering both results for the HP study and less charged GAGs, we would like to spark a discussion regarding the use of water models in theoretical GAG-related studies and whether any changes in the established approaches are required. Additionally, we would like to encourage other computational groups interested in GAGs to contribute to this potentially very important topic, especially considering this type of analysis of multiple GAG–protein complexes would be highly beneficial and could shed more light on this highly relevant subject in the GAG theoretical research.

## ■ ASSOCIATED CONTENT

### Data Availability Statement

All of the data (MD simulations) were obtained using AMBER suite (Amber20 and AmberTools). The data were then analyzed using R (statistics and plots), VMD (visualization of the obtained trajectories), and GIMP (figure preparation). The data used to generate MD trajectories are accessible in a publicly available repository Zenodo under the link <https://doi.org/10.5281/zenodo.7260781>. All software except the AMBER suite is free of charge. AMBER software can be obtained from <http://ambermd.org/GetAmber.php>. R can be downloaded from <https://www.r-project.org/>. GIMP can be downloaded from <https://www.gimp.org/>. VMD can be downloaded from <http://www.ks.uiuc.edu/Research/vmd/>.

### SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jcim.2c01472>.

Graphical representation of water model types used in this study together with their detailed description; visualization of the atoms chosen for the calculations of the dihedral angles and glycosidic linkages; visualization of the HP structures obtained during MD simulations; auxiliary graphs representing different parameters analyzed in this work; auxiliary tables gathering more information regarding investigated parameters of GAG molecules; and comparison of experimentally reported end-to-end distances (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Authors

Mateusz Marcisz – Faculty of Chemistry, University of Gdańsk, 80-308 Gdańsk, Poland; Intercollegiate Faculty of Biotechnology of UG and MUG, 80-307 Gdańsk, Poland; Email: [mateusz.marcisz@phdstud.ug.edu.pl](mailto:mateusz.marcisz@phdstud.ug.edu.pl)

Sergey A. Samsonov – Faculty of Chemistry, University of Gdańsk, 80-308 Gdańsk, Poland; [orcid.org/0000-0002-5166-4849](https://orcid.org/0000-0002-5166-4849); Email: [sergey.samsonov@ug.edu.pl](mailto:sergey.samsonov@ug.edu.pl)

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.jcim.2c01472>

### Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

### Funding

This research was funded by the National Science Centre of Poland (UMO-2018/30/E/ST4/00037) and supported by COST Action INNOGLY CA 18103, supported by COST (European Cooperation in Science and Technology).

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

The authors would like to thank Annemarie Danielsson for the technical assistance with preparation of the hydrogen bonds visualization. They would like to thank Umesh R. Desai (Virginia Commonwealth University) for the inspiration for the heparin volume calculations. The molecular dynamics simulations were performed on the PROMETHEUS cluster provided by Polish Grid Infrastructure (PL-GRID, grant names:

plggpugag software, plggagstr4) as well as on the local “piasek” cluster.

## ■ REFERENCES

- (1) Bu, C.; Jin, L. NMR Characterization of the Interactions Between Glycosaminoglycans and Proteins. *Front. Mol. Biosci.* **2021**, *8*, No. 646808.
- (2) Habuchi, H.; Habuchi, O.; Kimata, K. Sulfation Pattern in Glycosaminoglycan: Does It Have a Code? *Glycoconjugate. J.* **2004**, *21*, 47–52.
- (3) Varki, A.; Cummings, R. D.; Esko, J. D.; Stanley, P.; Hart, G. W.; Aebi, M.; Darvill, A. G.; Kinoshita, T.; Packer, N. H.; Prestegard, J. H.; Schnaar, R. L.; Seeberger, P. H. *Essentials of Glycobiology*, 3rd ed.; Cold Spring Harbor Laboratory Press: NY, 2015; pp 2015–2017.
- (4) Clerc, O.; Mariethoz, J.; Rivet, A.; Lisacek, F.; Perez, S.; Ricard-Blum, S. A Pipeline to Translate Glycosaminoglycan Sequences into 3D Models. Application to the Exploration of Glycosaminoglycan Conformational Space. *Glycobiology* **2019**, *29*, 36–44.
- (5) Vallet, S. D.; Clerc, O.; Ricard-Blum, S. Glycosaminoglycan–Protein Interactions: The First Draft of the Glycosaminoglycan Interactome. *J. Histochem. Cytochem.* **2021**, *69*, 93–104.
- (6) Karamanos, N. K.; Piperigkou, Z.; Theocharis, A. D.; Watanabe, H.; Franchi, M.; Baud, S.; Brézillat, S.; Götte, M.; Passi, A.; Vigetti, D.; Ricard-Blum, S.; Sanderson, R. D.; Neill, T.; Iozzo, R. V. Proteoglycan Chemical Diversity Drives Multifunctional Cell Regulation and Therapeutics. *Chem. Rev.* **2018**, *118*, 9152–9232.
- (7) Wight, T. N. A Role for Proteoglycans in Vascular Disease. *Matrix Biol.* **2018**, *71–72*, 396–420.
- (8) Marcisz, M.; Huard, B.; Lipska, A. G.; Samsonov, S. A. Further Analyses of APRIL/APRIL-Receptor/Glycosaminoglycan Interactions by Biochemical Assays Linked to Computational Studies. *Glycobiology* **2021**, *31*, 772–786.
- (9) Salbach, J.; Rachner, T. D.; Rauner, M.; Hempel, U.; Anderegg, U.; Franz, S.; Simon, J. C.; Hofbauer, L. C. Regenerative Potential of Glycosaminoglycans for Skin and Bone. *J. Mol. Med.* **2012**, *90*, 625–635.
- (10) 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.
- (11) Moustakas, A.; Souchelnytskyi, S.; Heldin, C. H. Smad Regulation in TGF- $\beta$  Signal Transduction. *J. Cell Sci.* **2001**, *114*, 4359–4369.
- (12) Morla, S. Glycosaminoglycans and Glycosaminoglycan Mimetics in Cancer and Inflammation. *Int. J. Mol. Sci.* **2019**, *20*, No. 1963.
- (13) Shi, D.; Sheng, A.; Chi, L. Glycosaminoglycan-Protein Interactions and Their Roles in Human Disease. *Front. Mol. Biosci.* **2021**, *8*, No. 639666.
- (14) Huynh, M. B.; Ouidja, M. O.; Chantepie, S.; Carpentier, G.; Maïza, A.; Zhang, G.; Vilares, J.; Raisman-Vozari, R.; Papy-Garcia, D. Glycosaminoglycans from Alzheimer’s Disease Hippocampus Have Altered Capacities to Bind and Regulate Growth Factors Activities and to Bind Tau. *PLoS One* **2019**, *14*, No. e0209573.
- (15) Risau, W. Mechanisms of Angiogenesis. *Nature* **1997**, *386*, 671–674.
- (16) Ma, S. N.; Mao, Z. X.; Wu, Y.; Liang, M. X.; Wang, D. D.; Chen, X.; Chang, P.; Zhang, W.; Tang, J. H. The Anti-Cancer Properties of Heparin and Its Derivatives: A Review and Prospect. *Cell Adhes. Migr.* **2020**, *14*, 118–128.
- (17) Paiardi, G.; Richter, S.; Oreste, P.; Urbinati, C.; Rusnati, M.; Wade, R. C. The Binding of Heparin to Spike Glycoprotein Inhibits SARS-CoV-2 Infection by Three Mechanisms. *J. Biol. Chem.* **2022**, *298*, No. 101507.
- (18) Clausen, T. M.; Sandoval, D. R.; Spliid, C. B.; Pihl, J.; Perrett, H. R.; Painter, C. D.; Narayanan, A.; Majowicz, S. A.; Kwong, E. M.; McVicar, R. N.; Thacker, B. E.; Glass, C. A.; Yang, Z.; Torres, J. L.; Golden, G. J.; Bartels, P. L.; Porell, R. N.; Garretson, A. F.; Laubach, L.; Feldman, J.; Yin, X.; Pu, Y.; Hauser, B. M.; Caradonna, T. M.; Kellman, B. P.; Martino, C.; Gordts, P. L. S. M.; Chanda, S. K.; Schmidt, A. G.;

