



The 8th International Electronic Conference on Medicinal Chemistry (ECMC 2022)

01-30 NOVEMBER 2022 | ONLINE

Stochastic dynamics mass spectrometric 3D structural analysis of N-glycans of fetal bovine serum – an experimental and theoretical study

Chaired by **DR. ALFREDO BERZAL-HERRANZ**;
Co-Chaired by **PROF. DR. MARIA EMÍLIA SOUSA**



pharmaceuticals



Bojidarka Ivanova ^{1,*}, and Michael Spiteller ¹

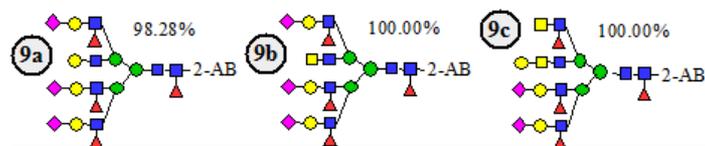
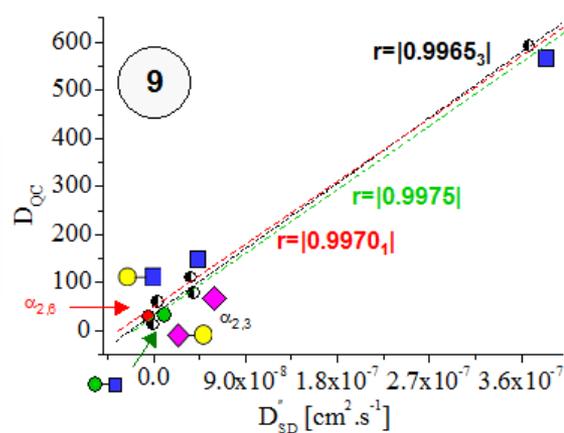
¹ Lehrstuhl für Analytische Chemie, Institut für Umweltforschung, Fakultät für Chemie und Chemische Biologie, Universität Dortmund, Otto-Hahn-Straße 6, 44221 Dortmund, Nordrhein-Westfalen, Deutschland.

* Corresponding author: B.Ivanova@infu.uni-dortmund.de; b_ivanova@web.de

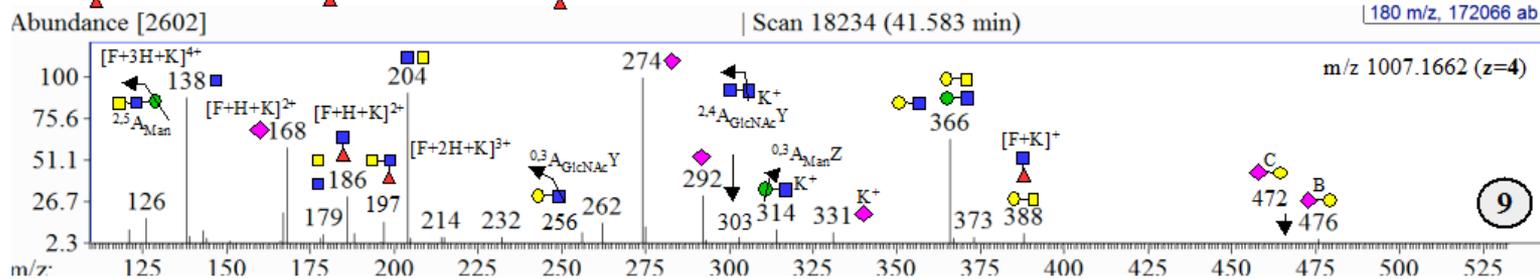
Stochastic dynamics mass spectrometric 3D structural analysis of N-glycans of fetal bovine serum – an experimental and theoretical study

Graphical Abstract

$$D_{QC} = \left[\frac{\prod_{i=1}^{3N} v_i^{(0)}}{\prod_{i=1}^{3N-1} v_i^{(s)}} \right] \times e^{-\frac{\Delta H}{R \times T}}$$



$$D_{SD}^{n,tot} = \sum_i^n D_{SD}^{n,i} = \sum_i^n \left(2.6388 \times 10^{-17} \times \left(\overline{I_i^2} - \left(\overline{I_i} \right)^2 \right) \right)$$



Abstract:

The contribution addresses problem of *mass spectrometric* (MS) 3D molecular and electronic structural analysis of mixtures of *glycans* of *fetal bovine serum*. The research task is unexpectedly difficult, due to random variation of non template-driven glycosylation and fucosylation processes; a lack of regioselective derivatization, causing for mixtures of polydisperse glycans toward length and skeletal modifications; isomers of oligomers and polymers, including linear and branches molecular structures, respectively. That fact difficult significantly their structural analysis, mass spectromerically. Furthermore, MS phenomena of carbohydrates (CBs) include reactions of intramolecular rearrangement and cyclization, proton and charge transfer effects, noncovalent binded self-associates, alkali metal ion adducts, and multiply charged species under tandem MS/MS operation mode. Despite, the study presents a new line of plausible solution to the problem. It employs our innovative stochastic dynamic MS model formula $D''_{SD} = 2.6388 \cdot 10^{-17} \cdot (\langle I^2 \rangle - \langle I \rangle^2)$ capable of exactly quantifying *fluctuations* and temporal behaviour of measurable variable intensity (I) of analyte peaks. It exactly and directly quantifies analyte concentration in solution and determines 3D molecular and electronic structures. The later task is less straightforward. It employs Arrhenius's model equation within the framework of his transition state theory and power capability of *quantum chemical* methods. The validity of the latter statements is examined, herein. The study, first, comes to grips MS collision-induced dissociation (CID) phenomena of mixtures of 2-aminobenzamide derivatized glycans. It utilizes *ab initio* and DFT static and *molecular dynamics*, *molecular mechanics*, and *chemometrics*, as well.

Keywords: 3D structural analysis; Glycals; Mass spectrometry; Stochastic dynamics

ECMC
2022

The 8th International Electronic
Conference on Medicinal Chemistry
01-30 NOVEMBER 2022 | ONLINE

Introduction

Naturally occurring carbohydrates (**CBs**) are abundant class of chemicals, amongst others, having great structural complexity and diversity of isomers of linear and branched oligomers and polymers [1]. *Circa* 70 % mammalian cell proteins are glycosylated [1,2]. There are three types of mammalian glycoproteins called N-, O- and C-glycans, respectively. The latter, less abundant structures are so-called *glycosylphosphatidylinositol lipid anchors*. Also, there are glycoconjugated derivatives such as oligosaccharides, proteoglycans and glycolipids, respectively. Owing to structural diversity of CBs — a hexasaccharide shows up to 10^{12} isomers — there has been concentrated enormous effort on characterizing biological function of glycans and its connection with disease development in humans [3,4]. There are four major categories of biological activity:

- (i) Structural and modulatory capability;
- (ii) extrinsic and
- (iii) intrinsic recognitions; and
- (iv) molecular mimicry, respectively [5].

The fields of *glycomics*, dealing with quantitative and structural determination of glycans, and *glycoproteomics*, determining macro- and micro-heterogeneities of glycoproteins gain considerable attention recently. **Monitoring of aberrant glycosylation — major process of posttranslational modifications of proteins involved into their transport and folding, as well as, targeting [6] and using biomarkers — allows diagnostics and prognostics of many diseases, because of glycans attached to proteins and lipids on surfaces of living cells interact with glycan-binding proteins.**

[1] D. Harvey, *Mass Spectrom. Rev.* 18 (1999) 349–451.

[2] M. Goli, A. Yu, B. Cho, S. Gautam, J. Wang, C. Gutierrez-Reyes, P. Jiang, W. Peng, Y. Mechref, LC-MS/MS in glycomics and glycoproteomics analyses, chapter 8; in El Rassi (Ed.) *Carbohydrate Analysis by Modern Liquid Phase Separation Techniques* (Second Edition) Elsevier, Amsterdam, (2021), pp. 391-441.

[3] A. Araujo, V. Castro, R. Reis, R. Pires, *ACS Med. Chem. Lett.* 12 (2021) 548–554.

[4] J. Munkley, *Oncology Lett.* 17 (2019) 2569–2575.

[5] Y. Xie, M. Butler, *Glycobiology* 32 (2022) 289–303.

[6] Y. Hu, T. Shihab, S. Zhou, K. Wooding, Y. Mechref, *Electrophoresis* 37 (2016) 1498–1505.

Glycan-binding proteins mediate physiological and pathophysiological processes associated with signal transduction, cell adhesion, immune responses, viral and bacterial infections, as well as, diseases such as inflammation, cancer, (auto)immune and neurodegenerative diseases, respectively [3,4,7–11].

The fundamental understanding of glycosylation processes can significantly improve human health diagnostics and therapeutics by monitoring of glycan biomarkers.

Particularly, highlights are on early stage detection of cancer, which can be carried out reliably by such biomarkers.

Enormous effort has been concentrated on assessing variation of glycosylation or fucosylation levels, as well as, glycans of proteins in serum or blood, in order to, monitor progression of different cancers.

Glycans have been suggested as biomarkers, however, their isomeric structures [12] have not been determined, so far. **Sialic acid bonding type ($\alpha_{2,6}$ - and $\alpha_{2,3}$ -linkages) has been connected with types of carcinoma cells. Presence of $\alpha_{2,6}$ -linked sialic acid has been found in biological fluids of patients with hepatitis B cirrhosis. Conversely, human samples of patients with hepatitis C cirrhosis show $\alpha_{2,3}$ -linked type sialic acids. Thus, profiling of isomeric glycans might effectively discriminate between diseases [13]. The same is true for *fucosylation process* as potential marker for cancers and diseases, which is connected more directly with changes in disease [14].**

[7] N. Suzuki, T. Abe, K. Hanzawa, S. Natsuka, Sci. Rep. 11 (2021) 6334.

[8] L. Wei, Y. Cai, L. Yang, Y. Zhang, H. Lu, Anal. Chem. 90 (2018) 10442–10449.

[9] S. Zhou, X. Dong, L. Veillon, Y. Huang, Y. Mechref, Anal. Bioanal. Chem. 409 (2017) 453–466.

[10] G. Lageveen-Kammeijer, N. de Haan, P. Mohaupt, S. Wagt, M. Filius, J. Nouta, D. Falck, M. Wuhrer, Nat. Commun. 10 (2019) 2137.

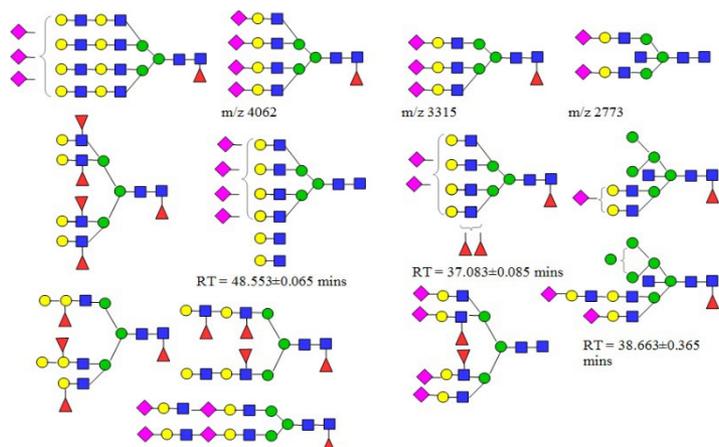
[11] C. Ashwoo, C. Lin, M. Thaysen-Andersen, N. Packer, J. Am. Soc. Mass Spectrom. 29 (2018) 1194–1209.

[12] N. Seo, H. Lee, M. Oh, G. Kim, S. Lee, J. Ahn, H. Cha, K. Kim, J. Kim, H. An, Frontiers Mol. Biosci. 8 (2021) 778851.

[13] Y. Huang, S. Zhou, J. Zhu, D. Lubman, Y. Mechref, Electrophoresis. 38 (2017) 2160–2167.

[14] H. Yin, J. Zhu, J. Wu, Z. Tan, M. An, S. Zhou, Y. Mechref, D. Lubman, Electrophoresis. 37 (2016) 2624–2632.

However, **glycosylation** and **fucosylation** represent non-template-driven processes [15], causing for diversity of glycan structures consisting of long or short molecules, branched or linear structures; furthermore, variously linked. They are classified into three major types in N-glycans: High-mannose, complex, and hybrid derivatives, containing a pentasaccharide core-structure (**Figure 1**) of (GlcNAc)₂ and (Man)₃-fragments. Conversely, O-glycans are connected to peptides *via* a (GalNAc)-residue [16]. Complex-type N-glycans, can be further divided into core, branching, and derivatives, having variable regions. In vertebrates, (Man)₃-core is very frequently $\alpha_{1,6}$ -fucosylated or there is bisecting (GlcNAc)-fragment attached to β -Man of (Man)₃-core. So far, there are detected up to five (GlcNAc)-containing branches connected with α -Man of (Man)₃-fragment. Penta-antennary structures are rarely found in mammals. They are determined in birds and fishes. Naturally occurring glycans usually contain several isomers with same mass values [7].



