Conformational Analysis of Heparin-Analogue Pentasaccharides by Nuclear Magnetic Resonance Spectroscopy and Molecular Dynamics Simulations

Gábor Balogh,[a#] Tamás Gyöngyösi,[b,c#] István Timári,[b#] Mihály Herczeg,[d,e] Anikó Borbás,[d] S. Kashif Sadiq,[f,g]* Krisztina Fehér,[c]* Katalin E. Kövér[b,c]*

[a]Division of Clinical Laboratory Science, Department of Laboratory Medicine, Faculty of Medicine, University of Debrecen, Nagyerdei krt. 98, H-4032 Debrecen, Hungary

[b]Department of Inorganic and Analytical Chemistry, University of Debrecen, Egyetem tér 1, H-4032 Debrecen, Hungary

[c]MTA-DE Molecular Recognition and Interaction Research Group, University of Debrecen, Egyetem tér 1, H-4032 Debrecen, Hungary

[d]Department of Pharmaceutical Chemistry, University of Debrecen, Egyetem tér 1, H-4032 Debrecen, Hungary

[e]Research Group for Oligosaccharide Chemistry of Hungarian Academy of Sciences, University of Debrecen, Egyetem tér 1, H-4032 Debrecen, Hungary

[f]Heidelberg Institute for Theoretical Studies, Schloss-Wolfsbrunnenweg 35, 69118 Heidelberg, Germany

[g]European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany

#these authors contributed equally

*corresponding authors

E-mail: kover@science.unideb.hu

E-mail: feher.krisztina@science.unideb.hu

E-Mail: kashif.sadiq@embl.de

1

Abstract

Elucidation and improvement of blood coagulant properties of heparin is the focus of intense research. In this study we performed conformational analysis using Nuclear Magnetic Resonance (NMR) spectroscopy and Molecular Dynamics (MD) simulations on heparin pentasaccharide analogue idraparinux, its disulfonatomethyl analogue, which features a slightly improved blood coagulation property and a trisulfonatomethyl analogue, in which activity has been totally abolished. As the ring conformation of the G subunit has been suggested as a major determinant of biological properties, we analyzed the sugar ring conformations and the dynamics of the interglycosidic linkages. We found that the conformation of the G ring is dominated by the ^2S_O skewed boat next to $^1C_4$ chair in all three derivatives. Both the thermodynamics and the kinetics of the conformational states were found to be highly similar in the three derivatives. Molecular kinetic analysis showed that the ^2S_O skewed boat state of the G ring is equally favorable in the three analogues resulting in similar ^2S_O populations. Also, the transition kinetics from $^1C_4$ chair to ^2S_O skewed boat was found to be comparable in the derivatives, which indicates a similar energy barrier between the two states of the G subunit. We also identified a slower conformational transition between the dominant $^4C_1$ chair and the boat conformations on the E subunit. Both G and E ring flips are also accompanied by changes along the interglycosidic linkages, which take place highly synchronous with the ring flips. These findings indicate that conformational plasticity of the G ring and the dominance of the ^2S_O skewed boat populations do not necessarily warrant the biological activity of derivatives and the impact of other factors need to be also considered.

Abbreviations

Antithrombin (AT)

Markov state model (MSM)

Molecular Dynamics (MD)

Nuclear Magnetic Resonance (NMR)

Nuclear Overhauser Effect (NOE)

Introduction

Heparin is a linear anionic polysaccharide belonging to the heparan sulfate family of glycosaminoglycans. It is composed of alternating α-D-glucosamine and hexuronic acids (β-D-glucuronic acid, βDGlcA, or α-L-iduronic acid, αLIdoA) that are variably *N*- and *O*-sulfated.[1] Heparin

and heparan sulfate have been shown to interact with at least 300 proteins and can regulate a range of biological processes including cell development and differentiation, inflammation, angiogenesis, neurodegenerative diseases, and viral and bacterial infections.[2-5] Heparin chains display high structural variability in size, sequence and degree of sulfation that results from post-polymerization, nontemplate-driven enzymatic modifications (*O*- and *N*-sulfation, C5 epimerization) on the precursor -(1→4)-βDGlcA-αDGlcNAc-(1→4)- polysaccharide.[6] The sequence and the sulfation pattern of heparin determine its three-dimensional structure and therefore its mode of interaction with proteins. Elucidation of the conformational and the dynamical properties of these complex molecules is essential to enable a better understanding of their structure–function relationships.

The best studied activity of heparin is the anticoagulant effect, which is due to its ability to bind to and enhance the inhibitory activity of the plasma protein antithrombin (AT) against serine proteases of the coagulation system, particularly thrombin and factor Xa.[7, 8] The structure of the specific pentasaccharide sequence responsible for the binding to AT is αDGlcNAc(6S)-βDGlcA-αDGlcNS(3S,6S)-αLIdoA(2S)-αDGlcNS(6S) with an unusual 3-*O*-sulfate group in the inner glucosamine.[9] The interaction of the pentasaccharide is mainly governed by specific electrostatic interactions between its sulfate and carboxylate groups and positively charged amino acids of the heparin-binding region of AT. [10] The molecular basis of this interaction was first revealed by crystal structure studies.[11, 12] using idraparinux (**1**, Figure 1), a synthetic high-affinity, non-glycosaminoglycan variant of the natural pentasaccharide, differing in the pattern of sulfation and full *O*-methylation.[13] Further studies with naturally occurring and synthetic heparin oligosaccharides indicated that the conformational plasticity of the L-iduronic acid, and its ability to adopt the ^2S_O skewed boat conformation is crucial for binding to AT.[14-18] Systematic NMR and computational experiments uncovered that the conformation of LIdoA is regulated by the sulfation patterns of neighboring sugar units.[19, 20]

Our research group[20, 21] and others[22, 23] investigated how the functional group modification of the individual monosaccharides of the AT-binding pentasaccharide affects the anticoagulant activity. We have found that substitution of the 6-*O*-sulfate ester groups of the **F** and **H** glucose units of idraparinux by sulfonatomethyl moieties hardly affects the AT-binding capability confirmed by the outstanding anticoagulant activity of pentasaccharide disulfonate **2** (Figure 1). However, quite surprisingly, introduction of a third sulfonatomethyl moiety onto the **D** glucose unit (compound **3**) proved to be detrimental to the biological activity. We assumed that this modification, although being distant, adversely affects the conformation or conformational dynamics of the L-iduronic-acid residue. In order to better understand the conformational consequences of the sulfate-sulfonatomethyl group exchange, we accomplished a detailed computational and NMR-based conformational analysis.
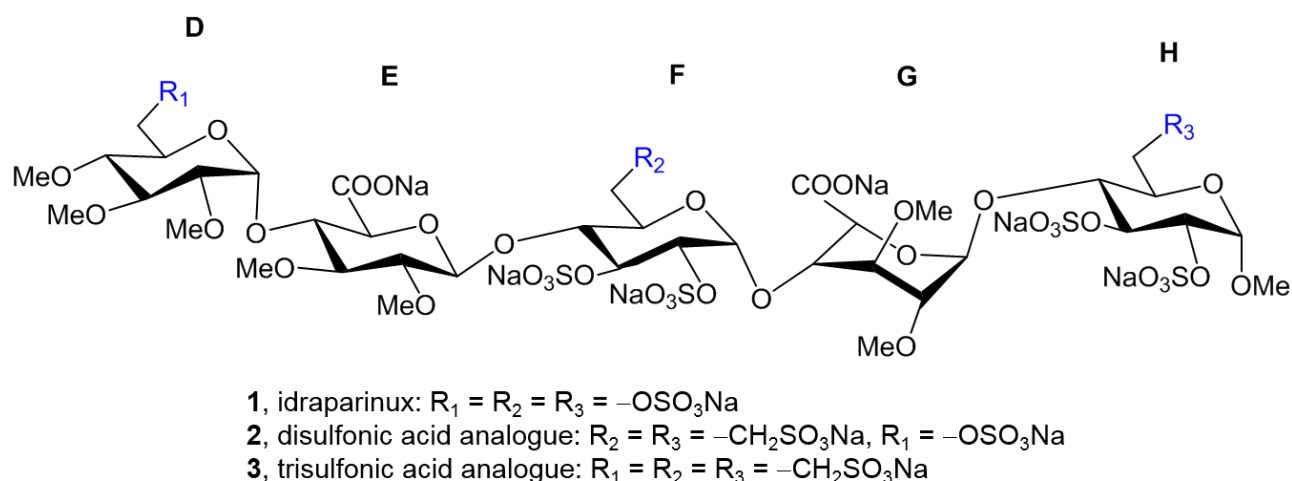
Figure 1 Structure of the investigated pentasaccharide idraparinux (**1**) with indication of the pertinent substitution pattern for analogues **2** and **3**. The sugar rings are denoted D, E, F, G and H, whereby D, F and H subunits are glucopyranosides, E is a D-glucuronic acid and G is an L-iduronic acid ring.