- Godula, K.; Leibel, S. L.; Jose, J.; Corbett, K. D.; Ward, A. B.; Carlin, A. F.; Esko, J. D. SARS-CoV-2 Infection Depends on Cellular Heparan Sulfate and ACE2. *Cell* **2020**, *183*, 1043–1057.e15.
- (19) Kim, S. Y.; Jin, W.; Sood, A.; Montgomery, D. W.; Grant, O. C.; Fuster, M. M.; Fu, L.; Dordick, J. S.; Woods, R. J.; Zhang, F.; Linhardt, R. J. Characterization of Heparin and Severe Acute Respiratory Syndrome-Related Coronavirus 2 (SARS-CoV-2) Spike Glycoprotein Binding Interactions. *Antiviral Res.* **2020**, *181*, No. 104873.
- (20) Liu, L.; Chopra, P.; Li, X.; Bouwman, K. M.; Tompkins, S. M.; Wolfert, M. A.; De Vries, R. P.; Boons, G. J. Heparan Sulfate Proteoglycans as Attachment Factor for SARS-CoV-2. *ACS Cent. Sci.* **2021**, *7*, 1009–1018.
- (21) Derler, R.; Gesslbauer, B.; Weber, C.; Strutzmann, E.; Miller, I.; Kungl, A. Glycosaminoglycan-Mediated Downstream Signaling of CXCL8 Binding to Endothelial Cells. *Int. J. Mol. Sci.* **2017**, *18*, No. 2605.
- (22) Penk, A.; Baumann, L.; Huster, D.; Samsonov, S. A. NMR and Molecular Modeling Reveal Specificity of the Interactions between CXCL14 and Glycosaminoglycans. *Glycobiology* **2019**, *29*, 715–725.
- (23) Nordsieck, K.; Baumann, L.; Hintze, V.; Pisabarro, M. T.; Schnabelrauch, M.; Beck-Sickingler, A. G.; Samsonov, S. A. The Effect of Interleukin-8 Truncations on Its Interactions with Glycosaminoglycans. *Biopolymers* **2018**, *109*, No. e23103.
- (24) Uciechowska-Kaczmarzyk, U.; Babik, S.; Zsila, F.; Bojarski, K. K.; Beke-Somfai, T.; Samsonov, S. A. Molecular Dynamics-Based Model of VEGF-A and Its Heparin Interactions. *J. Mol. Graphics Modell.* **2018**, *82*, 157–166.
- (25) Faham, S.; Hileman, R. E.; Fromm, J. R.; Linhardt, R. J.; Rees, D. C. Heparin Structure and Interactions with Basic Fibroblast Growth Factor. *Science* **1996**, *271*, 1116–1120.
- (26) Digabriele, A. D.; Lax, I.; Chen, D. I.; Svahn, C. M.; Jaye, M.; Schlessinger, J.; Hendrickson, W. A. Structure of a Heparin-Linked Biologically Active Dimer of Fibroblast Growth Factor. *Nature* **1998**, *393*, 812–817.
- (27) Wigén, J.; Elowsson-Rendin, L.; Karlsson, L.; Tykesson, E.; Westergren-Thorsson, G. Glycosaminoglycans: A Link between Development and Regeneration in the Lung. *Stem Cells Dev.* **2019**, *28*, 823–832.
- (28) Faye, C.; Moreau, C.; Chautard, E.; Jetne, R.; Fukai, N.; Ruggiero, F.; Humphries, M. J.; Olsen, B. R.; Ricard-Blum, S. Molecular Interplay between Endostatin, Integrins, and Heparan Sulfate. *J. Biol. Chem.* **2009**, *284*, 22029–22040.
- (29) Xu, D.; Esko, J. D. Demystifying Heparan Sulfate-Protein Interactions. *Annu. Rev. Biochem.* **2014**, *83*, 129–157.
- (30) Nagarajan, B.; Holmes, S. G.; Sankaranarayanan, N. V.; Desai, U. R. Molecular Dynamics Simulations to Understand Glycosaminoglycan Interactions in the Free- and Protein-Bound States. *Curr. Opin. Struct. Biol.* **2022**, *74*, No. 102356.
- (31) Imberty, A.; Lortat-Jacob, H.; Perez, S. Structural View of Glycosaminoglycan–Protein Interactions. *Carbohydr. Res.* **2007**, *342*, 430–439.
- (32) Petitou, M.; Casu, B.; Lindahl, U. 1976–1983, a Critical Period in the History of Heparin: The Discovery of the Antithrombin Binding Site. *Biochimie* **2003**, *85*, 83–89.
- (33) Sepuru, K. M.; Nagarajan, B.; Desai, U. R.; Rajarathnam, K. Structural Basis, Stoichiometry, and Thermodynamics of Binding of the Chemokines KC and MIP2 to the Glycosaminoglycan Heparin. *J. Biol. Chem.* **2018**, *293*, 17817–17828.
- (34) Kogut, M. M.; Marcisz, M.; Samsonov, S. A. Modeling Glycosaminoglycan–Protein Complexes. *Curr. Opin. Struct. Biol.* **2022**, *73*, No. 102332.
- (35) Marcisz, M.; Zacharias, M.; Samsonov, S. A. Modeling Protein-Glycosaminoglycan Complexes: Does the Size Matter? *J. Chem. Inf. Model.* **2021**, *61*, 4475–4485.
- (36) Marcisz, M.; Gaardl0s, M.; Bojarski, K. K.; Siebenmorgen, T.; Zacharias, M.; Samsonov, S. A. Explicit Solvent Repulsive Scaling Replica Exchange Molecular Dynamics (RS-REMD) in Molecular Modeling of Protein Glycosaminoglycan Complexes. *J. Comput. Chem.* **2022**, *43*, 1633–1640.