Fuc	▽	Fucose
Gal	○	Galactose
GalNAc	◻	N-acetyl galactosamine
Glc	⬢	Glucose
GlcNAc	⬠	N-acetylglucosamine
Man	◇	Mannose
NeuAc	◇	N-acetylneuraminic acid/ Sialic Acid
NeuGc	◇	N-glycolylneuraminic acid

Figure 1. Symbolic diagrams of complex multi-antennary N-glycans from feutin and IgG according to Harvey and co-workers [11]; m/z and chromatographic retention time (RTs) [mins] data.

Structural diversity of glycans enables us to encode rich biological information about cell functions. However, the molecular structural complexity and isomerism represent a major analytical challenge in *glycomics* and *glycoproteomics*. Also, glycans exhibit low affinity of glycan-binding proteins ($K_d \sim 1$ mM) [17,18]. Thus, development of *glycosciences* is insignificant comparing with fields of *genomics* and *proteomics* [17,18].

[15] M. Wührer, A. Deelder, Y. van der Burgt, *Mass Spectrom. Rev.* 30 (2011) 664–680.

[16] X. Sun, L. Tao, L. Yi, Y. Ouyang, N. Xu, D. Li, R. Linhardt, Z. Zhang, *J. Pharmaceut. Anal.* 7 (2017) 87–94.

[17] E. Mucha, A. Stuckmann, M. Marianski, W. Struwe, G. Meijera, K. Pagel, *Chem. Sci.* 10 (2019) 1272–1284.

[18] H. Lu, Y. Zhang, P. Yang, *Nat. Sci. Rev.* 3 (2016) 345–364.

ECMC
2022

The 8th International Electronic
Conference on Medicinal Chemistry
01–30 NOVEMBER 2022 | ONLINE

Methods of choice for quantifying glycans are reversed-phase liquid chromatography, high-performance ion-exchange chromatography employing in fluorescence detection, pulsed amperometric detection, or capillary electrophoresis [19]. The chromatographic analysis involves 2D or 3D maps. NMR spectroscopy [20] is utilized for structural determining. It requires a significant analyte amount of few milligrams, which makes the method unsuitable for so-called *minute amounts* of analytes which are extracted from antibodies or cancer cells. The simplest and most direct approach to study glycans involves **mass spectrometric quantification and 2D structural analysis** [16]. The method is several orders of magnitude higher sensitive than NMR. Also, it is applicable to trace sample amounts. There are used commercially available kits for labelling glycans, for instance, based on 2-aminobenzamide (2-AB.) The 2-AB is routinely employed in HILIC analysis with Fs-detection. (**HILIC is employed in the current study.**) It shows high resolution and selectivity for CB isomers [21,22]. 2-AB labelled analytes of fetuin have been described [19]. MS spectra of glycans are implemented into searching algorithms for analyte annotation [23], but the accuracy is low. The software frequently uses integrating peak area under MS fragment ions [2]. Many glycomics and glycoproteomics methods are similar to algorithms for *proteomics* database searching ones, which are adapted to *glycomics* and *glycoproteomics* [24]. The effectiveness of those approaches is *ca.* 20-50 %, studying mixtures [25,26]. Improvement of accuracy of searching algorithms has been achieved by employment in chromatographic and MS data on RTs and MS fragment ions. The RTs correlate linearly ($|r|=0.998$) with molecular structural complexity of glycans, showing that at high RTs and *m/z* values, there are chiefly multi-antennary (and hybrid) glycans [6]. Despite, superior method performances as robust analytical instrumentation, conventional MS based quantitative and molecular structural MS methods for data-processing or searching algorithms are incapable of distinguishing glycan isomers [25a]. There are effort in employing specific MS fragments produced from CID-MS/MS method [25], however, often, these ions show low abundance. Moreover, frequently, proteomic methods shows effectiveness less than 20 % examining CID product ions of peptides by automated searching algorithms [25b]. The low rate of determination is explained with the complex fragmentation chemistry of peptides showing competitive fragmentation paths, which are not accounted for the searching algorithms. Carbohydrates, however, also exhibit competitive processes of proton and charge transfer effects, intramolecular rearrangement and cyclization, thus making their chemical fragmentation reactions also complex [27] (below.)

[19] J. Ahn, J. Bones, Y. Yu, P. Rudd, M. Gilar, J. Chromatogr. B, 878 (2010) 403–408.

[20] Y. Yu, T. Tyrikos-Ergas, Y. Zhu, G. Fittolani, V. Bordonni, A. Singhal, R. Fair, A. Grafmgller, P. Seeberger, M. Delbianco, Angew. Chem. 131 (2019) 13261–13266.

[21] F. Higel, U. Demelbauer, A. Seidl, W. Friess, F. Soergel, Anal. Bioanal. Chem. 405 (2013) 2481–2493.

[22] R. Kozak, C. Tortosa, D. Fernandes, D. Spencer, Anal. Biochem. 486 (2015) 38–40.

[23] M. Ardejani, L. Noodleman, E. Powers, J. Kelly, Nat. Chem. 480 (2021) 480–487.

[24] D. Polasky, F. Yu, G. Teo, A. Nesvizhskii, Nat. Met. 17 (2020) 1125–1132.

[25] (a) C. Liew, C. Yen, J. Chen, S. Tsai, S. Pawar, C. Wu, C. Ni, Nat. Commun. Chem. 4 (2021) 92; (b) L. Yu, Y. Tan, Y. Tsai, D. Goodlett, N. Polfer, J. Proteome Res. 10 (2011) 2409–2416.

ECMC
2022

The 8th International Electronic
Conference on Medicinal Chemistry
01–30 NOVEMBER 2022 | ONLINE

Stochastic dynamic mass spectrometry – an innovative concept

In what follows in this section, we attempt to highlight what we mean to illustrate, herein, and how it contributes crucially to develop fields of *glycomics* and *glycoproteomics*. First, this study examines quantitatively functional relations among MS measurable variable *intensity* (I) of analyte fragment ions and 3D molecular and electronic structures of glycans, employing our innovative stochastic dynamic (**SD**) theory and model equations (1) and (2).

$$D_{SD}^{tot} = \sum_i^n D_{SD}^i = \sum_i^n \left(1.3194 \cdot 10^{-17} \cdot A^i \cdot \frac{\overline{I_i^2} - (\overline{I_i})^2}{(\overline{I_i} - \overline{I_i})^2} \right) \quad (1)$$

There is complete distinction between relations (1) and (2), and all known model relations used to *glycomics* and *glycoproteomics*, including those ones, employed in searching database algorithms.

$$D_{SD}^{n,tot} = \sum_i^n D_{SD}^{n,i} = \sum_i^n \left(2.6388 \times 10^{-17} \times \left(\overline{I_i^2} - (\overline{I_i})^2 \right) \right) \quad (2)$$

There are introduced functions between SD *diffusion* parameters (D_{SD}^i , $D_{SD}^{n,i}$) and measurable variable intensity (I) of i th MS peak of an analyte ion determined over short spans of scan time of a MS measurement. To argue for a great importance of formulas (1) and (2) for fields of *glycomics* and *glycoproteomics* does not mean that one should simply evaluate excellent-to-exact chemometric parameters reported, so far, to studies devoted to SD mass spectrometric 3D conformational and electronic structural analysis of LMWs [26–31]. The later strongly suggests that, in order to, fully grasp the fundamentally different theoretical approach which we provide in this paper, one should understand adequately theoretical background behind equations (1) and (2), itself. For the sake of brevity, we refer the reader's attention to works [26–31].

[26] B. Ivanova, M. Spiteller, *Anal. Chem. Letts.* 12 (2022) 322–348.

[27] B. Ivanova, M. Spiteller, *Bioorg. Chem.* 93 (2019) 103308.

[28] B. Ivanova, M. Spiteller, *Steroids* 164 (2020) 108750.

[29] B. Ivanova, M. Spiteller, *Anal. Chem. Letts.* 10 (2020) 703–721.

[30] B. Ivanova, M. Spiteller, *Rev. Anal. Chem.* 38 (2019) 1.

[31] B. Ivanova, M. Spiteller, *J. Mol. Liq.* 292 (2019) 111307.

ECMC
2022

The 8th International Electronic
Conference on Medicinal Chemistry
01–30 NOVEMBER 2022 | ONLINE

Utilization of equations (1) and (2) for purposes of exact MS based 3D structural determination of analytes is carried out complementarily with Arrhenius's model equation (3). Thus, correlating between theoretical and experimental diffusion parameters from perspective of chemometrics and statistical parameters, there is assessed quantitatively the statistical significance of relationship $D''_{SD}=f(D_{QC})$ of a selected set of MS peaks of molecular and fragment ions unique to given 3D molecular and electronic structures of analyte. At this point, we make sense of saying *unique fragment ions*, because of the relation $D''_{SD}=f(D_{QC})$ takes exact quantitative account, on the one hand, of physico-chemical measurable parameters mass-to-charge and MS peak intensity values of observable fragment species. On the other hand, its accounts exactly for 3D molecular and electronic structures of analytes producing corresponding observable MS ions, because of, equation (3) determines *free Gibbs energy* parameters of molecular species, which are unique to any 3D conformation and electronic structure of any analyte.

$$D_{QC} = \left[\frac{\prod_{i=1}^{3N} v_i^{(0)}}{\prod_{i=1}^{3N-1} v_i^{(s)}} \right] \times e^{-\frac{\Delta H}{R \times T}} \quad (3)$$

D_{QC} denotes *theoretical quantum chemical diffusion* parameter; $v_i^{(0)}$ and $v_i^{(s)}$ are frequencies of ionic species in ground and transition states; ΔH represents difference in enthalpies of the two states; T and R mean temperature and universal gas constant, respectively.

Experimental

Data availability

This work uses experimental data-set of MS measurements of glycans according to [32]. There is processed file **AB_23Apr20_0074_test_Fetuin.raw** [<https://zenodo.org/record/4705546>], which is public available free of charge.

Reference [32] provides detail on sample preparation techniques and methods for isotope labelling of glycans, as well as, experimental conditions of mass spectrometric measurements and analytical instrumentation.

Theory/computations

Comprehensive description of theoretical quantum chemical methods and theories used to obtain D_{QC} parameters according to equation (3) have been provided recently [26]. The current contribution utilizes same theoretical protocols program packages and software, respectively.

Chemometrics

Software R4Cal Open Office STATISTICS for Windows 7 was utilized. Statistical significance was checked by *t*- and *W*-tests. The model fit was determined upon, by *F*-test. Analysis of variance (ANOVA) tests were employed. The nonlinear fitting off, among MS datasets was carried out by means of searching method based on Levenberg-Marquardt algorithm. ProteoWizard 3.0.11565.0 (2017) [<https://proteowizard.sourceforge.io/>], mMass 5.5.0 [<http://www.mmass.org/download/>] and AMDIS 2.71 (2012) [<https://amdis.software.informer.com/2.7/>] software was utilized.

Database

The contribution uses database and searching algorithms of **GlycoWorkBench 2** [33] [<https://glycoworkbench.software.informer.com/2.1/>].

Nomenclature

Fragmentation ion assignment uses nomenclature by Dommon and Costello [34], Stephens and co-workers [35], as well as, Spina and co-workers, respectively [36] (**Figure 2**.)

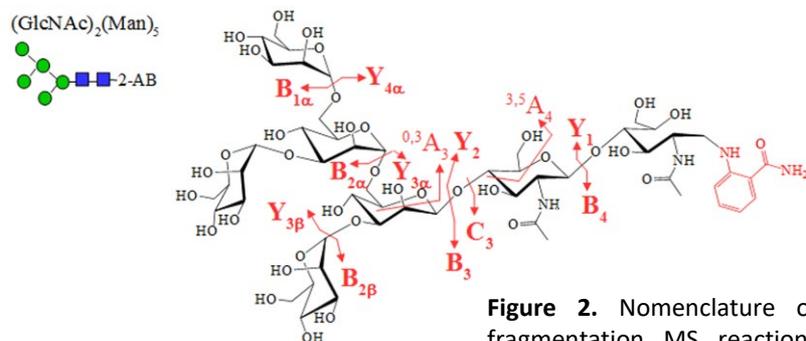


Figure 2. Nomenclature of fragmentation MS reactions of a high Man-type glycan

[32] R. Smith, Released N-glycans from IgG and Fetuin on Agilent QTOF and Orbitrap, (2021) DOI 10.5281/zenodo.4705545 [<https://zenodo.org/record/4705546>].

[33] A. Ceroni, K. Maass, H. Geyer, R. Geyer, A. Dell, S. Haslam, J. Prot. Res. 7 (2008) 1650–1659.