The dynamic nature of carbohydrates makes computational description of their conformational properties a challenging task.[24] As a result, structural ensembles rather than single conformations are best suited to describe carbohydrates, which exist in multiple transient interconverting conformations. In our previous report, we investigated the suitability of carbohydrate force fields in MD simulations, with special emphasis on enhanced sampling methods such as Gaussian accelerated molecular dynamics (GAMD), to characterize the conformational properties of idraparinux. These studies found that the CHARMM force field provides the best description.[25] Furthermore, we developed and carefully validated parameters to model a specific torsion angle parameter in CHARMM for the sulfonatomethyl moiety present in the modified derivatives.

As a consequence of the above mentioned intrinsic flexibility, the experimentally determined NMR parameters related to geometrical properties reflect time averaged values[26] and thus require simulations for their interpretations. Such as, Nuclear Overhauser Effects (NOEs) related to interatomic distances are subject to time averaging due to motions occurring on different timescales. Vicinal scalar coupling constants related to torsion angles are also time averaged by dynamics. Specifically, homonuclear ^{1}H,^{1}H couplings between ring protons are suitable to characterize the conformation and/or conformational dynamics of sugar rings, while heteronuclear ^{1}H,^{13}C couplings can describe the orientation of interglycosidic linkages.

In the present study we perform a rigorous conformational analysis of idraparinux (**1**), together with two of its novel derivatives (**2** and **3**) by comparing NMR data with structural and dynamic descriptors

obtained from an ensemble produced by 10 replicas of x 2 µs long independent unbiased MD simulations. The attainable conformations of the G (L-iduronic acid) subunit were analyzed and compared in the three analogues with particular focus on the pertinent homonuclear vicinal scalar coupling constants, the characteristic interatomic distances derived from NOEs and also by monitoring the evolution of the Cremer-Pople θ parameter during the MD simulations.

Beyond observation of conformational transitions, it is desirable and informative to reconstruct the underlying kinetics of such processes, especially to investigate whether such kinetics varies between related analogues. Although enhanced sampling methods accelerate the frequency of rarely observed conformational transitions, they also present challenges for the extraction of the underlying conformational kinetics.[27] The construction of Markov state models (MSMs) built on unbiased MD simulations is a powerful and well-established approach[28] to rigorously characterize the molecular kinetics of many kinds of processes ranging from conformational transitions to protein folding and biomolecular associations.[29-31] Therefore, here, we construct MSMs to analyze the conformational kinetics of G-ring puckering across the three analogues to provide further insight into the sought-after structure-dynamics-biological activity relationships (SDAR) of the derivatives.

Finally, we characterize and compare the dynamic behavior of the interglycosidic linkages of the derivatives using heteronuclear scalar coupling constants and by monitoring the corresponding torsion angles in the MD simulations. The potential role of conformational flips in the G ring in light of its correlated changes with the interglycosidic linkages, which induce to global conformational changes, is discussed.

Methods

Equilibrium MD simulations. We performed 10 replicas of unbiased equilibrium MD simulations for the three pentasaccharide derivatives (idraparinux **1**, and its sulfonic acid derivatives **2** and **3**) using the CHARMM carbohydrate force field. In the case of the idraparinux system, TIP3P solvent was generated and ions were added using the "Quick MD Simulator" of the CHARMM-GUI. Because the 6-deoxy-6-sulfonatomethyl modification is not supported by the CHARMM-GUI, a local installation of CHARMM[32] was used for the molecules **2** and **3** applying the force field parameters that were published in our previous article[25]. Aside from that, the preparatory steps were identical to the idraparinux system. The sizes of the cubic boxes were defined so that there was always a 10 Å space between the edges of the box and the pentasaccharide atoms. The number of water molecules in the three systems were 2948, 2944, and 2939, respectively.

Equilibration simulations were performed using the AMBER 18 software package. Each system was

minimized in a sequence of 4 stages each with 500 steps with restraints applied to non-heavy atoms of the solute of 10, 5, 1 and 0 kcal/mol/A^2 respectively. This was followed by equilibration using a timestep of 2 fs by constraining non-heavy atoms via the SHAKE algorithm, firstly with gradual heating from 10 K to 300 K for 1 ns at constant volume with non-heavy atom solute restraints of 10 kcal/mol/A^2, then in the NVT ensemble for 1 ns at 300 K using a Langevin thermostat with a damping constant of $\gamma = 5$ ps^{-1}. This was followed by a run in the NPT ensemble using a Berendsen barostat with p = 1 bar and a pressure relaxation time of 1 ps, maintaining these restraints, then sequentially relaxing restraints to 5, 1 and 0.5 kcal/mol/A^2 each for 100 ps and finally unrestrained equilibration in NPT for 5 ns. This was followed by a set of 10 replica production MD simulations each for 2 μs, for each system also in the unrestrained NPT ensemble and with a randomized velocity distribution in each case to enhance independent sampling. The Coulomb and vdW cutoffs were set to 1.0 nm. The PME method was used or long-range electrostatics calculations.

Trajectory analyses for calculation of hydrogen-hydrogen distances and dihedral angles were monitored with python scripts written using MDAnalysis[33]. The Cremer-Pople[34] θ and φ parameters were determined using an in-house written python code. Clustering was performed based on the Cremer-Pople θ parameter. Pearson product-moment correlation coefficients were determined using the numpy module in python.

Markov State Models. MSMs were constructed to determine the kinetics of the G-ring conformational transitions for each system by building on the pyEMMA software[35]. There is a well-established methodology[28] that describes the myriad ways in which MSMs can be constructed, validated and interpreted to provide insight into the kinetics of complex macromolecular processes that have many degrees of freedom. However, as a first step, MSMs require projecting the atomistic coordinates into a set of collective variables (CVs) that describe the conformational transition. Given the low-dimensionality of the CV-space, namely the Cremer-Pople parameters with which G-ring puckering can be followed as well as the exhibition of multiple transitions through the course of the MD simulations, a canonical treatment of MSM construction could be applied, which is summarized as follows. The CV-space is discretized into a set of non-overlapping microstates through geometrical clustering. The probability of transitioning between these microstates within a given time interval or lag time, can then be calculated from observing the MD data. The lag time is chosen such that the interstate jump processes are Markovian or memoryless. Diagonalization of the resulting transition matrix leads to a set of eigenvalues and eigenvectors that describe the rate of transitions, with the first eigenvector corresponding to the stationary (equilibrium) distribution and subsequent eigenvectors to various transition processes in order of increasing rate. Microstates can then be further clustered into coarse-grained macrostates on the basis of kinetic similarity using a formalism based on projected

Markov models that are approximated via hidden Markov models.[36] This yields the transition kinetics between a handful of resulting macrostates which then describe biologically meaningful conformational states of the molecular process. Here, the two-dimensional feature space of Cremer-Pople θ and φ parameters was used to characterize the transitions. Microstates were determined using the k-means++ clustering algorithm[35] using 200 clusters. MSMs were built using different lag times (τ) and the relaxation timescale was analyzed with varying τ. All three systems exhibited a converged separation of timescales after the first non-stationary eigenvector with a τ = 10 ns, corresponding to systems with two kinetically well-described metastable states. Subsequent MSMs were therefore built with this lag time. The stationary distribution across microstates was calculated. Potential of mean forces (PMFs) for G-ring puckering were determined by computing Boltzmann-weighted density distributions in the Cremer-Pople θ-φ space where each data point was weighted by its corresponding microstate equilibrium probability. A hidden Markov model (HMM) was constructed based on a two-state description from which the stationary distribution of both macrostates was calculated together with the rate constants corresponding to the transition between them. Relative free energies were calculated between the two macrostates using the relation $\Delta G = -k_BT \ln(\rho_2/\rho_1)$, where ρ_1 and ρ_2 are stationary probabilities of states 1 and 2, respectively, with a zero point assigned to the less favorable state.