- (37) Samsonov, S. A.; Zacharias, M.; Chauvot de Beauchene, I. Modeling Large Protein–Glycosaminoglycan Complexes Using a Fragment-Based Approach. *J. Comput. Chem.* **2019**, *40*, 1429–1439.
- (38) Boittier, E. D.; Burns, J. M.; Gandhi, N. S.; Ferro, V. GlycoTorch Vina: Docking Designed and Tested for Glycosaminoglycans. *J. Chem. Inf. Model.* **2020**, *60*, 6328–6343.
- (39) Eberhardt, J.; Santos-Martins, D.; Tillack, A. F.; Forli, S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J. Chem. Inf. Model.* **2021**, *61*, 3891–3898.
- (40) Nivedha, A. K.; Thieker, D. F.; Makeneni, S.; Hu, H.; Woods, R. J. Vina-Carb: Improving Glycosidic Angles during Carbohydrate Docking. *J. Chem. Theory Comput.* **2016**, *12*, 892–901.
- (41) Mottarella, S. E.; Beglov, D.; Beglova, N.; Nugent, M. A.; Kozakov, D.; Vajda, S. Docking server for the identification of heparin binding sites on proteins. *J. Chem. Inf. Model.* **2014**, *54*, 2068–2078.
- (42) van Zundert, G. C. P.; Rodrigues, J. P. G. L. M.; Trellet, M.; Schmitz, C.; Kastiris, P. L.; Karaca, E.; Melquinod, A. S. J.; van Dijk, M.; de Vries, S. J.; Bonvin, A. M. J. J. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* **2016**, *428*, 720–725.
- (43) Grosdidier, A.; Zoete, V.; Michielin, O. SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Res.* **2011**, *39*, W270–W277.
- (44) Uciechowska-Kaczmarzyk, U.; Chauvot de Beauchene, I.; Samsonov, S. A. Docking software performance in protein-glycosaminoglycan systems. *J. Mol. Graphics Modell.* **2019**, *90*, 42–50.
- (45) Sapay, N.; Cabannes, E.; Petitou, M.; Imberty, A. Molecular Modeling of the Interaction between Heparan Sulfate and Cellular Growth Factors: Bringing Pieces Together. *Glycobiology* **2011**, *21*, 1181–1193.
- (46) Nagarajan, B.; Sankaranarayanan, N. V.; Patel, B. B.; Desai, U. R. A Molecular Dynamics-Based Algorithm for Evaluating the Glycosaminoglycan Mimicking Potential of Synthetic, Homogenous, Sulfated Small Molecules. *PLoS One* **2017**, *12*, No. e0171619.
- (47) Singh, A.; Kett, W. C.; Severin, I. C.; Agyekum, I.; Duan, J.; Amster, I. J.; Proudfoot, A. E. I.; Coombe, D. R.; Woods, R. J. The Interaction of Heparin Tetrasaccharides with Chemokine CCL5 Is Modulated by Sulfation Pattern and pH. *J. Biol. Chem.* **2015**, *290*, 15421–15436.
- (48) Roy, R.; Jonniya, N. A.; Kar, P. Effect of Sulfation on the Conformational Dynamics of Dermatan Sulfate Glycosaminoglycan: A Gaussian Accelerated Molecular Dynamics Study. *J. Phys. Chem. B* **2022**, *126*, 3852–3866.
- (49) Samsonov, S. A.; Teyra, J.; Pisabarro, M. T. Docking Glycosaminoglycans to Proteins: Analysis of Solvent Inclusion. *J. Comput. Aided Mol. Des.* **2011**, *25*, 477–489.
- (50) Samsonov, S.; Teyra, J.; Pisabarro, M. T. A Molecular Dynamics Approach to Study the Importance of Solvent in Protein Interactions. *Proteins: Struct., Funct., Genet.* **2008**, *73*, 515–525.
- (51) Samsonov, S. A. Computational Analysis of Solvent Inclusion in Docking Studies of Protein–Glycosaminoglycan Systems. In *Computational Drug Discovery and Design*; Gore, M. et al., Eds.; Methods in Molecular Biology; Humana Press, 2018; Vol. 1762, pp 445–454.
- (52) Almond, A.; Sheehan, J. K.; Brass, A. Molecular Dynamics Simulations of the Two Disaccharides of Hyaluronan in Aqueous Solution. *Glycobiology* **1997**, *7*, 597–604.
- (53) Almond, A.; Sheehan, J. K. Glycosaminoglycan Conformation: Do Aqueous Molecular Dynamics Simulations Agree with x-Ray Fiber Diffraction? *Glycobiology* **2000**, *10*, 329–338.
- (54) Lang, E. J. M.; Baker, E. G.; Woolfson, D. N.; Mulholland, A. J. Generalized Born Implicit Solvent Models Do Not Reproduce Secondary Structures of de Novo Designed Glu/Lys Peptides. *J. Chem. Theory Comput.* **2022**, *18*, 4070–4076.
- (55) Verli, H.; Guimarões, J. A. Molecular dynamics simulation of a decasaccharide fragment of heparin in aqueous solution. *Carbohydr. Res.* **2004**, *339*, 281–290.
- (56) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. GLYCAM06: A