[34] B. Dommon, C. Costello, Glycoconjugate J 5 (1988) 397–409.

[35] E. Stephens, S. Maslen, L. Green, D. Williams, Anal Chem 76 (2004) 2343–2354.

[36] E. Spina, D. Romeo, G. Impallomeni, D. Garozzo, D. Waidelich, M. Glueckmann, Rapid Comm. Mass Spectrom. 18 (2004) 392–398.

ECMC
2022

The 8th International Electronic
Conference on Medicinal Chemistry
01–30 NOVEMBER 2022 | ONLINE

Results and discussion

Mass spectrometric data on fragmentation reactions of glycans – a brief description

Comprehensive analysis of $[M+Na]^+$ adduct of high Man-analyte $(GlcNAc)_2(Man)_5$ [37] shows a major cleavage path yielding to B_4 ion *via* loss of $(GlcNAc)$ -2-AB fragment (m/z 1036.) The internal fragmentation ions are at m/z 364 (Y_1), 833 (B_3) and 567 (Y_2), in addition to, product ions at m/z 874, 712, 550, 388 and 226, respectively. Peaks at m/z 671, 509 and 347 belong to B_3 ion, while peak at m/z 550 – to loss of $(Man)_3$ -residues of branching analytes. Peak at m/z 671 is, due to Man-loss of a 3-antenna structure. It is a $B_{2\beta}$ cleavage of B_3 -ion. Peak at m/z 671 is determined in 2-AB derivative of glycan $(GlcNAc)_2(Man)_6$, containing additional Man-substituent and 3-antenna structure. Peaks at m/z 671 (Man)₄, 833 (Man)₅ and 995 (Man)₆ are observed in hybrid or high Man-glycans. Fragments at m/z 583 and 569 are typical for core branching Man-derivatives; for instance, $(GlcNAc)_2(Man)_5$ -2-AB one. Ions at m/z 745, 583 and 421 show further Man-loss [37–40]. **The subsequent loss of Man-residues does not depend on labelled compound (Figures S1 and S2; s.i.)** ESI(+)-CID-MS/MS spectrum of $(GlcNAc)_2(Man)_5$ $[M+Na]^+$ adduct derivatized with 3-aminoquinoline shows loss of mass of 162. Fragmentation reactions are: m/z 1385.5 ($[M+Na]^+$) \rightarrow 1223.5 ($Y_{4\alpha}/Y_{3\beta}$); 1036 (B_4) \rightarrow 874 ($B_4/Y_{4\alpha}/Y_{3\beta}$), 833 (B_3) \rightarrow 671 ($B_3/Y_{4\alpha}$ and $B_3/Y_{3\beta}$), 671 ($B_3/Y_{4\alpha}$ and $B_3/Y_{3\beta}$) \rightarrow 509 ($(Man)_3^+$), 550 ($B_4/Y_{3\beta}$) \rightarrow 388 ($B_4/Y_{3\alpha}/Y_{3\beta}$), and 509 ($(Man)_3^+$) \rightarrow 347 ($(Man)_2^+$), respectively. The phenomenon does not depend on adducts types. Protonated analyte $(GlcNAc)_2(Man)_5$ -3-AQ $[M+H]^+$ shows reactions: m/z 1364 ($[M+H]^+$) \rightarrow 1201, 1201 \rightarrow 1039, 1039 \rightarrow 877, 877 \rightarrow 715, 528 \rightarrow 366 and 366 \rightarrow 204; m/z 682 ($[M+2H]^{2+}$) \rightarrow 601, 601 \rightarrow 520, 520 \rightarrow 439 and 439 \rightarrow 358, respectively [38]. The $[M+2Na]^{2+}$ adduct shows: m/z 693 ($[M+2Na]^{2+}$) \rightarrow 612, 612 \rightarrow 531, and 531 \rightarrow 450, respectively. The same is true for dications $[(GlcNAc)_2(Man)_6$ -3-AQ+H+Rb] $^{2+}$, $[(GlcNAc)_2(Man)_6$ -3-AQ+H+Na] $^{2+}$, $[(GlcNAc)_2(Man)_6$ -3-AQ+H+K] $^{2+}$, and $[(GlcNAc)_2(Man)_6$ -3-AQ+H+Li] $^{2+}$, respectively. On the base on available comprehensive description, briefly resented, so far, there are assigned ions of 2-AB derivatized $(GlcNAc)_2(Man)_9$ ion of glycan at m/z 772.3699 $z=3$ (RT=15.38 mins; **Figures S1 and S2.**) Characteristic is MS peak at m/z 833 [37]. 2-AB labelled analyte of type $(GlcNAc)_2(Man)_{10}$ can be found in databases of naturally occurring and synthetic glycans [33]. Peaks at m/z 1258 (1258.2), 1419.9 (1420.4), 1581.9 (1582.5), 1744.1 (1744.7), and 1904.3 (1906.8) correspond to $[M+Na]^+$ -adducts.

[37] D. Harvey, J. Am. Soc. Mass Spectrom. 11 (2000) 900–915.

[38] X. Sun, L. Tao, L. Yi, Y. Ouyang, N. Xu, D. Li, R. Linhardt, Z. Zhang, J. Pharmaceut. Anal. 7 (2017) 87–94.

[39] L. Ruhaak, G. Zauner, C. Huhn, C. Bruggink, A. Deelder, M. Wuhrer, Anal. Bioanal. Chem. 397 (2010) 3457–3481.

[40] D. Harvey, Analyst 125 (2000) 609–617.

However, there is a common to all type glycans MS peak at m/z 204 assigned to $[(\text{GlcNAc})]^+$ cation. Also, there are common peaks at 138 and 168 belonging to $[(\text{GlcNAc})]^+$ ion. The latter assignment is carried out looking at fragmentation paths of D-glucose and D-fructose producing ion at m/z 127 of 5-hydroxymethyl-2-furaldehyde [41,42]. MS peak at m/z 657 is frequently assigned to $[(\text{GlcNAc})-(\text{Gal})-(\text{Neu5Ac})]^+$ cation. The cleavage of (Neu5Ac)-fragment yields to $[(\text{GlcNAc})-(\text{Gal})]^+$ ion at m/z 366 [43]. Glycan (1) shows peak at m/z 366 $[(\text{GlcNAc})-(\text{Man})]^+$, however, its spectrum lacks of peak at m/z 657.

Figure 3 shows important common fragmentation ions of glycans.

The logic scheme of 3D structural analysis of glycans according to equations (2) and (3) is highlighted in Figure 4.

Our analysis concentrates on Fuc-glycans (1), (6), (7) and (9)–(11) (Figures 1, 5–7, and S4–S7) looking at Lewis_x (Gal-1→4(Fuc-1→3)GlcNAc) and Lewis_y ((Fuc-1→2)Gal-1→4(Fuc-1→3)GlcNAc) epitopes, because of they can be upregulated in cancer [44]. Like (1) showing subsequent loss of a Man-fragments, high-Fuc-containing N-glycans exhibit subsequent loss of Fuc-subunits ($\Delta m/z=146$.) Isomers of Fuc-derivatives exhibit peaks at m/z 680 and 522 of charged species or sodium adducts of (Fuc)(Gal)(GlcNAc)(Fuc), (Gal)(GlcNAc)(Fuc) or (Fuc)(Gal)(GlcNAc) ions [45]. 3-Antenna glycans show ions at m/z 424, 570, and 716 of (Gal)(GlcNAc)-fragment, containing zero, one, and two Fuc-residues. Peak at m/z 716 belongs to $^{13}\text{A}_{3a}$ fragmentation product of hepta-Fuc-triantennary glycan [73]. The MS peak at m/z 570 is more abundant, when mono-Fuc-substitution on individual glycan antennae mostly occurs. The study, herein, details on ions at m/z 591.5 ($z=3$) (6) and 590.8 ($z=2$) (7) (scans 5425 and 5443 of raw data [32]) as an effort to shed more light on this issue from perspective of functional relations among parameters of equations (1)–(3). Importantly, data-base based annotation is also carried out (Figure S3,) showing reliability within 90–100 % of proposed structures. However, chemometrics of twenty-four both m/z and absolute average intensity data on MS peaks of (6) and (7) yields to coefficients of linear correlation $|r|=1$ and 0.9889₄ (Figure S8.) ANOVA tests of variables of those analytes indicate a lack of significantly difference (Tables 1, S1 and S2.) Ions at m/z 504 and 731 belong to $[2\text{AB}-(\text{Man})(\text{GlcNAc})]^+$ and $[(\text{Man})(\text{GlcNAc})(\text{Man})(\text{GlcNAc})]^+$ ions [46]. Due to, importance of sialic acid bonding type, i.e. α -2,6 and α -2,3 linkages in glycans [47], herein, there is examined ions at m/z 308 and 472 (Figures 8 and 9.) Under negative MS polarity peak at m/z 306 belongs to $[\text{O}^4\text{A}_2\text{-CO}_2]^-$ anion of α -2,6 bonded sialic acid [48]. The MS peak at m/z 308 could be assigned to $[(\text{Gal})(\text{Fuc})]^+$ ion under positive MS polarity. Depending on complexity of molecular structure of analyte such as (8) MS peak at m/z 308 can be assigned to monomer $[\text{NeuGc}]^+$ or dimer $[(\text{NeuGc})_2]^{2+}$ species, as well.

[41] A. Ricci, B. Di Rienzo, F. Pepi, A. Troiani, S. Garzoli, P. Giacomello, J. Mass Spectrom. 50 (2015) 50, 228–234.

[42] F. Pepi, A. Ricci, S. Garzoli, A. Troiani, C. Salvitti, B. Di Rienzo, P. Giacomello, Carbohydr. Res. 413 (2015) 145–150.

[43] M. Nakano, D. Higo, E. Arai, T. Nakagawa, K. Kakehi, N. Taniguchi, A. Kondo, Glycobiology 19 (2009) 135–143.

[44] D. Harvey, W. Struwe, J. Am. Soc. Mass Spectrom. 29 (2018) 1179–1193.

[45] J. Hofmann, A. Stuckmann, M. Crispin, D. Harvey, K. Pagel, W. Struwe, Anal. Chem. 89 (2017) 2318–2325.

[46] N. de Haan, Y. Narimatsu, M. Aasted, I. Larsen, I. Marinova, S. Dabelsteen, S. Vakhrushev, H. Wandall, Anal. Chem. 94 (2022) 4343–4351.

[47] Y. Huang, S. Zhou, J. Zhu, D. Lubman, Y. Mechref, Electrophoresis. 38 (2017) 2160–2167.

[48] J. Zhao, S. Li, C. Li, S. Wu, W. Xu, Y. Chen, M. Shameem, D. Richardson, H. Li, Anal. Chem. 88 (2016) 7049–7059.

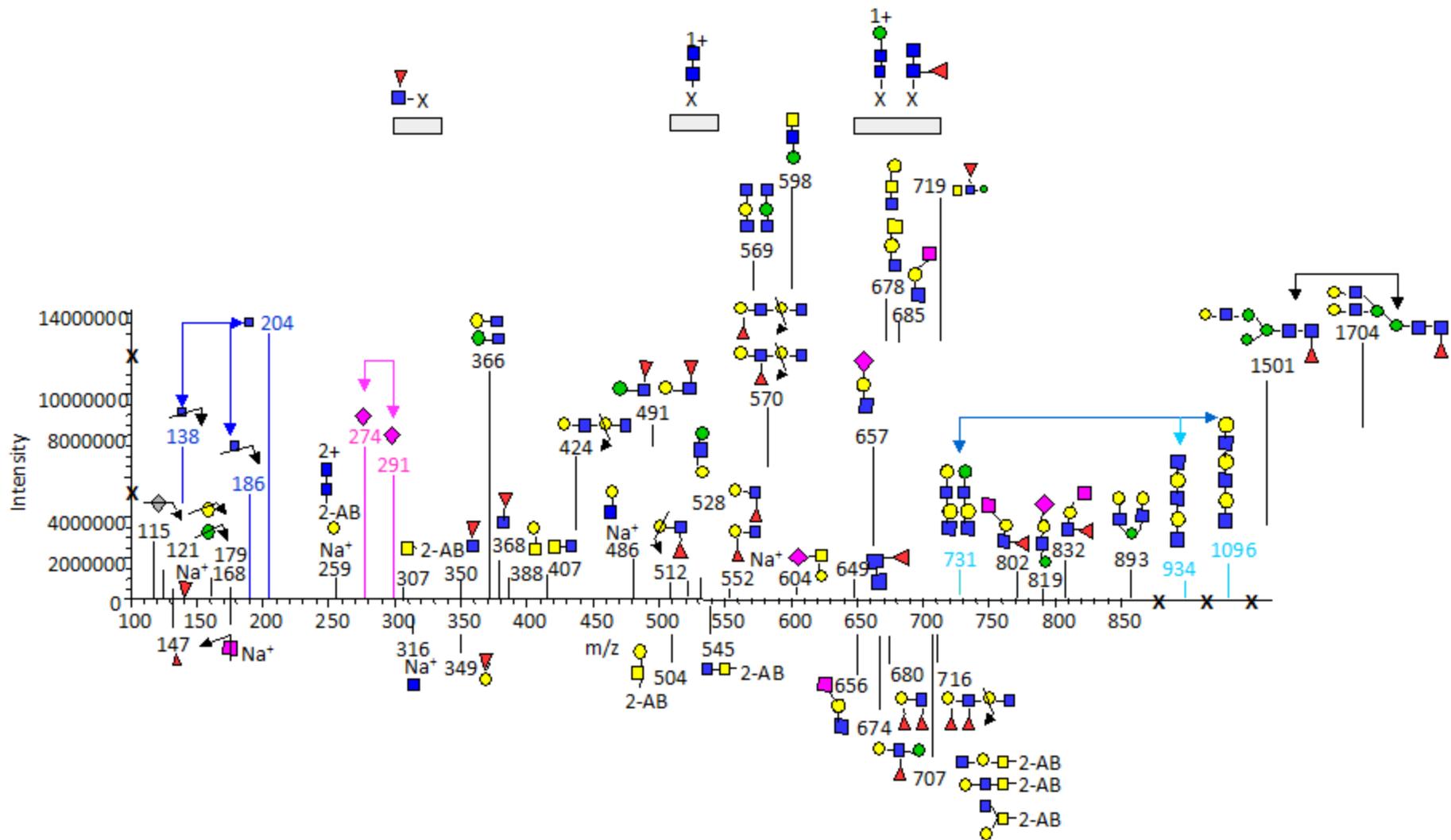


Figure 3. Characteristic fragmentation mass spectrometric ions, often found in spectra of glycans, depending on analyte molecular structure and its isomers; symbolic representation of carbohydrate sub-units according to [11].