NMR spectroscopy. The NMR samples contained ca. 20 mg of each derivative dissolved in 500 μL D$_2$O. NMR spectra were recorded on a Bruker Avance II 500 (^{1}H: 500.13 MHz; ^{13}C:125.76 MHz) spectrometer equipped with a 5 mm TXI probe or on a Avance NEO 700 (^{1}H: 700.25 MHz; ^{13}C: 176.08 MHz) spectrometer equipped with a 5 mm TCI prodigy probe at 278 K, 298 K and 310 K, respectively to monitor the impact of conformational exchange dynamics. The unambiguous assignment of ^{1}H and ^{13}C resonances was achieved by the combined use of 1D and 2D homo- and heteronuclear NMR spectroscopic methods including ^{1}H-^{1}H COSY, TOCSY, ROESY and ^{1}H-^{13}C HSQC experiments. NOE volume intensities were extracted from ROESY experiments recorded with 300 ms mixing time using the EASY ROESY implementation.[37] Structurally relevant $^3J_{H,H}$ couplings were determined using multiplets extracted from 1D ^{1}H and/or 2D phase sensitive TOCSY spectra.[38] Heteronuclear (^{1}H-^{13}C) vicinal coupling constants along the interglycosidic linkages were determined by PIPAP-HSQMBC[39] method.

Results and discussion

We determined ^{1}H,^{1}H coupling constants of all the sugar rings in the derivatives as listed in supplementary Table S1. Here we focus only on the analysis of the data measured for the G subunit (L-iduronic acid) in Table 1. All the $^3J_{1,2}$, $^3J_{2,3}$ and the $^3J_{3,4}$ couplings are rather similar within the

error of the measurement for the derivatives. Comparing these values with the relevant couplings reported by Hsieh et al.[40] for fixed G ring conformations determined by MD using GLYCAM 06[41] indicates a conformational averaging over different ring populations in all of them. Since the measured $^3J_{1,2}$ couplings (2.9 Hz, 3.0 Hz and 3.3 Hz) are between the large (7.6 Hz) and the small values (0.0 / 1.3 Hz) predicted for $^4C_1$ and $^1C_4$ / ^2S_O, respectively, it is not possible to determine which is the dominant conformation. The same is true for $^3J_{2,3}$. In the case of $^3J_{3,4}$ coupling, however, the measured values (3.0 Hz, 2.9 Hz, 3.2 Hz) point to the dominant presence of ^2S_O with a predicted value of 2.7 Hz, as the couplings for both chair conformations are predicted to be larger (5.5 Hz for $^1C_4$ and 9.5 Hz for $^4C_1$).

Table 1 Calculated proton-proton vicinal coupling constants for fixed G ring conformations ($^1C_4$, $^4C_1$ and ^2S_O) determined by Hsieh et al.[40] using MD (AMBER 14 with GLYCAM 06) and experimental proton-proton vicinal coupling constants determined for the G ring in the derivatives **1**, **2**, **3**. The coupling constants were determined from spectra with a digital resolution of 0.2 Hz.

	$^3J_{H,H}$ (Hz)	$^1C_4$	$^4C_1$	^2S_O	**1**	**2**	**3**
	$^3J_{1,2}$	0.0	7.6	1.3	2.9	3.0	3.3
G ring	$^3J_{2,3}$	3.7	8.5	1.4	5.7	4.9	6.0
	$^3J_{3,4}$	5.5	9.5	2.7	3.0	2.9	3.2

In order to validate the MD simulations by comparison with experimental data, we also calculated ^1H-^1H distances from NOEs as shown in supplementary Table S2. The majority of computed distances (17 for parent compound **1**, 17 for derivative **2** and 16 for derivative **3**) are in good agreement ($\pm$ 0.3 Å) with the experimental data considering uncertainty of NOE derived distances of $\pm$ 0.1 - 0.15 Å as well as the standard deviations of individual MD distances, which are between 0.1 – 0.9 Å. Larger deviations are observed on inter-ring NOEs, specifically, between the D/E ring and the G/H rings. Since the E and G rings are involved in conformational averaging, which are connected to simultaneous changes in the interglycosidic torsions (*vide infra*), small deviations in the conformer populations at these regions in the MD could result in poorer reproduction of inter-ring distances and thus account for the difference between the MD averaged and the experimentally derived NOE distances.

We calculated the ratios of proton-proton distances (G5-G2/ G5-G4) corresponding to key NOEs in the G subunit and compared them to distances obtained for fixed G ring conformations ($^1C_4$, $^4C_1$ and ^2S_O)[40] (Table 2). Considering ca. 0.1-0.15 Å uncertainty of the measurement, the values obtained for the three derivatives are comparable and close to the value characteristic for ^2S_O skewed boat ring conformation.

Table 2 Distances corresponding to key NOEs in the G subunit and ratios of these distances for fixed G ring conformations ($^1C_4$, $^4C_1$ and ^2S_O)[40] and for derivatives **1**, **2**, **3**. At the top major G ring conformations ($^1C_4$, $^4C_1$ and ^2S_O) are shown. The standard deviation of NOE derived distances are ca. 0.1-0.15 Å.
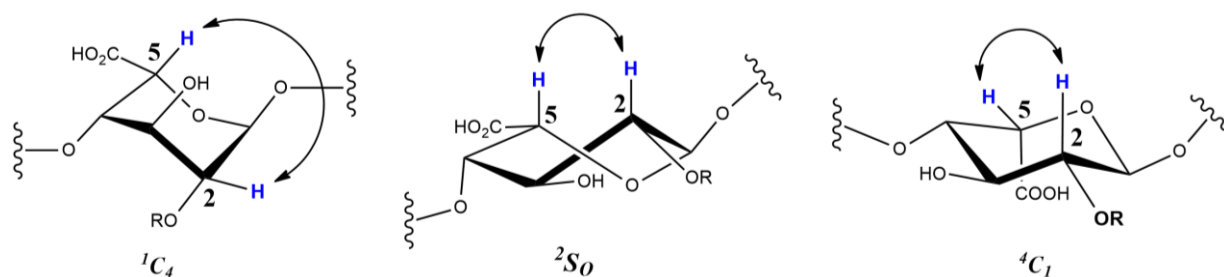

Distances (Å) and distance ratios of proton pairs corresponding to key NOEs	$^1C_4$	$^4C_1$	^2S_O	1	2	3
G5-G4	2.4	2.4	2.3	2.4	2.5	2.4
G5-G2	4.0	4.2	2.6	2.9	2.8	2.7
G5-G2 / G5-G4	1.67	1.75	1.13	1.21	1.12	1.13

We studied the ring conformations of the subunits by analyzing the distribution of the Cremer-Pople angles during the course of the equilibrium MD simulations as shown in Figure 2. For characterizing the conformation of the sugar rings, we calculated the Cremer-Pople parameters[34] for each rings of all derivatives. Namely, three Cremer-Pople parameters can be used to describe the puckering of a six-membered ring. The value of the θ parameter is close to 0° ($^4C_1$) or 180° ($^1C_4$) when the ring is in a chair conformation and ~90° in case of boat or skewed boat conformation. The φ angle distinguishes between the various boat and twist-boat conformations, while the third parameter, the total puckering amplitude, Q, describes the difference of the ring from a regular hexagon. For characterizing the

overall conformation of the pentasaccharides, the distribution of the interglycosidic torsion angles φ and ψ were also examined. The time evolutions of these angles are shown in supplementary Figure S1.