- Generalizable Biomolecular Force Field. *Carbohydrates. J. Comput. Chem.* **2008**, *29*, 622–655.
- (57) Izadi, S.; Onufriev, A. V. Accuracy Limit of Rigid 3-Point Water Models. *J. Chem. Phys.* **2016**, *145*, No. 074501.
- (58) Izadi, S.; Anandakrishnan, R.; Onufriev, A. V. Building Water Models: A Different Approach. *J. Phys. Chem. Lett.* **2014**, *5*, 3863–3871.
- (59) Teyra, J.; Samsonov, S. A.; Schreiber, S.; Pisabarro, M. T. SCOWLP Update: 3D Classification of Protein-Protein, -Peptide, -Saccharide and -Nucleic Acid Interactions, and Structure-Based Binding Inferences across Folds. *BMC Bioinf.* **2011**, *12*, No. 398.
- (60) Jana, M.; Bandyopadhyay, S. Conformational Flexibility of a Protein-Carbohydrate Complex and the Structure and Ordering of Surrounding Water. *Phys. Chem. Chem. Phys.* **2012**, *14*, 6628–6638.
- (61) Nurisso, A.; Blanchard, B.; Audfray, A.; Rydner, L.; Oscarson, S.; Varrot, A.; Imbert, A. Role of Water Molecules in Structure and Energetics of Pseudomonas Aeruginosa Lectin I Interacting with Disaccharides. *J. Biol. Chem.* **2010**, *285*, 20316–20327.
- (62) Sepuru, K. M.; Nagarajan, B.; Desai, U. R.; Rajarathnam, K. Molecular Basis of Chemokine CXCL5-Glycosaminoglycan Interactions. *J. Biol. Chem.* **2016**, *291*, 20539–20550.
- (63) Shanthamurthy, C. D.; Gimeno, A.; Leviatan Ben-Arye, S.; Kumar, N. V.; Jain, P.; Padler-Karavani, V.; Jimenez-Barbero, J.; Kikkeri, R. Sulfation Code and Conformational Plasticity of 1-Iduronic Acid Homo-Oligosaccharides Mimic the Biological Functions of Heparan Sulfate. *ACS Chem. Biol.* **2021**, *16*, 2481–2489.
- (64) Liu, Q.; Brady, J. W. Anisotropic Solvent Structuring in Aqueous Sugar Solutions. *J. Am. Chem. Soc.* **1996**, *118*, 12276–12286.
- (65) Pagnotta, S. E.; McLain, S. E.; Soper, A. K.; Bruni, F.; Ricci, M. A. Water and Trehalose: How Much Do They Interact with Each Other? *J. Phys. Chem. B* **2010**, *114*, 4904–4908.
- (66) Pol-Fachin, L.; Verli, H. Depiction of the Forces Participating in the 2-O-Sulfo- $\alpha$ -1-Iduronic Acid Conformational Preference in Heparin Sequences in Aqueous Solutions. *Carbohydr. Res.* **2008**, *343*, 1435–1445.
- (67) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Pairwise Solute Descreening of Solute Charges from a Dielectric Medium. *Chem. Phys. Lett.* **1995**, *246*, 122–129.
- (68) Onufriev, A.; Bashford, D.; Case, D. A. Modification of the Generalized Born Model Suitable for Macromolecules. *J. Phys. Chem. B* **2000**, *104*, 3712–3720.
- (69) Onufriev, A.; Bashford, D.; Case, D. A. Exploring Protein Native States and Large-Scale Conformational Changes with a Modified Generalized Born Model. *Proteins: Struct., Funct., Bioinf.* **2004**, *55*, 383–394.
- (70) Mongan, J.; Simmerling, C.; McCammon, J. A.; Case, D. A.; Onufriev, A. Generalized Born Model with a Simple, Robust Molecular Volume Correction. *J. Chem. Theory Comput.* **2007**, *3*, 156–169.
- (71) Nguyen, H.; Roe, D. R.; Simmerling, C. Improved Generalized Born Solvent Model Parameters for Protein Simulations. *J. Chem. Theory Comput.* **2013**, *9*, 2020–2034.
- (72) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926.
- (73) Neria, E.; Fischer, S.; Karplus, M. Simulation of Activation Free Energies in Molecular Systems. *J. Chem. Phys.* **1996**, *105*, 1902.
- (74) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem. A* **1987**, *91*, 6269–6271.
- (75) Horn, H. W.; Swope, W. C.; Pitner, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an Improved Four-Site Water Model for Biomolecular Simulations: TIP4P-Ew. *J. Chem. Phys.* **2004**, *120*, 9665–9678.
- (76) Mahoney, M. W.; Jorgensen, W. L. A Five-Site Model for Liquid Water and the Reproduction of the Density Anomaly by Rigid, Nonpolarizable Potential Functions. *J. Chem. Phys.* **2000**, *112*, 8910.
- (77) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
- (78) Mulloy, B.; Forster, M. J.; Jones, C.; Davies, D. B. N.M.R. and Molecular-Modelling Studies of the Solution Conformation of Heparin. *Biochem. J.* **1993**, *293*, 849–858.
- (79) Huige, C. J. M.; Altona, C. 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.
- (80) Sattelle, B. M.; Shakeri, J.; Almond, A. Does microsecond sugar ring flexing encode 3D-shape and bioactivity in the heparanome? *Biomacromolecules* **2013**, *14*, 1149–1159.
- (81) Bojarski, K. K.; Sieradzan, A. K.; Samsonov, S. A. Molecular dynamics insights into protein-glycosaminoglycan systems from microsecond-scale simulations. *Biopolymers* **2019**, *110*, No. e23252.
- (82) Ferro, D. R.; Provasoli, A.; Ragazzi, M.; Casu, B.; Torri, G.; Bossennec, V.; Perly, B.; Sinay, P.; Petitou, M.; Choay, J. Conformer populations of L-iduronic acid residues in glycosaminoglycan sequences. *Carbohydr. Res.* **1990**, *195*, 157–167.
- (83) van Boeckel, C. A. A.; van Aelst, S. F.; Wagenaars, G. N.; Mellema, J. R.; Paulsen, H.; Peters, T.; Pollex, A.; Sinnwell, V. Conformational analysis of synthetic heparin-like oligosaccharides containing  $\alpha$ -L-idopyranosyluronic acid. *Recl. Trav. Chim. Pays-Bas* **1987**, *106*, 19–29.
- (84) Samsonov, S. A.; Bichmann, L.; Pisabarro, M. T. Coarse-grained model of glycosaminoglycans. *J. Chem. Inf. Model.* **2015**, *55*, 114–124.
- (85) Case, D. A.; Belfon, K.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Giambasu, G.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Harris, R.; Izadi, S.; Izmailov, S. A.; Kasavajhala, K.; Kovalenko, A.; Krasny, R.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Man, V.; Merz, K. M.; Miao, Y.; Mikhailovskii, O.; Monard, G.; Nguyen, H.; Onufriev, A.; Pan, F.; Pantano, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Skrynnikov, N. R.; Smith, J.; Swails, K.; Walker, R. C.; Wang, J.; Wilson, L.; Wolf, R. M.; Wu, X.; Xiong, Y.; Xue, Y.; York, D. M.; Kollman, P. A. *AMBER 2020*; University of California: San Francisco, 2020.
- (86) Khan, S.; Gor, J.; Mulloy, B.; Perkins, S. J. Semi-rigid solution structures of heparin by constrained X-ray scattering modelling: new insight into heparin-protein complexes. *J. Mol. Biol.* **2010**, *395*, 504–521.
- (87) Pavlov, G.; Finet, S.; Tatarenko, K.; Korneeva, E.; Ebel, C. Conformation of heparin studied with macromolecular hydrodynamic methods and X-ray scattering. *Eur. Biophys. J.* **2003**, *32*, 437–449.
- (88) Onufriev, A.; Case, D. A.; Bashford, D. Effective Born Radii in the Generalized Born Approximation: The Importance of Being Perfect. *J. Comput. Chem.* **2002**, *23*, 1297–1304.
- (89) Nagarajan, B.; Sankaranarayanan, N. V.; Desai, U. R. Rigorous Analysis of Free Solution Glycosaminoglycan Dynamics Using Simple, New Tools. *Glycobiology* **2020**, *30*, 516–527.
- (90) Core\_Team\_(2013). *R: A Language and Environment for Statistical Computing*; R. R Development Core Team, 2013.
- (91) Pluhařová, E.; Mason, P. E.; Jungwirth, P. Ion pairing in aqueous lithium salt solutions with monovalent and divalent counter-anions. *J. Phys. Chem. A* **2013**, *117*, 1766–1773.
- (92) Holmes, S. G.; Nagarajan, B.; Desai, U. R. 3-O-Sulfation Induces Sequence-Specific Compact Topologies in Heparan Sulfate That Encode a Dynamic Sulfation Code. *Comput. Struct. Biotechnol. J.* **2022**, *20*, 3884–3898.
- (93) Guvench, O.; Whitmore, E. K. Sulfation and Calcium Favor Compact Conformations of Chondroitin in Aqueous Solutions. *ACS Omega* **2021**, *6*, 13204–13217.
- (94) Neamtu, A.; Tamba, B.; Patras, X. Molecular dynamics simulations of chondroitin sulfate in explicit solvent: point charge water models compared. *Cellul. Chem. Technol.* **2013**, *47*, 191–202.