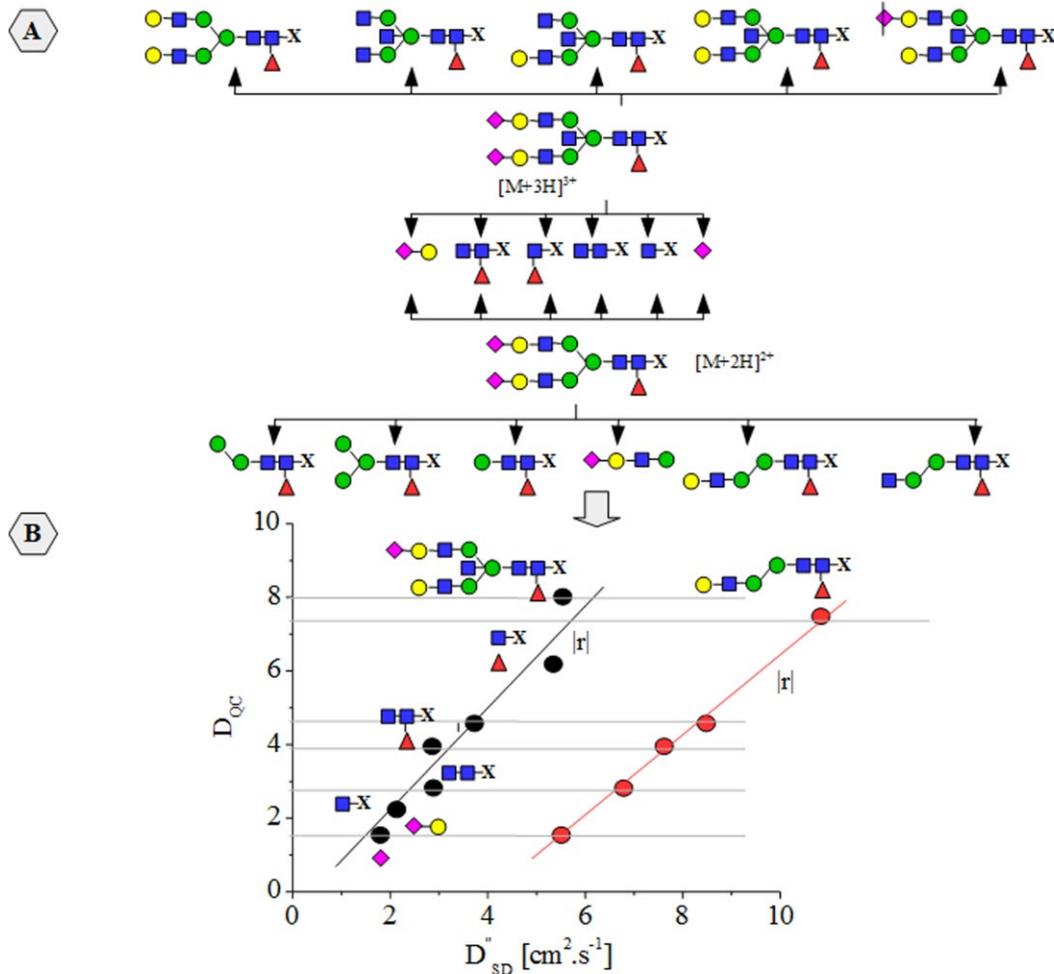


Figure 4. Fragmentation schemes of N-glycans yielding to common and characteristic peaks (A); functional relation between D_{SD}^* and D_{QC} data on equations (2) and (3) (B).

So far, mass spectra have not been extensively used to identify functional groups within the framework of given molecule [49]. However, as results from current study shall illustrate persuasively this approach has both descriptive and predictive power, thus making equations (1)–(3) prominent tools for annotation and 3D structural determination of glycans; furthermore, exactly from perspective of chemometrics.

[49] Y. Lin, L. Smith, Biomed. Mass Spectrom. 5 (1978) 604–611.

ECMC
2022

The 8th International Electronic
Conference on Medicinal Chemistry
01–30 NOVEMBER 2022 | ONLINE

There are examined following fragmentation ions of glycans (1)–(13) of experimental raw data [32]:

- (1) m/z 772.3699 (z=3); RT = 15.382 mins; (6) m/z 591.4823 (z=3); RT = 19.301 mins; (10) m/z 1107.4603 (z=3); RT = 22.868 mins;
 (2) m/z 1470.9091 (z=4); RT = 60.722 mins; (7) m/z 590.8190 (z=2); RT = 19.323 mins; (11) m/z 1107.7969 (z=3); RT = 22.699 mins;
 (3) m/z 1366.3770 (z=3); RT = 38.838 mins; (8) m/z 1006.5351 (z=3); RT = 60.721 mins; (12) m/z 872.35 (z=2) at RT = 10.979 mins; and
 (4) m/z 1107.1256 (z=3); RT = 22.608 mins; (9) m/z 1007.1662 (z=4); RT = 41.583 mins; (13) m/z 1136.96 (z=2) at RT = 10.911 mins;
 (5) m/z 1238.9983 (z=2); RT = 10.968 mins;

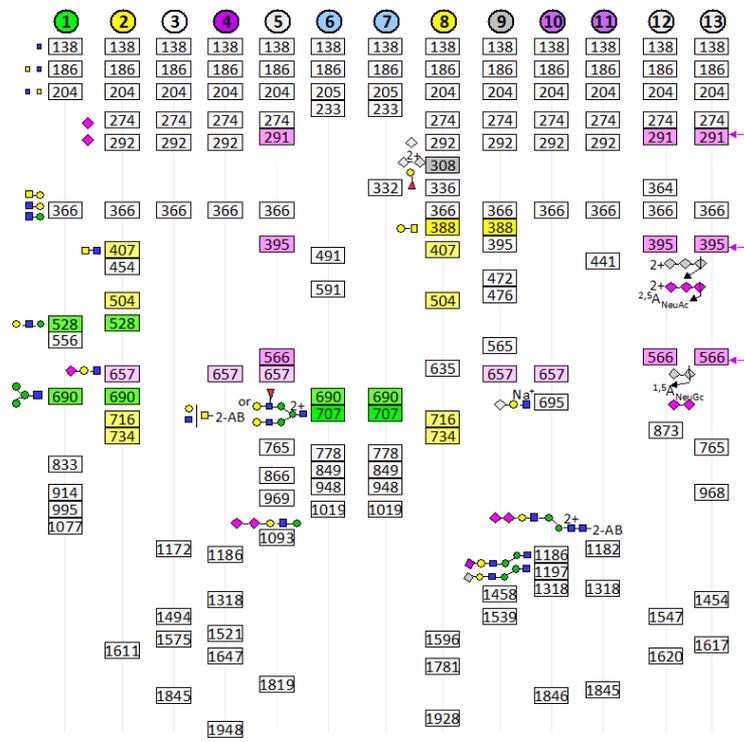
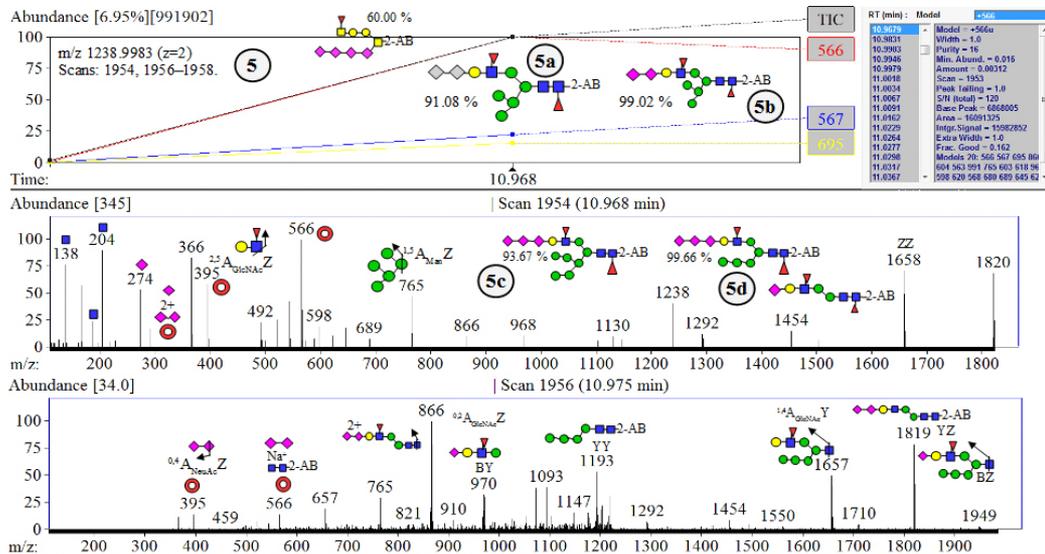


Figure 5. Mass spectra (screenshot) of fragmentation ion of (5); annotation of peaks according to GlycoWorkbench 2 [33]; symbolic representation of possible structures (5a)–(5d) with percentage probability of reliable assignment according to automated matching algorithms.

Table 1. ANOVA data

Data on analytes (4) and (11) (Table 3)					
Dataset	N	Mean	sd(yEr@)	se(yEr@)	
	10	742.24418	612.13826	193.57511	
	10	742.2469	612.14448	193.57708	
Source	DoF	SS	MS	F-value	P-value
Model	1	3.70046.10 ⁻⁵	3.70046.10 ⁻⁵	9.87535.10 ⁻¹¹	0.99999
Error	18	6744907.00	374717.055		
At the 0.05 level, the population means are not significantly different.					
Levene's Test for Equal Variance					
Model	1	234.718766	234.718766	2.19644.10 ⁻⁹	0.99996
Error	18	1.9235386.10 ¹²	1.068632.10 ¹¹		
At the 0.05 level, the population variations are not significantly different					
Data on analytes (4) and (10) (Table 3)					
Dataset	N	Mean	sd(yEr@)	se(yEr@)	
	10	742.24418	612.13826	193.57511	
	10	742.23243	612.09684	193.56202	
Source	DoF	SS	MS	F-value	P-value
Model	1	6.90606.10 ⁻⁴	6.906061.10 ⁻⁴	1.84315.10 ⁻⁹	0.99997
Error	18	6744382.14	374687.897		
At the 0.05 level, the population means are not significantly different.					
Levene's Test for Equal Variance					
Model	1	10412.3269	10412.3269	9.7465.10 ⁻⁸	0.99975
Error	18	1.9229643.10 ¹²	1.06831.10 ¹¹		
At the 0.05 level, the population variations are not significantly different					
Data on intensities of peaks of analytes (6) and (7) (Table 2)					
Dataset	N	Mean	sd(yEr@)	se(yEr@)	
	24	198560.92921	198998.50833	40620.40041	
	24	147612.30282	151463.44282	30917.3458	
Source	DoF	SS	MS	F-value	P-value
Model	1	3.11491504.10 ¹⁰	3.11491504.10 ¹⁰	0.99611	0.32347
Error	46	1.438456.10 ¹²	3.1270790.10 ¹⁰		
At the 0.01 level, the population means are not significantly different.					
Means Comparison using Bonferroni Test					
Dataset	Mean	Difference between means	Simultaneous confidence intervals		Significant at 0.01 level
			Lower limit	Upper limit	
Data1_(6)	198560.9				
Data1_(7)	147612.3	50948.6264	-86218.06194	188115.31474	No