In the course of the equilibrium MD simulation of **1** (shown in supplementary Figure S1a), a similar picture emerges as in the previous GAMD simulations[25], only the conformational transitions occur less frequently in the equilibrium MD. The D subunit is in $^4C_1$ chair conformation during the entire simulation without changing a $^1C_4$ state and very rarely to skewed boat conformations. The E ring is mainly in $^4C_1$ chair with a single transition to a boat conformation in two of the trajectories for a couple of nanoseconds. The F and H rings are stably maintaining $^4C_1$ chair conformation. The G subunit displays changes between ^2S_O skewed boat and $^1C_4$ chair conformations, but less frequently than in the GAMD simulations[25]. The ring conformations of the two sulfonatomethyl derivatives, **2** and **3**, show qualitatively similar conformational behavior as the parent compound, **1** (columns 1, 2, 3 in Figure 2).
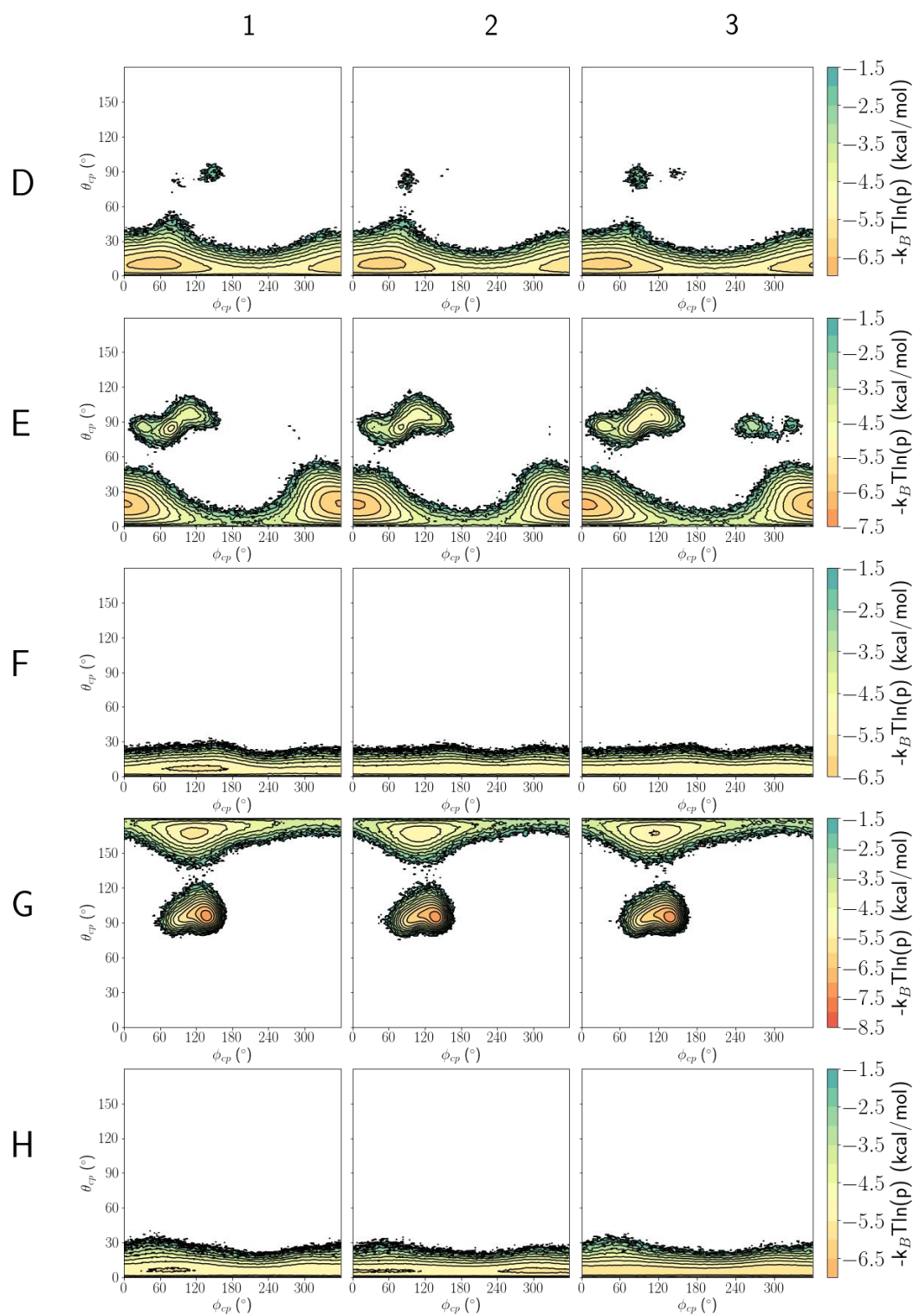
Figure 2 Maps of the Cremer-Pople θ and φ parameters during the 10 replica equilibrium MD simulations for the different subunits in derivatives **1**, **2** and **3**.

In order to quantify the thermodynamics and kinetics of the G ring conformational transitions, a Markov state model was constructed in the 2D Cremer-Pople θ-φ space using 200 microstates. Relaxation analysis showed a timescale separation between the first and second non-stationary

eigenvector, thus supporting a two-metastable state model for each system (Figure 3a). The stationary distribution weighted potential of mean force (PMF) was calculated on the Cremer-Pople landscape (Figure 3b). This shows that the ^2S_O skewed boat conformation is characterized by a deep energy well with PMF of ~3 kcal/mol in each system and a shallower well for the $^1C_4$ chair conformation. The relative free energies of the two conformations as well as their transition kinetics were calculated using a hidden Markov model (HMM). Accordingly, the ^2S_O skewed boat conformation was found to be the thermodynamically dominant state for all derivatives with a ΔG of -0.7, -1.1 and -1.0 kcal/mol for derivatives **1**, **2** and **3,** respectively. The difference between the two metastable states, however, is only marginal compared to the thermal noise (k_BT ~ 0.6 kcal/mol). The relative populations of the $^1C_4$ chair versus the ^2S_O skewed boat states are thus comparable for the three derivatives and cannot account for the difference observed in their biological activities. Besides, both the forward and backward transition kinetics from ^2S_O skewed boat to $^1C_4$ chair were similar with a k_{for} of ~10.8, 14.7 and 10.7 μs^{-1} and a k_{rev} of 3.3, 2.2 and 2.1 μs^{-1} for **1**, **2** and **3**, respectively. Thus, the rate of conformational flip indicates a similar conformational barrier between the $^1C_4$ chair and the ^2S_O skewed boat conformations for the three derivatives.
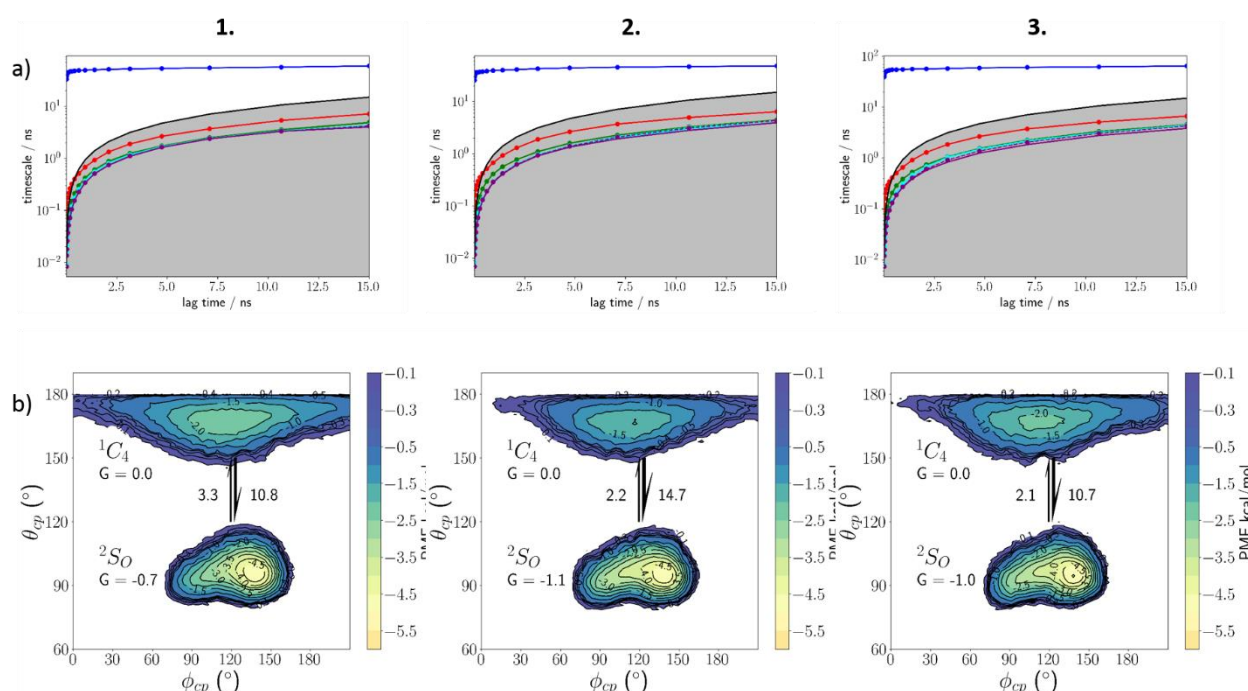
Figure 3 Markov state model (a) Relaxation analysis of ring puckering showing convergence of the slowest-order kinetics (blue line) and significant timescale separation from subsequent faster-order kinetics (other colored lines), suggesting two metastable states. Timescales of all faster-order transitions are smaller than the lag times used to measure them (grey region). (b) Stationary distribution weighted potential of mean force (PMF) calculated on the Cremer-Pople landscape.