## Publication D7

Modeling glycosaminoglycan-protein  
complexes





# Modeling glycosaminoglycan–protein complexes

Małgorzata M. Kogut<sup>a</sup>, Mateusz Marcisz<sup>a</sup> 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](mailto:sergey.samsonov@ug.edu.pl))

<sup>a</sup> 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.

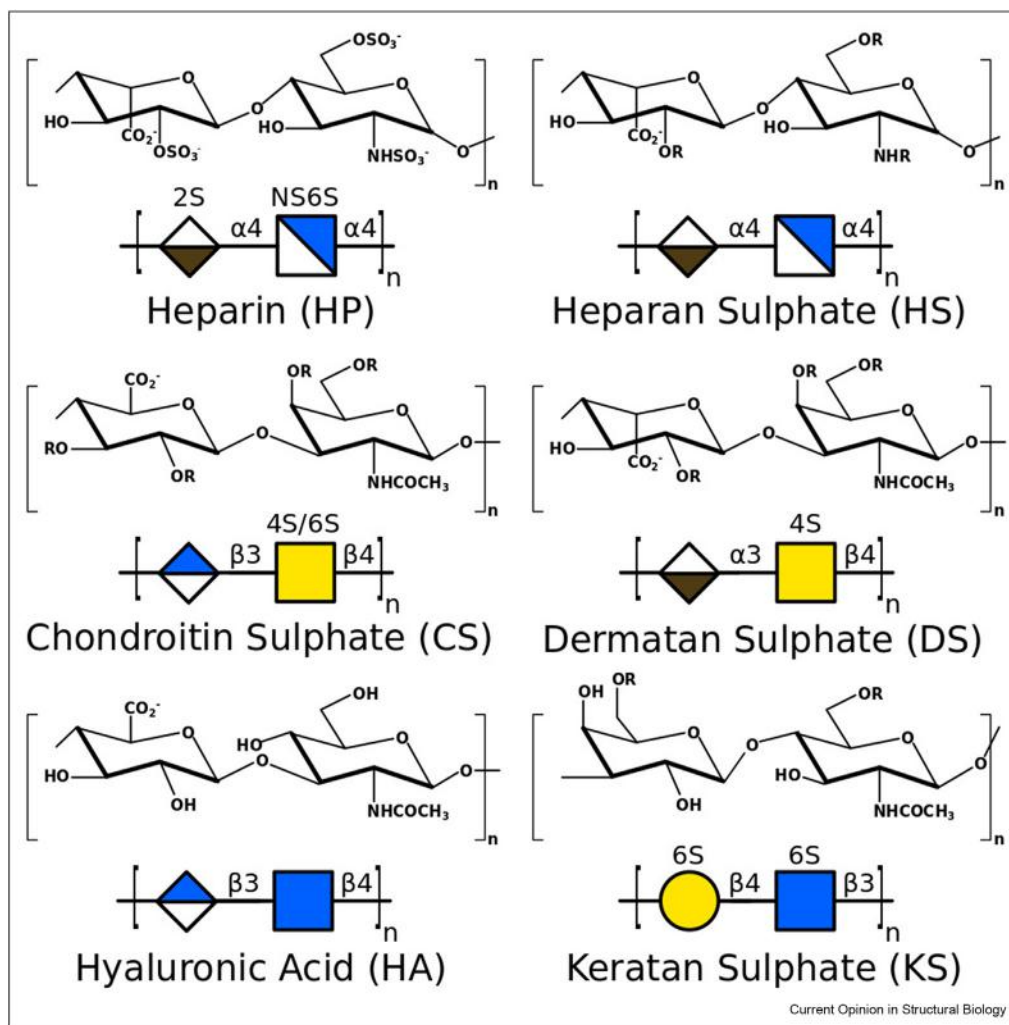
## Introduction

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



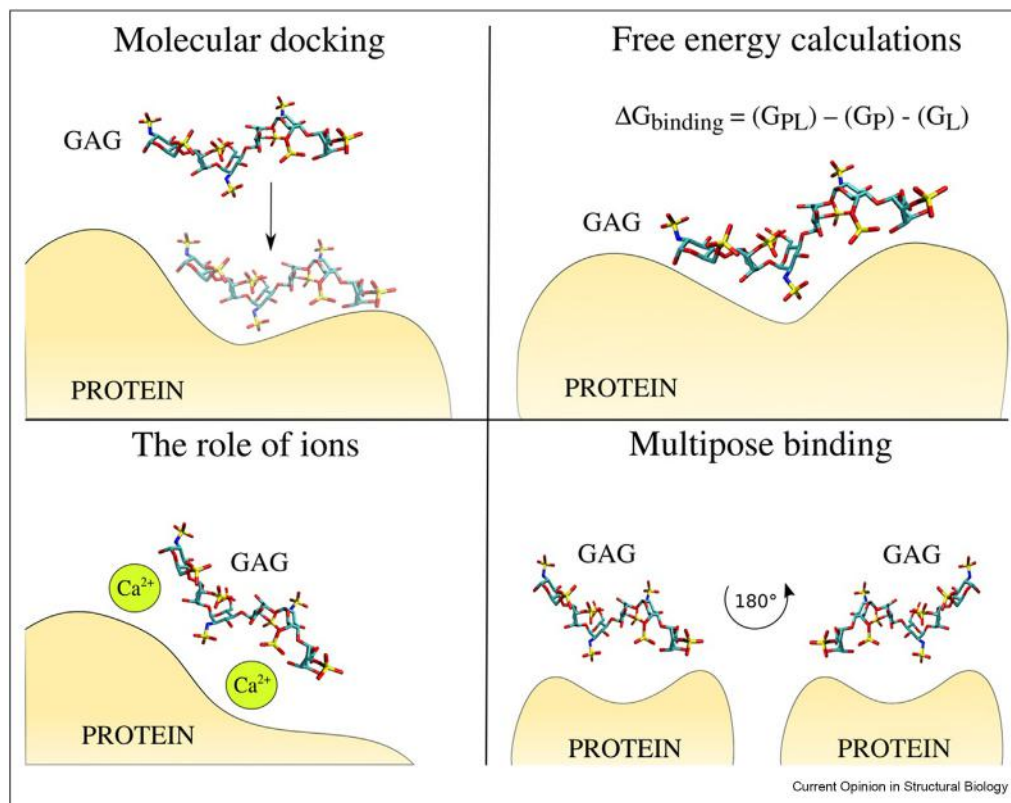
Repeating disaccharide units of glycosaminoglycans with their SNFG representation.