In (8) it would be too harsh to assign fragment, producing ion at m/z 308 to only [(NeuGc)₂]²⁺ dication on the base on single compound analysis. Moreover, searching data-base [33] shows 94.73 % coverage of proposed structure and MS pattern. Also, there is a lack of literature data on this problem. Gaps can be found for assigning peaks, not only of (NeuGc)₂ and oligomer (n<10) sub-units, but also of (NeuAc)_n (n<10) moieties. **According to [33] dimer and oligomer self-associates of (NeuGc) and (NeuAc) are often found [50–52].** Glycans (1)-(13) show percentage coverage between experimental spectra and structures, having (NeuAc)₂ or (NeuGc)₂ units as follows: (2), 99.29%, (NeuGc)₂ sub-unit; (3), 100.00%, (NeuAc)₂; (4), 100.00%, (NeuGc)₂ or (NeuAc)₂; (5), 91.08% (NeuGc)₂ or 99.02 % (NeuAc)₂; (8), 94.73%, (NeuGc)₂; and (10), 100.00%, (NeuGc)₂ moiety, respectively. **In total, from eleven analytes 54.54% exhibit (NeuGc)₂ or (NeuAc)₂ units with statistical probability 91.08–100%.** Homooligomers of CBs in glycans are broadly acknowledged in the literature. Highlighted is (Neu5Ac)₃ trimer [53]. Thus, we are interested in examining its characteristic MS ions; if any: thus, allowing us its unambiguous assignment, despite, complexity of molecular structure. In doing so, it seems for us both natural and plausible to, further, extend the number of analytes including different core glycans, but also containing oligomer (NeuGc) or (NeuAc) sub-units. Glycans (12) and (13) show coverage between experimental and proposed data on the base on database searching software [33] of: (12), 89.73%, (NeuGc)₄; and (13) 69.77% (NeuGc)₄ and 88.29% (NeuAc)₃, respectively. The (13) annotation of 100 % of MS observable peaks has been achieved looking at structures (13d) and (13e). A variation of 3D structure causes for a decreasing in accuracy. Despite, similarity of MS patterns of (5) and (13) annotation of (13c) to MS variables of former analyte shows accuracy 93.67%.

- [50] L. Cao, J. Diedrich, Y. Ma, N. Wang, M. Pauthner, S. Park, C. Delahunty, J. McLellan, D. Burton, J. Yates, J. Paulson, Nature 13 (2018) 1196. [52] Z. Klammer, P. Hsueh, D. Ayala-Talavera, B. Haab, Mol. Cell. Proteom. 18 (2019) 28–40. [53] K. Jiang, A. Aloor, J. Qu, C. Xiao, Z. Wu, C. Ma, L. Zhang, P. Wang, Anal. Bioanal. Chem. (2022) 550, 1–12. DOI 10.1007/s00216-021-06960-x
- [51] I. Breloy, S. Pacharra, P. Ottis, D. Bonar, A. Grahn, F. Hanisch, J. Biol. Chem. 287 (2012) 18275–18286. DOI 10.1074/jbc.M112.038286

Structure of (5) shows accuracy 99.66 %. **Figure 6** illustrates that (5), (13), and O-glycan (12) exhibit peaks at m/z 291, 395 and 566, respectively. Depending on $(\text{NeuGc})_2$ or $(\text{NeuAc})_2$ units, those peaks can be assigned to ion $[\text{NeuAc}]^+$ [53], $[(\text{NeuAc})_2]^{2+}$ or BZ fragment $[(\text{NeuGc})]^+$ (m/z 291;) $^{2,5}\text{A}_{\text{NeuAc}}$ dication $[(\text{NeuAc})_2]^{2+}$ or dimer $[(\text{NeuGc})_2]^{2+}$ (m/z 395;) and $^{1,5}\text{A}_{\text{NeuGc}}$ fragment $[\text{NeuGc}]^+$, or BZ ion $[(\text{NeuAc})_2]^+$ (m/z 566), respectively. In order to aim significant subtlety in our statement, regarding, assignment of MS ions to 3D molecular fragments of glycans, in **Figure 6** are provided evidences for a linear relation with a statistical significance, when there are correlated absolute intensity data on (5), (12) and (13) of ions at m/z 291, 395 and 566, respectively. **The chemometrics show that, despite, core N- or O-glycan structure those MS peaks correspond to same structural sub-unit.** Chemometrics of peaks of (5) and (13) at $m/z=100\text{--}1800$ show $|r|=0.6862$, thus assuming that MS spectrum of (13) belongs to (13e). The annotation of structure (13e) to MS pattern of (5) shows 60% accuracy. Despite, similarity between spectra of (5) and (13) ratios of intensities of MS peaks are significantly different. The peak at m/z 366 can be assigned to $^{0,4}\text{A}_{\text{NeuAc}}$ of NeuAc-self-associate in (5), as well. Of course, still we have not a solution of the general problem of unambiguous assignment of these MS peaks to $(\text{NeuAc})_2$ or $(\text{NeuGc})_2$ units. Rather, result illustrates a reliability of the general concept of correlation between MS intensity of molecular fragments with adequate parameters reflecting 3D molecular and electronic structure of molecular species producing MS ions.

An evidence, which can provide a warrant for accepting latter statement comes from correlation between D''_{SD} and D_{QC} data on fragments and their characteristic MS ions, thus contributing crucially to the room of current debates about accurate and proper assignment of MS fragments of glycans to their molecular and electronic structures. In order to increase in reliability of our statements, there are examined D_{QC} data on a series of possible 3D structural fragments *per* any observable MS peak within the framework of a set of possible glycan structures exhibiting accuracy of annotated MS peaks $> 90\%$. For instance, those are structures (5a)–(5d) and (13a)–(13d). We can remarked that (8) does not exhibit peaks at m/z 291, 395 and 566, despite that, according to [33] there is probability of 94.73 % of unit $(\text{NeuGc})_2$. The same is true for peak at m/z 308 of $[(\text{Gal})(\text{Fuc})]^+$ in (8), but is not observed in (9), despite, the fact that coverage between MS peaks and molecular structure according to [32] shows 100 %. In order to tackle those variations among database searching annotations [33] and observable MS phenomena, assignment of (9) is carried out manually, thus agreeing well with data on [54]; showing that at $\text{RT}=65$ mins there are 2-AB labelled glycans, having Fuc-residue. Of course, we should underline that (9) reveals peak at m/z 407 of $[(\text{Gal})(\text{GalNAc})]^+$ cation, as well. It can be associated with $[(\text{HexNAc})_2]^+$ [33]. Ion at m/z 407 is associated with oxonium-ion in O-linked N,N'-diacetyllactosamine (LacdiNAc)-modified glycans [55].

[54] F. Higel, U. Demelbauer, A. Seidl, W. Friess, F. Soergel, *Anal Bioanal Chem* 405 (2013) 2481–2493.

[55] I. Breloy, S. Pacharra, P. Ottis, D. Bonar, A. Grahn, F. Hanisch, *J. Biol. Chem.* 287 (2012) 18275–18286..

Further: How chemometrics evaluate relation $D''_{SD} = f(D_{OC})$, thus responding to reliable assignment of MS spectra, exhibiting virtually similar MS patterns? Those are trications of **(4)**, **(10)**, and **(11)**, respectively. They exhibit peak at m/z 657 of $[(\text{Neu5Ac})(\text{Gal})(\text{GalNAc})]^+$ cation. Analyte **(10)** also shows peak at m/z 695 of adduct $[(\text{Neu5Gc})(\text{Gal})(\text{GalNAc})+\text{Na}]^+$. There are peaks at $m/z=1180-1200$ of ions $[(\text{Neu5Ac})(\text{Gal})(\text{GalNAc})(\text{Man})_2(\text{GalNAc})]^+$ and $[(\text{Neu5Gc})(\text{Gal})(\text{GalNAc})(\text{Man})_2(\text{GalNAc})]^+$, as well. Analyte **(11)** shows three different structure producing coverage of 100% between MS pattern and molecular structure according to [33]. A statistically representative set of MS data on ten ions of those analytes are virtually identical from perspective of chemometrics (**Figure 10**). The correlation between measurable variables of **(4)**, **(10)** and **(11)** shows $|r|=1-0.9278$. ANOVA tests indicate that those datasets are not statistically significantly different. In other words, like in case of MS ions at m/z 291, 395 and 566 of **(5)**, **(12)** and **(13)**, there is mutually by pairs of compounds virtually identical both m/z and MS intensity quantities. Analogously, as in **(5)** and **(13)**, there are examined a set of possible assignments of MS ions of **(4)**, depending on three structures showing accuracy of 100 % of annotated peaks according to database-searching algorithm [33]. Owing to the fact that, temporal behaviour of MS intensity obeys a certain law and this law is equation (2) [27–31], as well as, there is linear correlation between D''_{SD} and D_{OC} parameters of ionic species [27–31], results lastly discussed assumes that there are virtually identical fragment ions. Looking at $sd(yEr\pm)$ values, it becomes clear that virtual identity of measurable variables is particularly highlighted in MS ions at m/z 138, 168, 204 and 274 showing $sd(yEr\pm)$ at five digit. Thus, we see the problem why the data-base matching algorithm [33] produces 100% reliability of structural identification of three different structures of **(10)** toward its MS measurable data. Obviously enough, there is needed employment in different method determining relations between measurable variables of glycans and not only 2D, but also 3D molecular and electronic structures of those analytes. The presented assignment is carried out according to [56,57]. Our innovative equation (2) and its complementary use to Arrhenius's equation (3) are possible way around this problem. Formulas (2) and (3) are actual physicochemical laws.

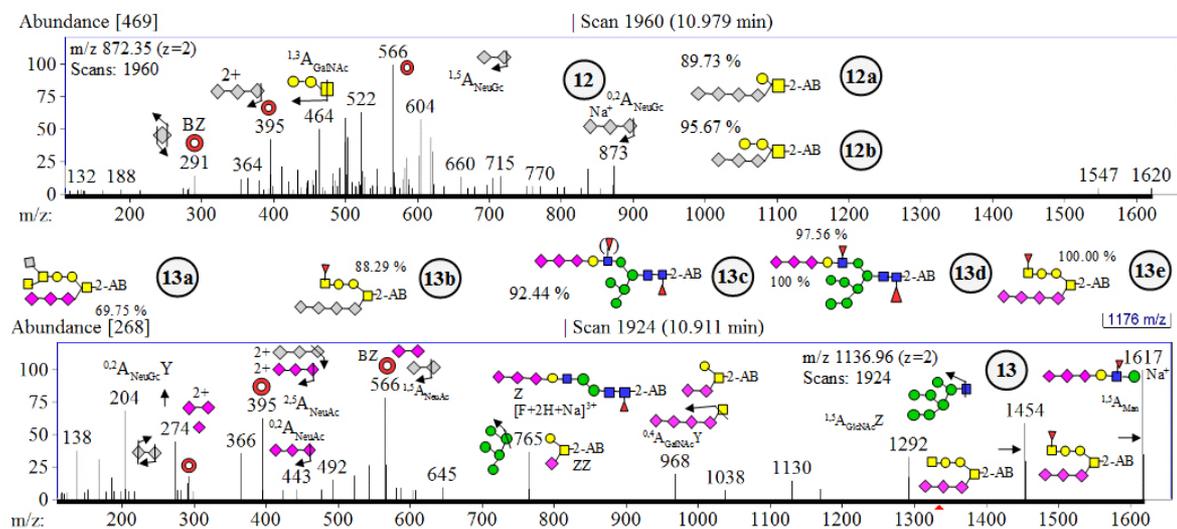


Figure 6. Mass spectra (screenshot) of ions of **(12)** and **(13)**; annotation of peaks according to GlycoWorkbench 2 [33]; symbolic representation of structures (12a), (12b), and (13a)-(13e) with percentage probability of assignment according to automated matching algorithms.

[56] C. Calvano, T. Cataldi, J. Koege, A. Monopoli, F. Palmisano, J. Sundermeyer, *J. Am. Soc. Mass Spectrom.* 28 (2017) 1666–1675.

[57] L. Hammad, D. Derryberry, Y. Jmeian, Y. Mechref, *Rapid Commun. Mass Spectrom.* 24 (2010) 1565–1574.