Thus, these findings regarding conformational plasticity of the G subunit for the derivatives free in solution could not contribute to the explanation as to why disulfonatomethyl derivative **2** is more active than the parent compound **1** and why the trisulfonatomethyl derivate **3** has lost its activity completely.

The biological activities of the heparin analogue pentasaccharides were proposed to be largely determined by the conformation of the G ring.[14-18] This view has, however, been recently challenged. Results of Liu, Guerrini and co-workers[40, 42] revealed that predominant population of the ^2S_O skew boat conformer of the iduronic acid (G ring) in free form is not a prerequisite for the activation of AT. They demonstrated that a synthetic heparin hexasaccharide with iduronate residue displaying 73% of $^1C_4$ conformer in solution could efficiently activate antithrombin and the iduronic acid adopts ^2S_O conformation only when bound to AT. Our present results demonstrate that the predominant ^2S_O conformation of the G iduronic subunit as observed for derivative **3** is also not a guarantee for biological activity.

Finally, to characterize the global conformation of the pentasaccharides that is strongly determined by the torsion angles around the interglycosidic linkages, we measured the heteronuclear $^1H,^{13}C$ vicinal coupling constants across all four linkages in the three derivatives. These NMR data are summarized in Table 3 together with averaged values of the corresponding interglycosidic torsions extracted from the MD simulations. Furthermore, we also extracted the corresponding interglycosidic torsion angles from the MD simulations shown as ϕ, ψ map in Figure 4 with their time evolution given in supplementary Figure S1.
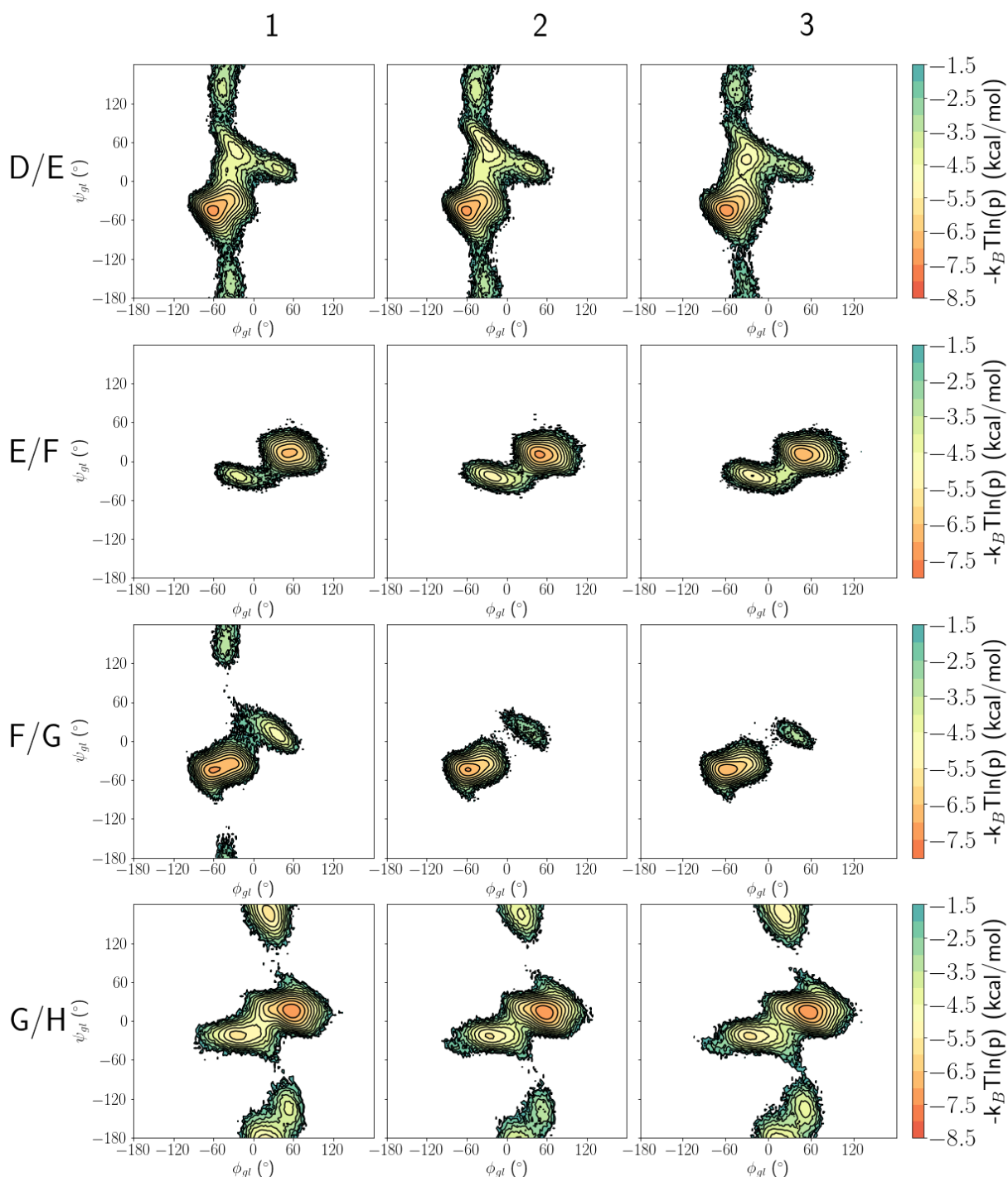
The coupling constant values are all very similar (within ± 0.1 Hz) in the three derivatives with the exception of the couplings between subunits G and H. The similarity observed in the conformationally averaged coupling constants suggests a comparable population distribution of staggered conformations around the D/E, E/F and F/G interglycosidic linkages. This is indeed confirmed by the similar appearance of the ϕ, ψ maps in Figure 4. The F/G linkage is an exception as for the ψ torsion shows a low energy region at +/- 180° for derivative **1**, which is absent in **2** and **3**. However, this minimum has a relatively small population, which does not significantly affect the average of the ψ dihedral (-39°, -42°, -42° for **1**, **2** and **3**, respectively) and only increases its standard deviation in **1** (± 14, ± 8, ± 8 for **1**, **2** and **3**, respectively). Thus, the dihedrals in the D/E, E/F and G/H linkages are mostly very similar for the three derivatives both according to the NMR data as well as by the MD simulations.

For the G/H linkage, somewhat different coupling constants are obtained for the trisulfonatomethyl derivative **3**, which are 0.4 - 0.5 Hz higher for $^3J_{H1-C1-O-C4}$ and 0.7 - 0.9 Hz smaller for $^3J_{C1-O-C1-H4}$ than for **1** and **2**. These differences are neither reproduced by the MD torsion angle averages, which are

48º, 52º, 51º for ϕ angle and 18º, 13º, 15º for the ψ angle for **1**, **2** and **3**, nor by the distribution of ϕ, ψ angles in the maps of Figure 4. As these ϕ, ψ maps show, the potential energy surface of the G/H interglycosidic linkage along the torsions is characterized by four low energy basins. Insufficient sampling could slightly distort the shape of these energy minima and alter their populations leading to inaccuracies in the averaged torsions of the MD trajectories.