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 <sup>a</sup>	Unlimited	Unlimited <sup>b</sup>	Unlimited

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

<sup>a</sup> Docking GAGs longer than dp8 makes docking results unreliable.

<sup>b</sup> Poses limited to the surface electrostatic potential of the receptor.

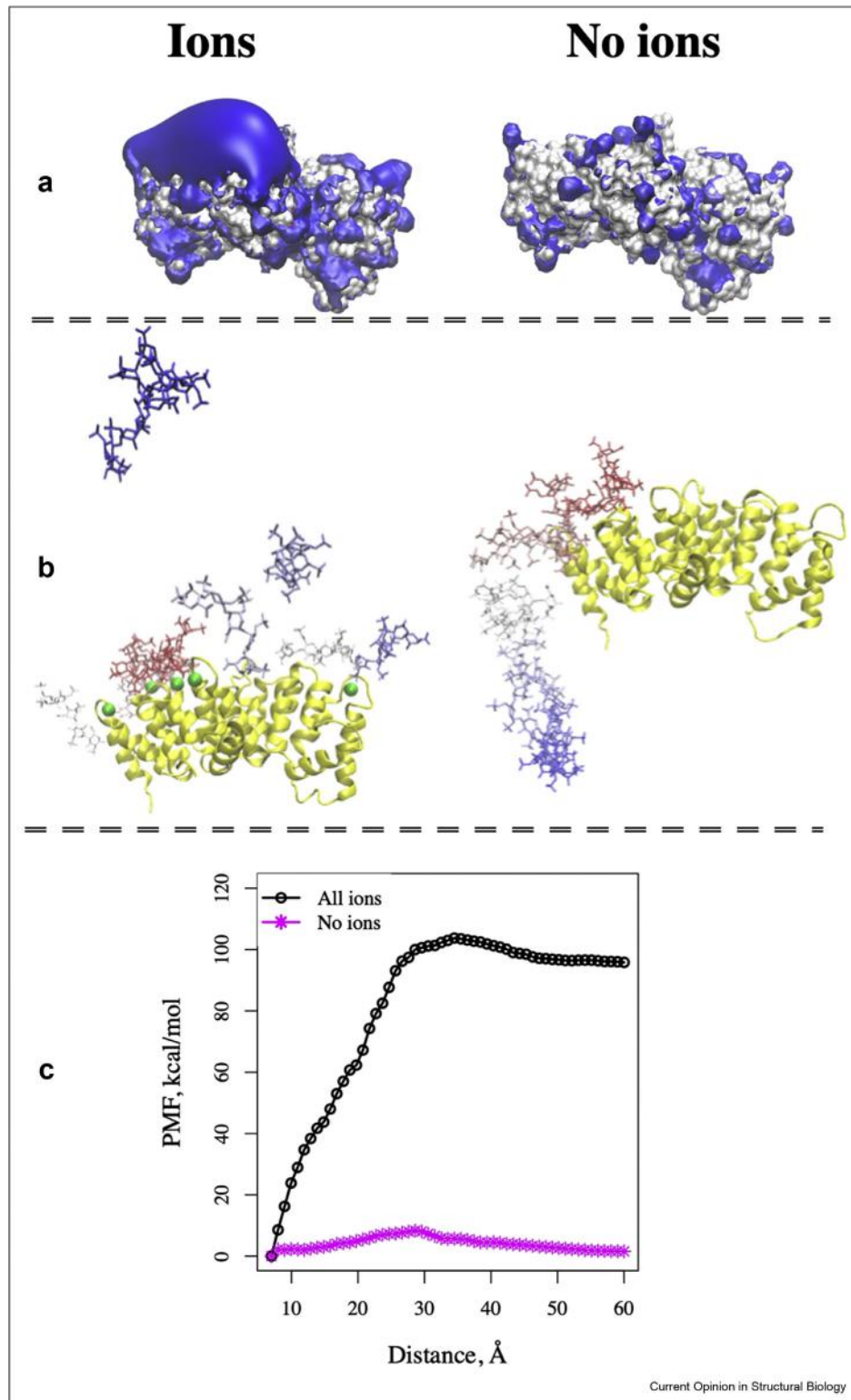
by properly ranking GAG-binding affinities. In addition, they allowed for atomistic interpretation of the experimental data within interdisciplinary studies [34,35]. LIE (linear interaction energy) is another, although less popular, technique, for energy analysis of GAGs, and was applied in several GAG 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]. Zn<sup>2+</sup> attenuation of the binding affinity was demonstrated for endostatin-heparin interactions [46]. Recently, several computational studies

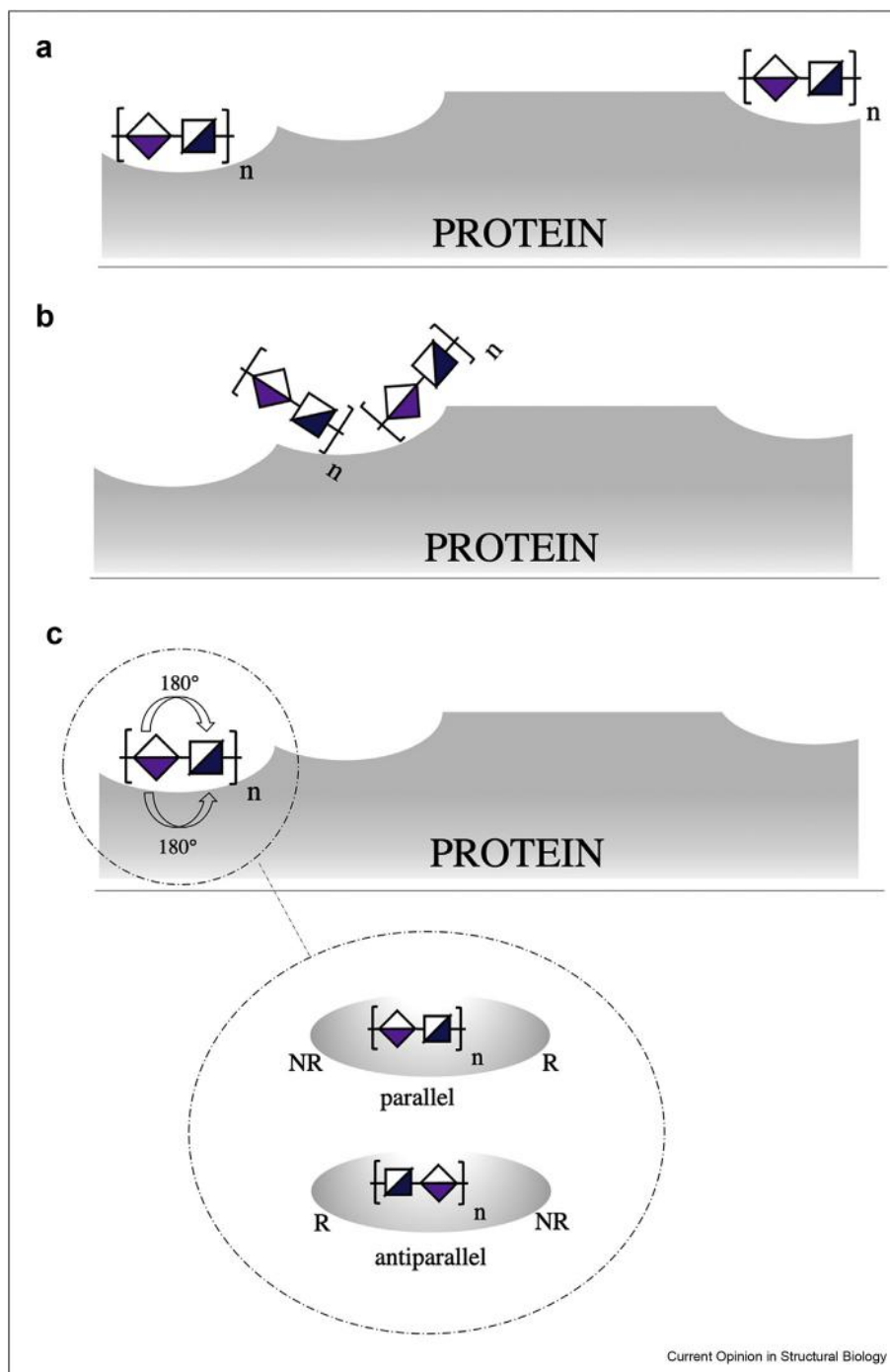
highlighted the importance of ions in the context of molecular modeling of protein-GAG complexes and their free energy analysis to improve the protocols used to treat the ionic environment in GAG-binding studies. The effect of ions on protein-GAG binding using modeling approaches was endeavored by Potthoff 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  $\text{Ca}^{2+}$  calculated with the Poisson-Boltzmann surface area approach. **(b)** The heparin dissociation pathway for PDB ID 2HYV complex in the presence and absence of  $\text{Ca}^{2+}$  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, Milanesi 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 3<sup>rd</sup> 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. Multhaup 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, Bonnardel 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 built 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  $\beta$  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