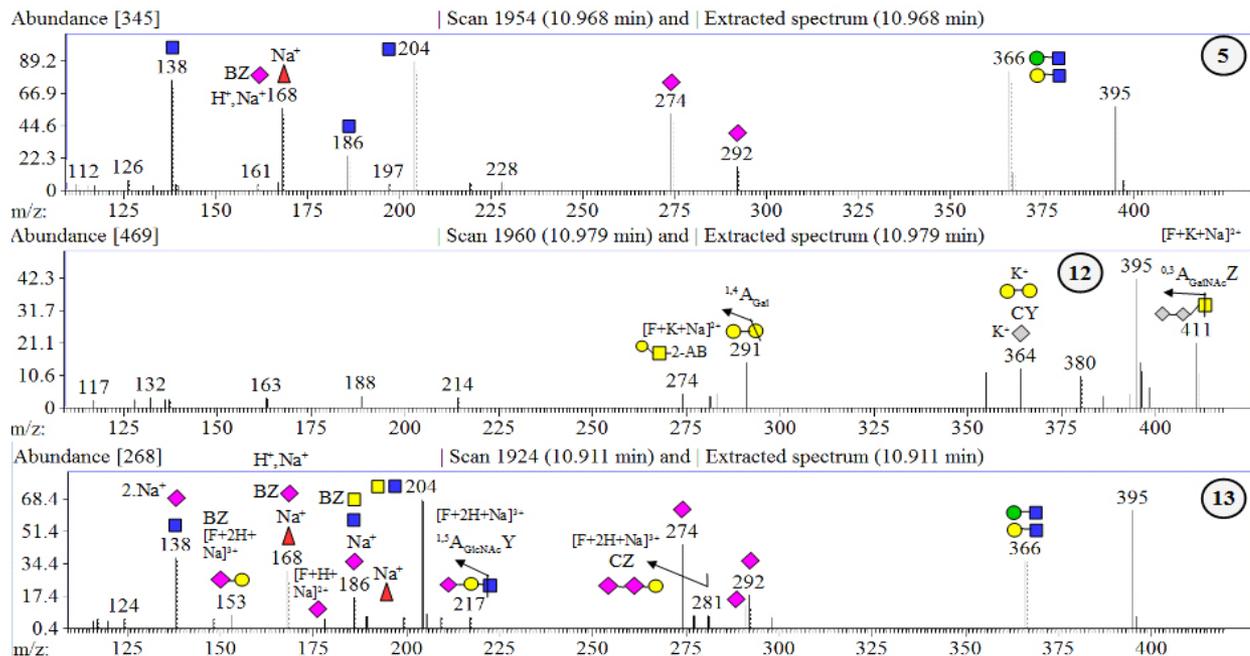


Figure 6. Mass spectra of ions of **(5)**, **(12)** and **(13)** within m/z 100-500; annotation of peaks according to GlycoWorkbench 2 [33] and literature research.

Also, different core molecular structure in assigning common MS peaks is a straightforward example of plausible and statistically representative assignment of species of so complex objects, as glycans. This is not the case, however, looking at MS peak at m/z 274. It is often assigned to $[Neu5Ac]^+$ ion [53], exhibiting a loss of solvent water molecule. However, examining of chemical reactivity of (Neu5Ac)-containing glycans show that there is a lactone formation of Neu5Ac-CBs [2], which should be observed between (Neu5Ac)-(Gal) residues. Thus, there is an additional question, regarding, mechanistic aspects of loss of solvent water molecule from Neu5Ac-fragment. Moreover, there are common peaks at m/z 657 and 366 of $[(GalNAc)-(Gal)-(Neu5Ac)]^+$ and $[(GalNAc)-(Gal)]^+$ cations [52]. Despite, the later assignment to peak at m/z 366, there can be annotated cation of $[(GalNAc)-(Man)]^+$, as well. For this reason, the study correlates between D''_{SD} and D_{OC} data on two $[(GalNAc)-(Gal)]^+$ and $[(GalNAc)-(Man)]^+$ 3D molecular and electronic structures, in addition to, type of (Gal)-(Neu5Ac) chemical bonding ($\alpha_{2,6}$ and $\alpha_{2,3}$). Moreover, peaks at m/z 274 and 366 can be found in spectra of O-glycans, where the former peak can be associated with dimer $[(NeuGc)_2]^+$. Also, peak at m/z 657 is often assigned to chemically bonded CB sequence $[(Neu5Ac)-(Gal)-(GalNAc)]^+$ [52] of O-glycans [58]. It is important to mention that hypothesised O-glycan (12) is also known from fetuin [53].

[58] M. Windwarder, F. Altmann, J. Proteom 108 (2014) 258–268.

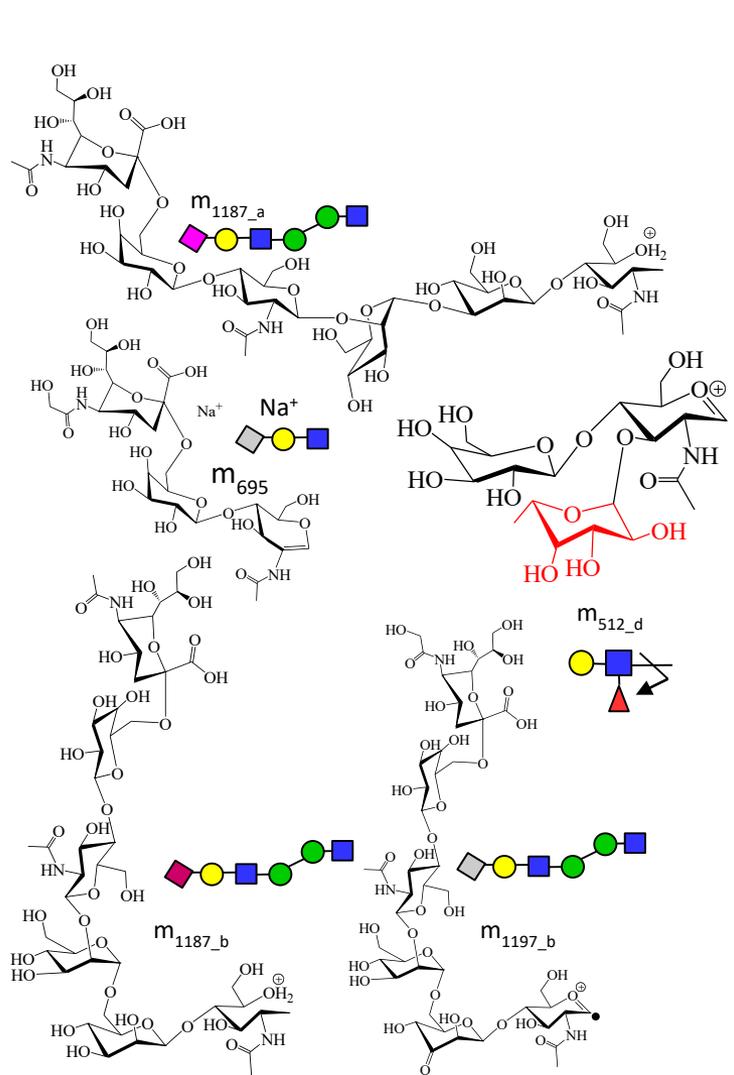


Figure 8. A selected set of fragment ions examined, herein: chemical and symbolic diagrams of the carbohydrate units.

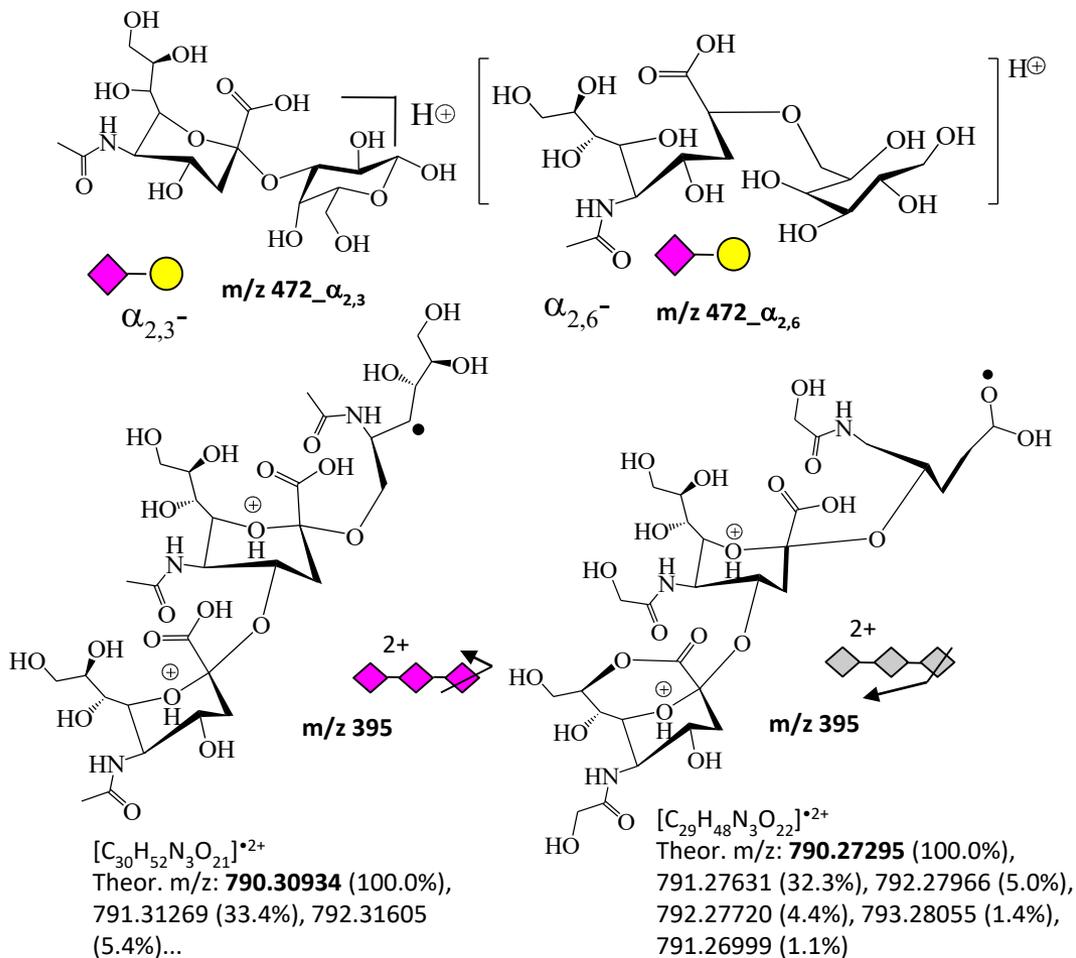


Figure 9. A selected set of fragment ions examined, herein: chemical and symbolic diagrams of carbohydrate units; theoretical m/z data on species and ratios of peaks of isotope shape given as percentage.

ECMC
2022

The 8th International Electronic
Conference on Medicinal Chemistry
01-30 NOVEMBER 2022 | ONLINE

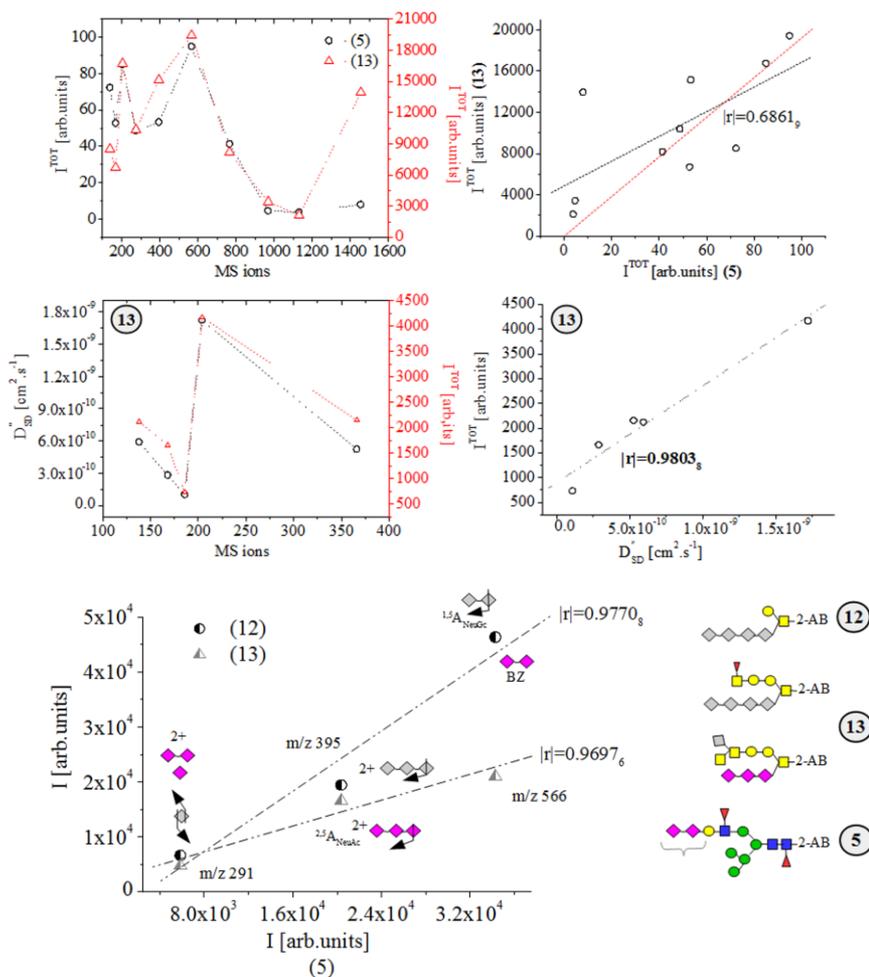


Figure 10. Relations among total mass spectrometric intensity (I^{TOT} [arb.units]) data and stochastic dynamic MS parameters according equation (2) of analytes (5), (12) and (13); symbolic presentation of possible carbohydrate structures according to [11]; chemometrics.