Table 3 Experimental heteronuclear ^{1}H,^{13}C vicinal coupling constants (Hz) measured along the interglycosidic linkages for the derivatives **1**, **2** and **3**. N.a. stands for data not available. Averaged torsion angles and their standard deviations from MD are in parentheses. The corresponding torsion angles are H1-C1-O-C4 for φ and C1-O-C1-H4 for ψ with O standing for the interglycosidic oxygen atom, atoms before the O belong to the preceding unit and after to the following unit. The coupling constants were determined from spectra with digital resolution of 0.2 Hz.

| Interglycosidic linkage	Coupling constant	Corresponding torsion angle	1 (Hz, degree)	2 (Hz, degree)	3 (Hz, degree)
D/E	$^3J_{\text{H1-C1-O-C4}}$	φ	3.2 (-52º ± 14)	3.1 (-52º ± 15)	3.0 (-53º ± 13)
	$^3J_{\text{C1-O-C1-H4}}$	ψ	4.1 (-38º ± 22)	n.a. (-37º ± 26)	n.a. (-39º ± 21)
E/F	$^3J_{\text{H1-C1-O-C4}}$	φ	3.3 (54º ± 15)	3.4 (49º ± 19)	3.4 (46º ± 22)
	$^3J_{\text{C1-O-C1-H4}}$	ψ	5.1 (12º ± 10)	5.2 (9º ± 11)	5.3 (8º ± 12)
F/G	$^3J_{\text{H1-C1-O-C4}}$	φ	3.4 (-47º ± 18)	3.3 (-52º ± 14)	3.3 (-52º ± 14)
	$^3J_{\text{C1-O-C1-H4}}$	ψ	3.8 (-39º ± 14)	3.7 (-42º ± 8)	3.7 (-42º ± 8)
G/H	$^3J_{\text{H1-C1-O-C4}}$	φ	4.1 (48º ± 24)	4.2 (52º ± 21)	4.6 (51º ± 22)
	$^3J_{\text{C1-O-C1-H4}}$	ψ	5.5 (18º ± 43)	5.7 (13º ± 21)	4.8 (15º ± 35)

Figure 4 Maps of the interglycosidic φ and ψ angles during the 10 replica equilibrium MD simulations for the different subunits in derivatives **1**, **2** and **3.** The interglycosidic torsion angles are defined according to the heteronuclear vicinal coupling constants in Table 3 measured between H and C atoms involved in the interglycosidic linkage as H1-C1-O-C4 for φ and C1-O-C1-H4 for ψ with O standing for the interglycosidic oxygen atom, atoms before the O belong to the preceding unit and after to the following unit.

We also examined the timing of the conformational changes between the ring flips and changes in the interglycosidic linkages by calculation of Pearson correlation coefficients between the Cremer-Pople θ parameter and the interglycosidic φ and ψ dihedral angles as shown in Figure 5.

Notably, the conformational changes of the E ring from $^4C_1$ chair to ^2S_O skewed boat and back to $^4C_1$ chair are correlated with changes of the φ and ψ torsion angles in the following E/F interglycosidic linkage (indicated by medium blue squares) and somewhat anticorrelated with that of the preceding D/E linkage (shown by pink squares). The significance of this observation, however, cannot be ascertained since the E ring flip is not sampled adequately in the simulations (only occur in 2 of the 10 replicas for analogue **1**, with 4 flips for derivative **2** and 5 for **3** as shown in supplementary Figure 1).

The conformation flips of the G subunit from ^2S_O skewed boat to $^1C_4$ chair and back appear to be only correlated with the φ torsion angle of the following G/H linkage (indicated by a dark blue square), while very weak or no correlation exists with the ψ dihedral or both angles of the preceding F/G linkage. This suggests that local changes in the conformation of the G ring might induce a transition in the global conformation (or *vice versa*) of the pentasaccharide via changes in the interglycosidic angles, particularly affecting the reducing end of the molecule.
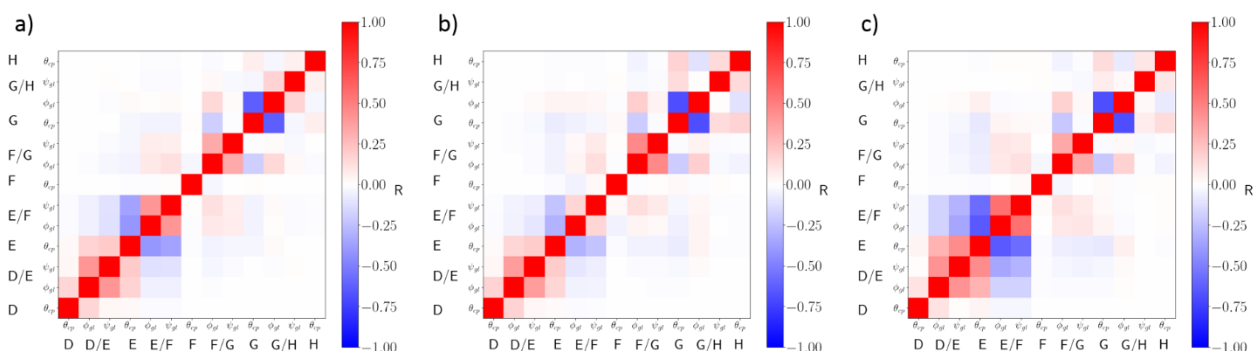

Figure 5 Pearson correlation between the Cremer-Pople θ parameter and in the interglycosidic φ and ψ dihedral angles between the subunits a) for **1**, b) for **2** and c) for **3**. Correlated changes are characterized by values close to 1 (in red), anticorrelated changes are indicated by values close to -1 (in blue) and no correlation is present with values close to 0 (in white).

In order to classify the overall conformation of the pentasaccharides, we calculated root mean square deviations (RMSD) of the conformations in the trajectories relative to a common reference structure, which is the heavy atom frame of the bound state pentasaccharide (PDB structure 1E03) as shown in

supplementary Figure S2. Unfortunately, the RMSD is unable to differentiate between the expected conformational states of the pentasaccharides based on ring flips and changes in the interglycosidic torsions. The distribution of the RMSD values appears to be very similar between the three analogs.