Mateusz Marcisz  
University of Gdańsk, Faculty of Chemistry  
Department of Theoretical Chemistry  
Wita Stwosza 63, Gdańsk, Poland  
mateusz.marcisz@phdstud.ug.edu.pl

Gdańsk, 1.06.2023

## STATEMENT

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

### **Evaluation of replica exchange with repulsive scaling approach for docking glycosaminoglycans.**

Martyna Maszota-Zieleniak, Mateusz Marcisz, Małgorzata M Kogut, Till Siebenmorgen, Martin Zacharias, Sergey A Samsonov.

Published in: Journal of Computational Chemistry, 2021, 42(15), 1040-1053..

I contributed to the preparation of this manuscript by providing data, writing, and revising the manuscript.

Percentage contribution: 40%

### **Advanced molecular dynamics approaches to model a tertiary complex APRIL/TACI with long glycosaminoglycans."**

Mateusz Marcisz, Martyna Maszota-Zieleniak, Bertrand Huard, and Sergey A. Samsonov

Published in: Biomolecules 2021, 11(9), 1349.

I contributed to the preparation of this manuscript by providing data, writing and revising the manuscript.

Percentage contribution: 40%

### **Repulsive Scaling Replica Exchange Molecular Dynamics in Modeling Protein-Glycosaminoglycan Complexes**

Mateusz Marcisz, Martyna Maszota-Zieleniak, and Sergey A. Samsonov.

Published in: Proteoglycans: Methods and Protocols. New York, NY: Springer US, 2023. 153-167

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

Percentage contribution: 70%

**Explicit solvent repulsive scaling replica exchange molecular dynamics (RS-REMD) in molecular modeling of protein-glycosaminoglycan complexes.**

Marcisz, M., Gaardløs, M., Bojarski, K. K., Siebenmorgen, T., Zacharias, M., & Samsonov, S. A

Published in: Journal of Computational Chemistry, 2022, 43(24), 1633-1640

I contributed to the preparation of this manuscript by providing data, writing and revising the manuscript.

Percentage contribution: 40%

**Modeling protein–glycosaminoglycan complexes: does the size matter?**

Mateusz Marcisz, Martin Zacharias, Sergey A. Samsonov

Published in: Proteoglycans: Journal of Chemical Information and Modeling 2021, 61(9), 4475-4485.

I contributed to the preparation of this manuscript by providing data, writing and revising the manuscript.

Percentage contribution: 80%

**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%

**Solvent models benchmark for molecular dynamics of glycosaminoglycans**

Mateusz Marcisz, Sergey A. Samsonov.

Published in: Journal of Chemical Information and Modeling. 2023, 63(7), 2147-57.

I contributed to the preparation of this manuscript by providing data, writing and revising the manuscript.

Percentage contribution: 80%



.....  
Signature

Dr hab. prof. UG Sergey Samsonov  
University of Gdańsk, Faculty of Chemistry  
Department of Theoretical Chemistry  
Wita Stwosza 63, Gdańsk, Poland  
martyna.maszota-zieleniak@ug.edu.pl

Gdańsk, 1.06.2023

## STATEMENT

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

### **Evaluation of replica exchange with repulsive scaling approach for docking glycosaminoglycans.**

Martyna Maszota-Zieleniak, Mateusz Marcisz, Małgorzata M Kogut, Till Siebenmorgen, Martin Zacharias, Sergey A Samsonov.

Published in: Journal of Computational Chemistry, 2021, 42(15), 1040-1053..

I contributed to the preparation of this manuscript by providing data, writing, and revising the manuscript.

### **Advanced molecular dynamics approaches to model a tertiary complex APRIL/TACI with long glycosaminoglycans."**

Marcisz, Mateusz, Martyna Maszota-Zieleniak, Bertrand Huard, and Sergey A. Samsonov

Published in: Biomolecules 2021, 11(9), 1349.

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

### **Repulsive Scaling Replica Exchange Molecular Dynamics in Modeling Protein-Glycosaminoglycan Complexes**

Marcisz, Mateusz, Martyna Maszota-Zieleniak, and Sergey A. Samsonov.

Published in: Proteoglycans: Methods and Protocols. New York, NY: Springer US, 2023. 153-167

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

### **Explicit solvent repulsive scaling replica exchange molecular dynamics (RS REMD) in molecular modeling of protein glycosaminoglycan complexes.**

Marcisz, M., Gaardløs, M., Bojarski, K. K., Siebenmorgen, T., Zacharias, M., & Samsonov, S. A

Published in: Journal of Computational Chemistry, 2022, 43.24: 1633-1640

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

## **Modeling protein–glycosaminoglycan complexes: does the size matter?**

Mateusz Marcisz, Martin Zacharias, Sergey A. Samsonov

Published in: Proteoglycans: Journal of Chemical Information and Modeling 2021, 61.9: 4475-4485.

I contributed to the preparation of this manuscript by 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.

## **Solvent models benchmark for molecular dynamics of glycosaminoglycans**

Marcisz, Mateusz, Sergey A. Samsonov.