Determination of stochastic dynamic diffusion parameters of glycans

Tables S3 and S4 show intensity data on ions *per* short spans of scan time. Table S5 shows parameters of equations (1) and (2). (Calculation tasks are detailed on [27–31]; Figure S9) $\ln P_1$ parameter is a constant ($\ln P_1 = 17.053$), despite, analyte, ionization method, and experimental conditions. Table S5 provides new pro-argument for the latter statement. The temporal behaviour of measurable variable *intensity* of peaks of analyte ions obeys, again, same physicochemical law or equation (1). Since, the sub-section also addresses a questions asking whether experimental variable *intensity* obeys equation (2) as a certain law, as well as, Figure S10 correlates between D_{SD}^+ and D_{SD}^i data on formulas (1) and (2). The values $|r|=0.9976_4$ and 0.8818 (8) indicate compatibility of those equations. The variability of $|r|$ -coefficient from $|r|=1$ is explained with error contribution to D_{SD}^i values, due to SineSqr fitting data-processing of relation $(I-\langle I \rangle)^2=f(t)$. As has been touched on this issues still in introductory section, the development of simplistic equation (2) from equation (1) as approach used to obtain D^i excludes from error contribution to D_{SD}^i parameters, due to curve-fitting approach with SineSqr function. Despite, the fact that there is deviation from exact $|r|=1$ of relation $D_{\text{SD}}^+ = f(D_{\text{SD}}^i)$ the two equations, expressing functionality between diffusions and measurable MS variable *intensity* of ion remains true — there is $\ln P_1=17.053$ of MS spectra of glycans, as well. (Within the framework of the Box-Mueller method used to write equation (1), the parameter P_1 represses a random number. See equation (A2.2) (p.291) [59].) Absolute data on $|r|=1$, examining function $D_{\text{SD}}^+ = f(D_{\text{SD}}^i)$ is achieved looking at ions at m/z 291, 292, 395 and 566 of (5), (12) and (13), respectively. Therefore, for those cases, there is absolute compatibility between D_{SD}^+ and D_{SD}^i parameters of equations (1) and (2). The same relation of ions of (13) at m/z 138, 153, 168, 186 204, 274 281, and 366 yields to $|r|=0.9921_8$.

Function between D_{SD}'' data on glycans by pairs shows $|r|=0.9999_8$ ((5)/(13)) and $|r|=0.5145_4$ ((12)/(13)). The decreasing in statistical significance in (12) and (13) agrees well with proposed annotation of MS ions. If peaks at m/z 291/292, 395 and 566 belong to monomer and dimer of (NeuAc)-unit(s), then, this explains the excellent value $|r|=0.9999_8$ of (5) and (13), having (NeuAc)₂-unit annotated at a reliability of 99.02 %. The (12) lacks of (NeuAc)₂-fragment, but there is oligomer (NeuGc)_n. The obtained $|r|=0.5145_4$ assumes presence of different (NeuAc)₂ and (NeuGc)_n self-associates of (5) and (12). Relation $D_{SD}''=f(D_{SD}')$ of (10) yields to $|r|=0.9312_1-0.9984_8$, while (11) shows $|r|=0.9983_7$. D_{SD}'' data on (6) and (7) yield to $|r|=0.9852$.

Relation $D_{SD}''=f(I^{TOT})$ of a set of ions of (8) shows $|r|=0.97$. Owing to the fact that D_{SD}'' parameter is determined precisely *per* short span of scan time ($t=0.18$ s; (8)), it becomes clear that employment in I^{TOT} values of both quantitative and 3D structural analyses causes for deviation from $|r|=1$, as well. Moreover, chemometrics of $D_{SD}''=f(I^{TOT})$ function of (12) and (13) show $|r|=0.1841_8$ and 0.7000_8 . Data on (10) produce $|r|=0.6512_4-0.9833_2$. High $|r|$ -parameter is obtained, excluding from ion at m/z 274. Results from (11) show $|r|=0.9921$, examining ions at m/z 138, 168, 204 and 274, respectively.

Theoretical data

Relations among molecular and electronic structures as well as energetics of characteristic fragmentation ions of glycans

Figures 11–13 depict relation among **molecular structure** \leftrightarrow **electronic structure** \leftrightarrow **energetics**. It determines most stable 3D conformation at ground (GS) and transition (TS) states (Table S6.) Frequently, charged CBs show intramolecular proton and charge transfer effects, as well as, rearrangement. Series of mono- and dications of monomers and dimers of (NeuAc)_n and (NeuGc)_n ($n=1-2$) show intramolecular cyclization. The discussed in the literature oxonium cation of peak at m/z 292 of [(NeuAc)]⁺ has covalently bonded O-centre from C⁷-OH group and C²-(COOH) group oxonium ion. The CT-effect results in stabilization of C²⁺-cation. Same is true for ion m_{292_b} , exhibiting thermodynamically favourable cyclic product ($\Delta E^{TOT}=|0.006201|$ a.u.) A cyclization reaction has been found looking at most stable 3D ionic structure of m_{274_b} cation (Figure 12(A)), as well. Again, initial state of oxonium cation appears less preferred. There is a covalent bonding of O-centre of C⁷-OH position accompanied with a CT effect. Cation m_{274_a} is less stable than ionic species m_{274_b} at $\Delta E=|0.036|$ a.u. Conversely, stable oxonium cation is found studying CBs ions belonging to MS peak at m/z 366 (Figure 11.) Its stabilization also depends on type of CBs monomer and its isomers; if any. The latter discrepancy between literature data and results from our study regarding thermodynamic stability of oxonium cations of CBs is associated with (i) a rather lack of a large systematic data on geometry and thermochemistry of MS species of CBs: and (ii) their complex fragmentation chemistry, as well as, (iii) many possible ion-molecule dissociation reactions. There is a lack of systematic studies devoted to theoretical chemical analysis of fragmentation reactions of even small CBs. Mass spectrometric fragmentation chemistry of glycans requires systematic effort, in order to be comprehensively understood. Assignment of MS peaks of glycans to corresponding fragment species represents a nontrivial research task.

[59] A. Satoh, Introduction to practice of molecular simulation (2011) Elsevier, Amsterdam, pp. 1–322.

ECMC
2022

The 8th International Electronic
Conference on Medicinal Chemistry
01–30 NOVEMBER 2022 | ONLINE

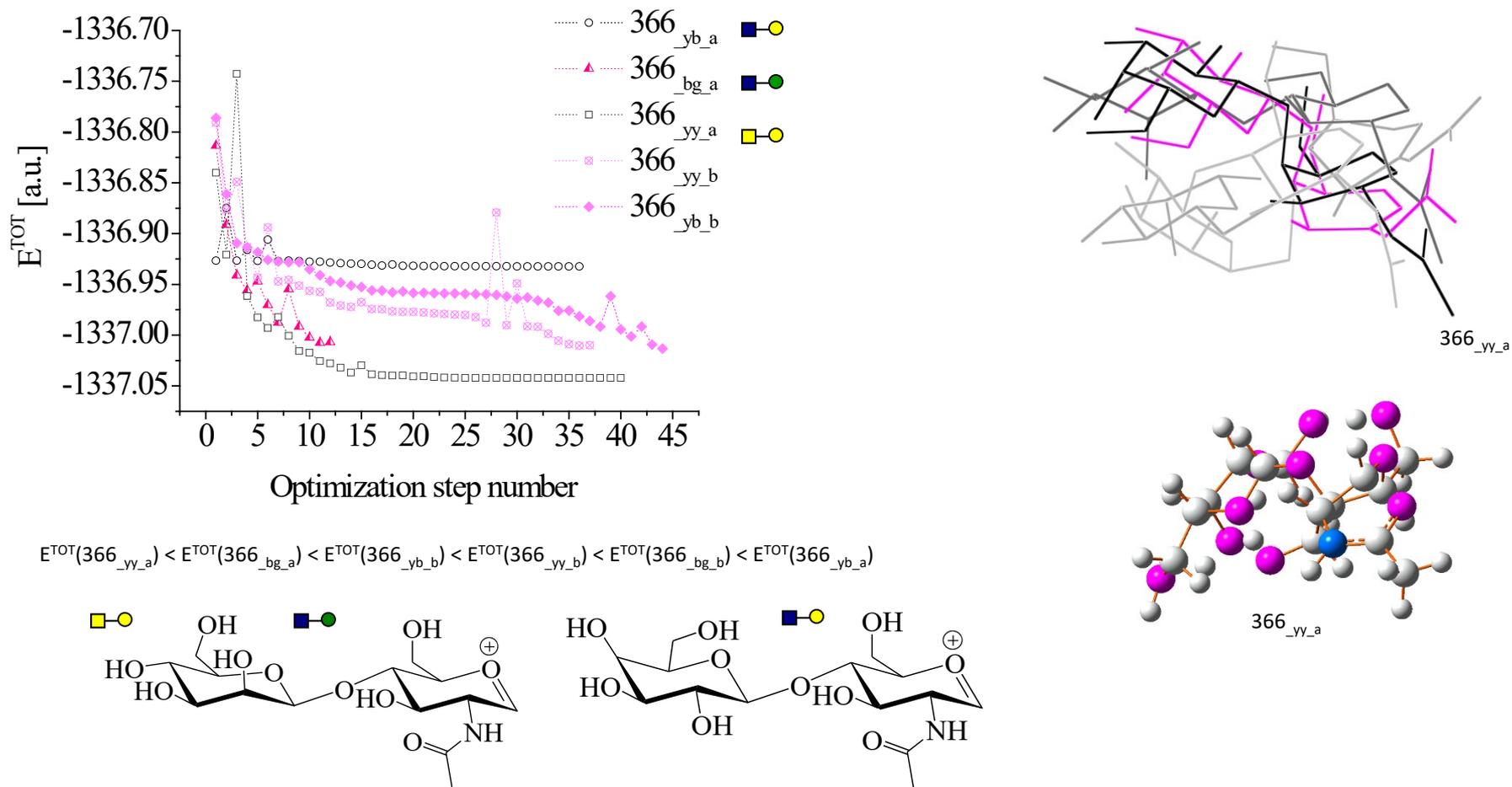


Figure 11. DFT (M062X) ionic optimization of different cations of MS peak at m/z 366: Total energy (E^{TOT} [kcal.mol⁻¹]) versus optimization step number; symbolic and chemical diagrams of carbohydrates; 3D molecular conformations.