The structural ensembles of the pentasaccharides obtained in the MD simulation could be most suitably partitioned along the two observed correlated conformational flips, one slower flip on the non-reducing end and one faster process on the reducing end. Correspondingly, the ensemble can be divided into four groups according to the conformation of the E ring, which can be in major ($^{4}C_{1}$, Cremer-Pople $\theta < 45°$) or minor ($^{2}S_{O}$, $60° < \theta < 120°$) states, and to the conformation of the G ring which can be major ($^{2}S_{O}$, $60° < \theta < 120°$) and minor ($^{1}C_{4}$, $\theta > 150°$) states. Representative structures of the four conformational groups for each derivative are shown in Figure 6. The group featuring chair conformation in the E subunit and a skewed boat in the G ring provides the largest cluster, the other two groups - one with both E and G rings either in chair or skewed boat conformations - are less populated, and finally, the smallest cluster is with the E subunit in skew boat and the G ring in chair conformation. The conformational groups are structurally rather similar for the three derivatives and are largely extended. Unfortunately, the lack of multiple transitions in the E ring precluded constructing an MSM of this four-state system. Thus, subsequent simulations over a longer timescale are needed that may elucidate the biological relevance, if any, of the differences in E ring kinetics observed in the various derivatives.
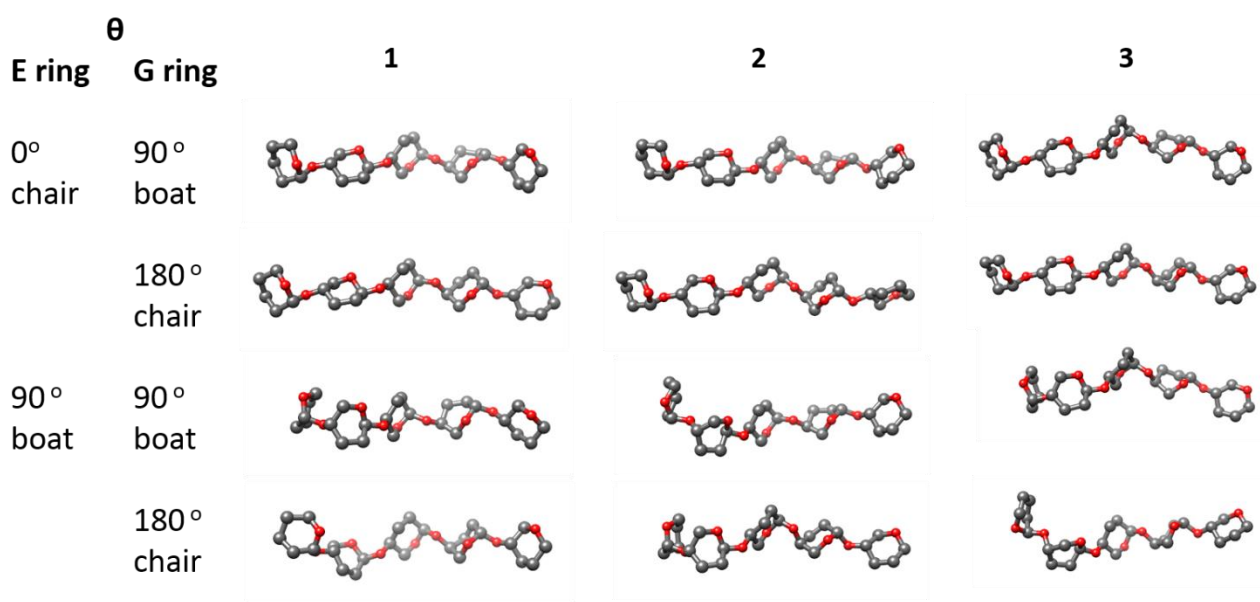
Figure 6 Representative structures of the four conformational groups for the three derivatives.

Conclusions

Understanding how the structural and dynamical properties of heparin analogues affect their blood coagulant properties is of immense importance. In this contribution we investigated the effect of sulfate-sulfonatomethyl group substitution on the conformational dynamics of the sugar rings and that of interglycosidic linkages by NMR spectroscopy and MD simulations.

In the course of our investigations, we observed that there are two major conformational changes with different time scales present in the pentasaccharides: one occurs at the non-reducing end (units D-E-F) with a slower timescale, while the other influences the reducing end (units G-H) with a faster timescale. The conformational changes in the ring conformations simultaneously translate to conformational flips in the neighboring interglycosidic linkages. It is not clear which conformational change is the cause and the effect: ring conformational change initiates interglycosidic linkage flips or the overall conformational change causes the ring transitions. There is, however, an asymmetry for E and G rings. The E ring conformational flips are accompanied by changes in both D/E and E/F linkages, while the conformational change of the G ring is only correlated with flips in the G/H linkage. Whether the different conformational states of E ring with associated changes in the D/E and E/F interglycosidic bonds may play a role in determining the biological activity is still unclear and to resolve this ambiguity longer MD simulations are needed.

There is no noticeable difference observed in the conformational dynamic behavior of the G subunit between the three derivatives which could be correlated with their different biological activity. Also, the overall conformations of the pentasaccharides as reflected by the arrangements of the interglycosidic linkages appear rather similar. For more than 30 years, relationship between the preferred ^2S_O conformation of the G subunit and anticoagulant activity was regarded as a major paradigm for the interpretation of biological properties of heparin analogues. It has recently been shown[40, 42] that the solution phase conformation of the G subunit is not necessarily correlated with biological properties. Our results also confirm this observation and show that the dominant ^2S_O conformation of the G subunit is not necessarily sufficient to produce an active analogue. There is more and more evidence showing that the conformational preference of the G subunit alone cannot be used to predict biological activity, thus this SDAR needs to be treated with care or revised.

The inactivity of derivative **3** with additional 6-sulfonatomethyl substitution at ring D may more likely be related to the sequence of AT binding events. According to a recent study[38], the complexation with AT is initiated at the D-E-F terminal of the pentasaccharide, which is then followed by conformational change in AT and lastly binding of the G-H end occurs. The additional 6-sulfonatomethyl group of **3** might obstruct the association of the D-E-F terminal to AT, thus exterminating the subsequent events of binding and therefore preventing its tight binding to AT,

which would be required for the allosteric activation of AT against its target coagulation factors. The atomic level details of the binding interactions of the three heparin analogue derivatives will be the topic of a future study.

Data and Software Availability

NMR data analyzed in this work is available at: https://unidebhu-my.sharepoint.com/:f:/g/personal/koverk_o365_unideb_hu/EshAA_MznK9PnqRvzeKadKkBFoqx M9xe2YCzoCdtTcLvYw?e=Bv31ad . NMR data analysis can be carried out using TOPSPIN 3.5 and is available at https://bruker-labscape.store/products/topspin-for-processing free of charge for academic users. The input files for carrying out all MD simulations using the AMBER 18 software package as well as the Cremer-Pople and glycosidic dihedral output data are available at

Supporting Information Available

Supplementary Table S1 Experimental homonuclear vicinal coupling constants in the different rings determined for the derivatives 1, 2 and 3.

Supplementary Table S2 Comparison of the distances from equilibrium MD corresponding to Nuclear Overhauser Effects (NOE) between ^{1}H nuclei.

Supplementary Figure S1 Time evolution of the Cremer-Pople θ and φ parameters and that of interglycosidic φ (in red) and ψ (in blue) torsion angles during the 10 replica equilibrium MD simulations for the different subunits for derivatives 1 in a), for 2 in b) and for 3 in c).

Supplementary Figure S2 Calculated RMSD values of the conformations in the trajectories using bound pentasaccharide in 1E03 as a reference.

References

1.   Lever, R.; Mulloy, B.; Page, C. P., Heparin : a Century of Progress. **2012**.
2.   Weiss, R. J.; Esko, J. D.; Tor, Y., Targeting heparin and heparan sulfate protein interactions. *Organic and Biomolecular Chemistry* **2017**, 15, 5656-5668.
3.   Lanzi, C.; Cassinelli, G., Heparan Sulfate Mimetics in Cancer Therapy: The Challenge to Define Structural Determinants and the Relevance of Targets for Optimal Activity. *Molecules* **2018**, 23, 2915.
4.   Page, C., Heparin and Related Drugs: Beyond Anticoagulant Activity. *ISRN Pharmacology* **2013**, 2013, 13.
5.   Jing, Z.; Yan, K.; Fuming, Z.; Robert, J. L., Impact of Temperature on Heparin and Protein Interactions. *Biochemistry & Physiology: Open Access* **2018**, 7, 1-5.
6.   Kreuger, J.; Kjellén, L., Heparan sulfate biosynthesis: regulation and variability. *Journal of Histochemical Cytochemistry* **2012**, 60, 898-907.
7.   Olson, S. T.; Richard, B.; Izaguirre, G.; Schedin-Weiss, S.; Gettins, P. G., Molecular mechanisms of antithrombin-heparin regulation of blood clotting proteinases. A paradigm for understanding proteinase regulation by serpin family protein proteinase inhibitors. *Biochimie* **2010**, 92, 1587-96.
8.   Gray, E.; Hogwood, J.; Mulloy, B., The anticoagulant and antithrombotic mechanisms of heparin. *Handbook of experimental pharmacology* **2012**, 43-61.
9.   Lindahl, U.; Thunberg, L.; Bäckström, G.; Riesenfeld, J.; Nordling, K.; Björk, I., Extension and structural variability of the antithrombin-binding sequence in heparin. *Journal of Biological Chemistry* **1984**, 259, 12368-76.
10.   Guerrini, M.; Mourier, P.; Torri, G.; Viskov, C., Antithrombin-binding oligosaccharides:

Structural diversities in a unique function? *Glycoconjugate journal* **2014**, 31.