Published in: Journal of Chemical Information and Modeling. 2023, 63.7: 2147-57.

  
.....  
Signature

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

Gdańsk, 31.05.2023

## STATEMENT

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

### **Evaluation of replica exchange with repulsive scaling approach for docking glycosaminoglycans.**

Martyna Maszota-Zieleniak, Mateusz Marcisz, Małgorzata M Kogut, Till Siebenmorgen, Martin Zacharias, Sergey A Samsonov.

Published in: Journal of Computational Chemistry, 2021, 42(15),1040-1053.


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)



Dr Martyna Maszota-Zieleniak  
University of Gdańsk, Faculty of Chemistry  
Department of Theoretical Chemistry  
Wita Stwosza 63, Gdańsk, Poland  
martyna.maszota-zieleniak@ug.edu.pl

Gdańsk, 29.05.2023

## STATEMENT

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

### **Evaluation of replica exchange with repulsive scaling approach for docking glycosaminoglycans.**

Martyna Maszota-Zieleniak, Mateusz Marcisz, Małgorzata M Kogut, Till Siebenmorgen, Martin Zacharias, Sergey A Samsonov.

Published in: Journal of Computational Chemistry, 2021, 42(15), 1040-1053..

I contributed to the preparation of this manuscript by providing data, writing, and revising the manuscript.

### **Advanced molecular dynamics approaches to model a tertiary complex APRIL/TACI with long glycosaminoglycans."**

Marcisz, Mateusz, Martyna Maszota-Zieleniak, Bertrand Huard, and Sergey A. Samsonov

Published in: Biomolecules 2021, 11(9), 1349.

I contributed to the preparation of this manuscript by providing data, writing, and revising the manuscript.

### **Repulsive Scaling Replica Exchange Molecular Dynamics in Modeling Protein-Glycosaminoglycan Complexes**

Marcisz, Mateusz, Martyna Maszota-Zieleniak, and Sergey A. Samsonov.

Published in: Proteoglycans: Methods and Protocols. New York, NY: Springer US, 2023. 153-167

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



Signature

Dr. Till Siebenmorgen  
Technical University of Munich  
Center for Functional Protein Assemblies  
Ernst-Otto-Fischer-Straße 8  
85748 Garching, Germany  
till.siebenmorgen@tum.de

Garching, 31.05.2023

## STATEMENT

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

### **Evaluation of replica exchange with repulsive scaling approach for docking glycosaminoglycans.**

Martyna Maszota-Zieleniak, Mateusz Marcisz, Małgorzata M Kogut, Till Siebenmorgen, Martin Zacharias, Sergey A Samsonov.

Published in: Journal of Computational Chemistry, 2021, 42.15: 1040-1053.

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

### **Explicit solvent repulsive scaling replica exchange molecular dynamics (RS-REMD) in molecular modeling of protein-glycosaminoglycan complexes.**

Marcisz, M., Gaardløs, M., Bojarski, K. K., Siebenmorgen, T., Zacharias, M., & Samsonov, S. A

Published in: Journal of Computational Chemistry, 2022, 43.24: 1633-1640

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



Signature

Dr Krzysztof Kamil Bojarski  
Politechnika Gdańska  
Wydział Chemiczny  
Gabriela Narutowicza 11/12  
80-233 Gdańsk  
krzysztof.bojarski@pg.edu.pl

Gdańsk, 14.06.2023

## STATEMENT

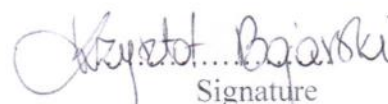
I declare that I am a co-author of the following publications included in the PhD thesis:

**Explicit solvent repulsive scaling replica exchange molecular dynamics (RS-REMD) in molecular modeling of protein-glycosaminoglycan complexes.**

Marcisz, M., Gaardløs, M., Bojarski, K. K., Siebenmorgen, T., Zacharias, M., & Samsonov, S. A

Published in: Journal of Computational Chemistry, 2022, 43.24: 1633-1640

I contributed to the preparation of this manuscript by providing data, writing and revising the manuscript.

  
Signature

Professor Martin Zacharias  
Technical University of Munich  
Center for Functional Protein Assemblies  
Ernst-Otto-Fischer-Straße 8  
85748 Garching, Germany  
zacharias@tum.de

Garching, 31.05.2023

## STATEMENT

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

### **Evaluation of replica exchange with repulsive scaling approach for docking glycosaminoglycans.**

Martyna Maszota-Zieleniak, Mateusz Marcisz, Małgorzata M Kogut, Till Siebenmorgen, Martin Zacharias, Sergey A Samsonov.

Published in: Journal of Computational Chemistry, 2021, 42.15: 1040-1053.

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

### **Modeling protein–glycosaminoglycan complexes: does the size matter?**

Mateusz Marcisz, Martin Zacharias, and Sergey A. Samsonov

Published in: Proteoglycans: Journal of Chemical Information and Modeling 2021, 61.9: 4475-4485.

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

### **Explicit solvent repulsive scaling replica exchange molecular dynamics (RS-REMD) in molecular modeling of protein-glycosaminoglycan complexes.**

Marcisz, M., Gaardløs, M., Bojarski, K. K., Siebenmorgen, T., Zacharias, M., & Samsonov, S. A

Published in: Journal of Computational Chemistry, 2022, 43.24: 1633-1640

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



Signature

PhD Margrethe Gaardløs  
Ingelbrecht Knudssøns gate 4B  
0365 Oslo  
margrethe.gaardlos@gmail.com

Oslo, 01.06.2023

## STATEMENT

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

**Explicit solvent repulsive scaling replica exchange molecular dynamics (RS-REMD) in molecular modeling of protein-glycosaminoglycan complexes.**

Marcisz, M., Gaardløs, M., Bojarski, K. K., Siebenmorgen, T., Zacharias, M., & Samsonov, S. A

Published in: Journal of Computational Chemistry, 2022, 43.24: 1633-1640

I contributed to the preparation of this manuscript by providing data, writing and revising the manuscript.

  
.....  
Signature

Professor Bertrand Huard  
University of Grenoble-Alpes  
Translational Innovation in Medicine and Complexity  
38700 La Tronche, France  
bertrand.huard@univ-grenoble-alpes.fr

Grenoble, 29.05.2023

## STATEMENT

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

**Advanced molecular dynamics approaches to model a tertiary complex APRIL/TACI with long glycosaminoglycans."**

Mateusz Marcisz, Martyna Maszota-Zieleniak, Bertrand Huard, and Sergey A. Samsonov

Published in: *Biomolecules* 2021, *11*(9), 1349.

I contributed to the preparation of this manuscript by providing data, writing and revising the manuscript.

**Further analyses of APRIL/APRIL-receptor/glycosaminoglycan interactions by biochemical assays linked to computational studies.**

Marcisz, M., Huard, B., Lipska, A. G., & Samsonov, S. A.

Published in: *Glycobiology* 2021, *31*(7), 772-786.

I contributed to the preparation of this manuscript by providing data, writing and revising the manuscript.



.....

Signature