11.   Jin, L.; Abrahams, J. P.; Skinner, R.; Petitou, M.; Pike, R. N.; Carrell, R. W., The anticoagulant activation of antithrombin by heparin. *Proc Natl Acad Sci U S A* **1997**, 94, 14683-14688.

12.   McCoy, A. J.; Pei, X. Y.; Skinner, R.; Abrahams, J. P.; Carrell, R. W., Structure of beta-antithrombin and the effect of glycosylation on antithrombin's heparin affinity and activity. *J Mol Biol* **2003**, 326, 823-33.

13.   Petitou, M.; van Boeckel, C. A. A., A synthetic antithrombin III binding pentasaccharide is now a drug! What comes next? *Angewandte Chemie-International Edition* **2004**, 43, 3118-3133.

14.   Ferro, D. R.; Provasoli, A.; Ragazzi, M.; Casu, B.; Torri, G.; Bossennec, V.; Perly, B.; Sinaÿ, P.; Petitou, M.; Choay, J., Conformer populations of l-iduronic acid residues in glycosaminoglycan sequences. *Carbohydrate Research* **1990**, 195, 157-167.

15.   van Boom, J. H.; Sakairi, N.; van Der Marel, G. A.; Basten, J. E. M.; van Boeckel, C. A. A., Synthesis of a Conformationally Constrained Heparin-like Pentasaccharide. *Chemistry – A European Journal* **1996**, 2, 1007-1013.

16.   Das, S. K.; Mallet, J.-M.; Esnault, J.; Driguez, P.-A.; Duchaussoy, P.; Sizun, P.; Hérault, J.-P.; Herbert, J.-M.; Petitou, M.; Sinaÿ, P., Synthesis of Conformationally Locked Carbohydrates: A Skew-Boat Conformation of L-Iduronic Acid Governs the Antithrombotic Activity of Heparin. *Angewandte Chemie International Edition* **2001**, 40, 1670-1673.

17.   Das, S. K.; Mallet, J.-M.; Esnault, J.; Driguez, P.-A.; Duchaussoy, P.; Sizun, P.; Herault, J.-P.; Herbert, J.-M.; Petitou, M.; Sinaÿ, P., Synthesis of Conformationally Locked l-Iduronic Acid Derivatives: Direct Evidence for a Critical Role of the Skew-Boat 2S0 Conformer in the Activation of Antithrombin by Heparin. *Chemistry – A European Journal* **2001**, 7, 4821-4834.

18.   Hricovíni, M.; Guerrini, M.; Bisio, A.; Torri, G.; Petitou, M.; Casu, B., Conformation of heparin pentasaccharide bound to antithrombin III. *The Biochemical journal* **2001**, 359, 265-72.

19.   Guerrini, M.; Elli, S.; Gaudesi, D.; Torri, G.; Casu, B.; Mourier, P.; Herman, F.; Boudier, C.; Lorenz, M.; Viskov, C., Effects on molecular conformation and anticoagulant activities of 1,6-anhydrosugars at the reducing terminal of antithrombin-binding octasaccharides isolated from low-molecular-weight heparin enoxaparin. *J Med Chem* **2010**, 53, 8030-8040.

20.   Demeter, F.; Gyöngyösi, T.; Bereczky, Z.; Kövér, K. E.; Herczeg, M.; Borbás, A., Replacement of the L-iduronic acid unit of the anticoagulant pentasaccharide idraparinux by a 6-deoxy-L-talopyranose - Synthesis and conformational analysis. *Scientific Reports* **2018**, 8, 10.

21.   Herczeg, M.; Lazar, L.; Bereczky, Z.; Kover, K. E.; Timari, I.; Kappelmayer, J.; Liptak, A.; Antus, S.; Borbas, A., Synthesis and Anticoagulant Activity of Bioisosteric Sulfonic-Acid Analogues of the Antithrombin-Binding Pentasaccharide Domain of Heparin. *Chemistry-a European Journal* **2012**, 18, 10643-10652.

22.   van Boeckel, C. A. A.; Petitou, M., The Unique Antithrombin III Binding Domain of Heparin: A Lead to New Synthetic Antithrombotics. *Angewandte Chemie International Edition in English* **1993**, 32, 1671-1690.

23.   Cros, S.; Petitou, M.; Sizun, P.; Pérez, S.; Imberty, A., Combined NMR and molecular modeling study of an iduronic acid-containing trisaccharide related to antithrombotic heparin fragments. *Bioorganic and Medicinal Chemistry* **1997**, 5, 1301-9.

24.   Imberty, A.; Perez, S., Structure, conformation, and dynamics of bioactive oligosaccharides: Theoretical approaches and experimental validations. *Chemical Reviews* **2000**, 100, 4567-4588.

25.   Balogh, G.; Gyöngyösi, T.; Timári, I.; Herczeg, M.; Borbás, A.; Fehér, K.; Kövér, K. E., Comparison of Carbohydrate Force Fields Using Gaussian Accelerated Molecular Dynamics Simulations and Development of Force Field Parameters for Heparin-Analogue Pentasaccharides. *Journal of Chemical Information Modeling* **2019**, 59, 4855-4867.

26.   Torchia, D. A., NMR studies of dynamic biomolecular conformational ensembles. *Prog Nucl Magn Reson Spectrosc* **2015**, 84, 14-32.

27.   Bruce, N. J.; Ganotra, G. K.; Kokh, D. B.; Sadiq, S. K.; Wade, R. C., New approaches for computing ligand–receptor binding kinetics. *Current Opinion in Structural Biology* **2018**, 49, 1-10.

28.   Prinz, J. H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F., Markov models of molecular kinetics: generation and validation. *The Journal of chemical physics* **2011**, 134, 174105.

29.   Sadiq, S. K.; Noé, F.; De Fabritiis, G., Kinetic characterization of the critical step in HIV-1 protease maturation. *Proc Natl Acad Sci U S A* **2012**, 109, 20449-20454.

30.   Noé, F.; Schütte, C.; Vanden-Eijnden, E.; Reich, L.; Weikl, T. R., Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proceedings of the National Academy of Sciences* **2009**, 106, 19011-19016.

31.   Plattner, N.; Doerr, S.; De Fabritiis, G.; Noé, F., Complete protein-protein association kinetics in atomic detail revealed by molecular dynamics simulations and Markov modelling. *Nature Chemistry* **2017**, 9, 1005-1011.

32.   Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E., GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, 1-2, 19-25.

33.   Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O., MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry* **2011**, 32, 2319-2327.

34.   Cremer, D.; Pople, J. A., General definition of ring puckering coordinates. *Journal of the American Chemical Society* **1975**, 97, 1354-1358.

35.   Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Pérez-Hernández, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J.-H.; Noé, F., PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *Journal of Chemical Theory and Computation* **2015**, 11, 5525-5542.

36.   Noé, F.; Wu, H.; Prinz, J.-H.; Plattner, N., Projected and hidden Markov models for calculating kinetics and metastable states of complex molecules. *The Journal of chemical physics* **2013**, 139, 184114.

37.   Thiele, C. M.; Petzold, K.; Schleucher, J., EASY ROESY: Reliable Cross-Peak Integration in Adiabatic Symmetrized ROESY. *Chemistry-a European Journal* **2009**, 15, 585-588.

38.   Uhrin, D.; Batta, G.; Hruby, V. J.; Barlow, P. N.; Kover, K. E., Sensitivity- and gradient-enhanced hetero (omega(1)) half-filtered TOCSY experiment for measuring long-range coupling constants. *Journal of Magnetic Resonance* **1998**, 130, 155-161.

39.   Castañar, L.; Saurí, J.; Williamson, R. T.; Virgili, A.; Parella, T., Pure in-phase heteronuclear correlation NMR experiments. *Angewandte Chemie-International Edition* **2014**, 53, 8379-82.

40.   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. *Scientific Reports* **2016**, 6, 8.

41.   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. *Journal of Computational Chemistry* **2008**, 29, 622-55.

42.   Stancanelli, E.; Elli, S.; Hsieh, P. H.; Liu, J.; Guerrini, M., Recognition and Conformational Properties of an Alternative Antithrombin Binding Sequence Obtained by Chemoenzymatic Synthesis. *Chembiochem* **2018**, 19, 1178-1188.