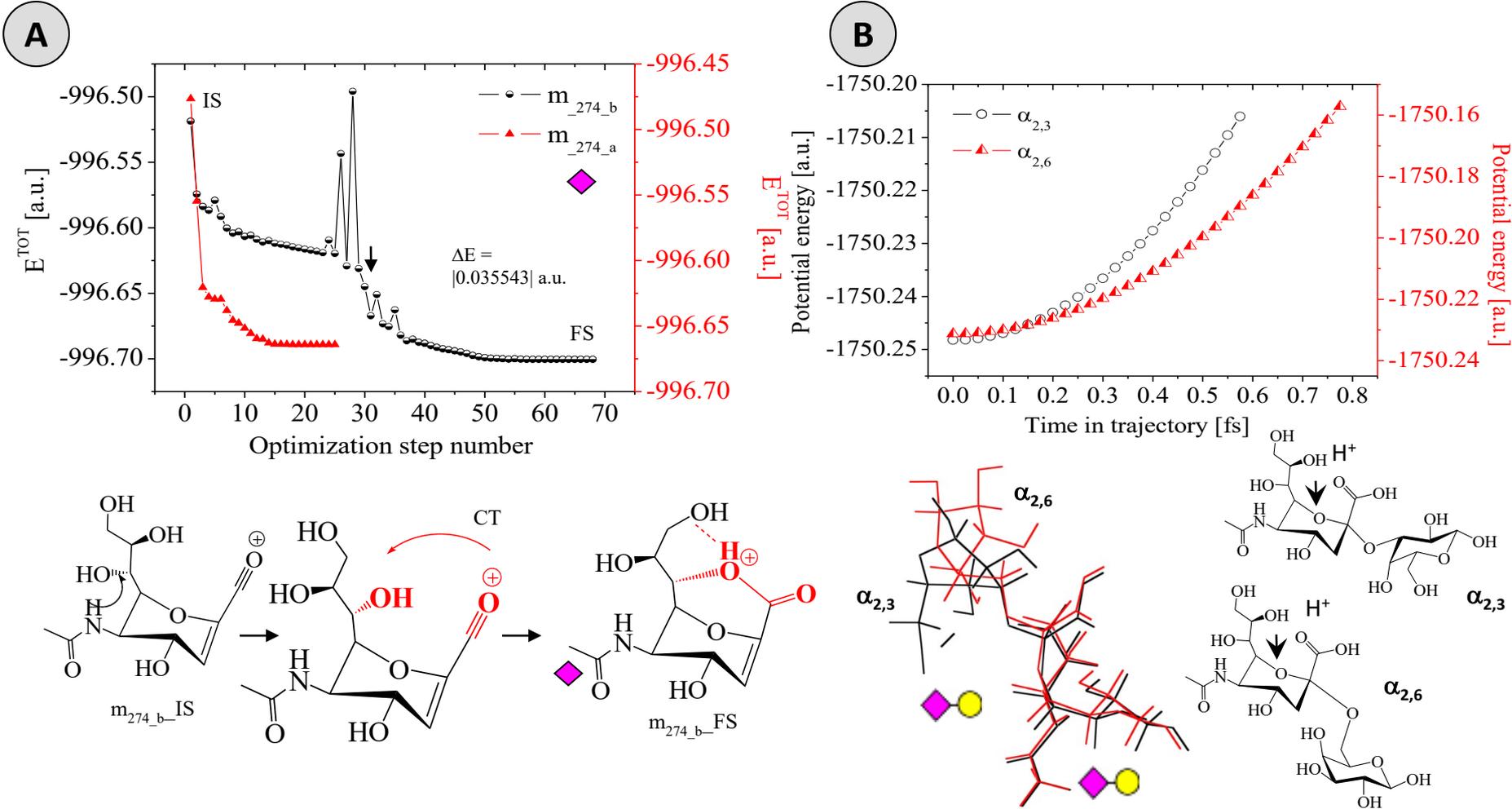


Figure 12. Density functional theory (DFT) ionic optimization at M062X level of theory of different cations of MS peak at m/z 247: Total energy (E^{TOT} [kcal.mol⁻¹]) versus optimization step number; difference in energy (ΔE); chemical diagrams of ion 274_b i ist initial (IS) and final (FS) steps; charge transfer (CT) reaction (A); MD data on ion at m/z 472 of [(Gal)(NeuAc)⁺]: Potential energy [a.u.] versus time in trajectory; chemical and symbolic diagrams of ions; 3D molecular structure of dications (B).

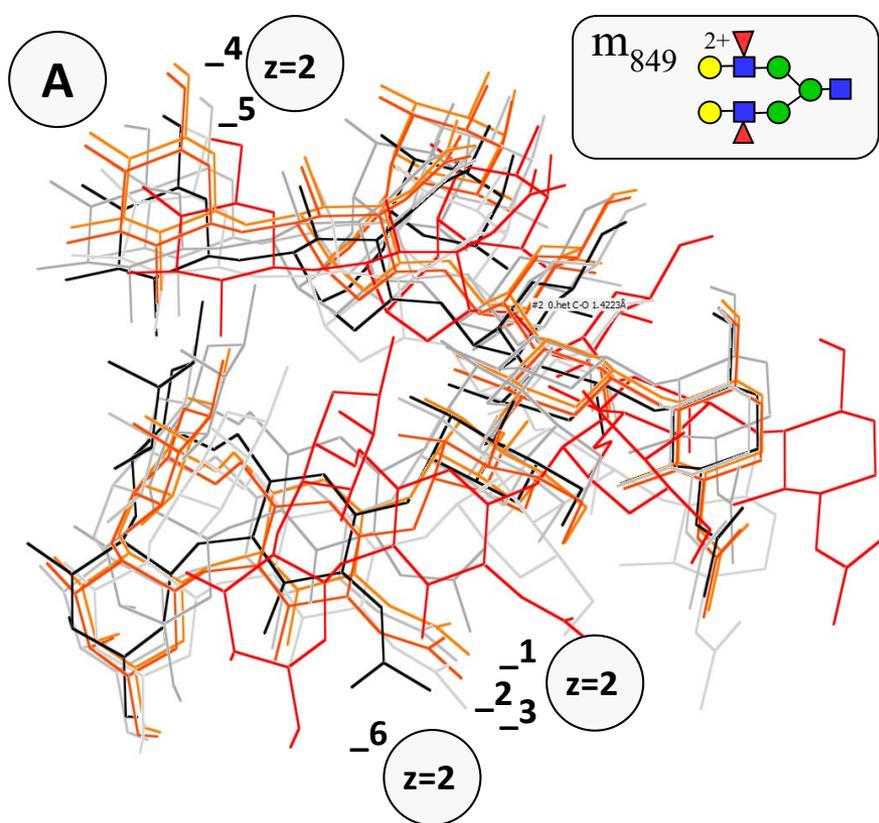
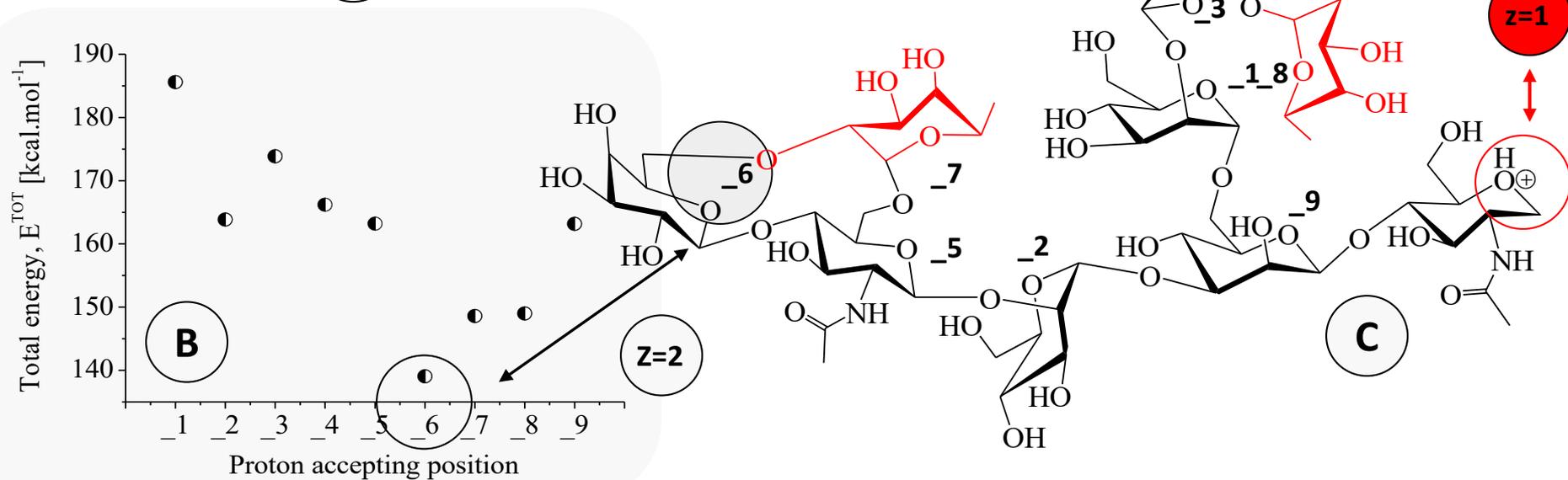


Figure 13. Molecular mechanics/molecular dynamics (MM/MD) data on dication ($z=2$) of MS peak at m/z 849 depending on proton position of second proton: 3D molecular conformations of charged species (A); total energy (E^{TOT} [kcal.mol $^{-1}$]) with respect to proton position of second proton of the dication (B); chemical diagram of ion at m/z 849 (C); symbolic representation of CB-units of dication.



Determination of quantum chemical diffusion parameters of glycans according to Arrhenius's theory

D_{QC} -data on formula (3) are obtained *via* high accuracy static and MD data employing in adiabatic Born-Oppenheimer MD, Atom-Centered Density Matrix Propagation

(ADMP) MD, DFT, and diabatic MD, respectively. Those methods provide comparable accuracy of 3D molecular and electronic structures, thermochemistry, diffusion, and binding energy with so-called quantum chemical *gold standard* or CCSD(T) approach to CBS limit [60]. M06-2X 6-311++G(2d,p) method shows comparable energetics of organics and cation-radicals with CCSD(T)/aug-ccpVTZ and CCSD(T)-CBS approaches ($\Delta E = |2| - |1|$ kJ.mol⁻¹) [61]. The 3D structural analysis is carried out by chemometrics, evaluating correlation between D''_{SD} - and D_{QC} -data on equations (2) and (3) ($D''_{SD} = f(D_{QC})$). The $v_i^{(s)}$ term, mentioned above, reflects change in vibrational entropy when charged ionic molecular system moves from GS to TS, written in harmonic approximation. Change of vibrational entropy reflects entropy contribution to *free Gibbs energy* of activation process. The imaginary frequency of saddle point is excluded from calculations of D_{QC} -parameters [62]. **Tables S6-S8** summarize atomic coordinates used to compute vibrations in GS and TS, as well as, vibrational modes used to calculate D_{QC} -data on **Table 2**.

m/z	Form	D_{QC}	m/z	Form	D_{QC}
138	138_b 	135.5037 ₇₉	168	168_b 	56.6548
				168_c 	137.8581
186	186_a 	110.8916			
	186_FucNa ⁺  Na ⁺	141.7072			
291	291_a 	184.8312 ₅	204	204_a 	590.6082
	291_c 	44.9376		204_b 	53.6261 ₀₄
292	292_a 	32.6495 ₉	274	274_a 	120.3697 ₅
	292_b 	78.2906 ₈		274_b 	174.4939 ₉
366	366_bg_a  	18.8329 ₈	472	472_α _{2,6}  	18.2031
	366_bg_b  	160158.2365		472_α _{2,3}  	12.6738 ₉₄
	366_yb_a  	59.6433 ₈			
	366_yb_b  	117.5081			
	366_yy_a  	73.8457			

Table 2. Theoretical D_{QC} data on equation (3) accounting for various forms of fragmentation species annotated to mass spectrometric observable peaks of glycans

[60] B. Boychuk, Y. Jeong, S. Wetmore, J. Chem. Theory Comput. 17 (2021) 5392–5408.

[61] Y. Liu, A. Dang, J. Urban, F. Turecek, Angew. Chem. Int. Ed. 59 (2020) 7772–7777.

[62] A. Nitzan, Chemical dynamics in condensed phases: Relaxation, transfer and reactions in condensed molecular systems, Oxford University Press, Oxford, (2006) pp. 1–712.

Correlative analysis between mass spectrometric stochastic dynamic and quantum chemical diffusion data on glycans

Figure 14 depicts correlation between MS and D_{QC} data on equations (2) and (3) of CBs ions including those depicted in **Figures 8** and **9**. There are evaluated data on **Tables 2** and **S3 - S8**. Excellent chemometrics show $|r|=0.9892-0.9977_7$. The figure use D_{QC} data on species of ions at m/z 366 and 472 depending on CB-units and isomers. The $\alpha_{2,3}$ isomer of ion at m/z 472 is more stable comparing with $\alpha_{2,6}$ one, but there is increasing in $|r|=0.9978$ analyzing MS ions of analyte (**9**). The result provides a forceful argument for assigning MS peak at m/z 472 to cation $m_{472_{\alpha_{2,3}}}$ of (**9**). Same logic scheme is used to assign peak at m/z 366 to cation $[(\text{Man})(\text{NeuAc})]^+$ instead of, to $[(\text{Gal})(\text{NeuAc})]^+$ or $[(\text{Gal})(\text{GalAc})]^+$ species, despite, the fact that oxonium cation of $[(\text{Gal})(\text{GalAc})]^+$ is most stable among ions belonging to MS peak at m/z 366 (**Figure 11**.) At this point, we can underline that one crucial advantage of our general line of complementary employment in equations (2) and (3) *via* logic scheme shown in **Figure 4** is that it drastically increases MS capability of determining 3D structurally enantiomers, despite, their exactly same energies and physico-chemical properties, excluding from their optical rotation [63]. Owing to the fact, that MS intensity variables of ions reflect physicochemical properties, it is obvious that MS shape of enantiomers are virtually identical, however, excluding from fluctuations of observable parameters. Exactly same energetics of such species unable us to employ only GS *free Gibbs energy* parameters, in order to, distinguish between enantiomers. Since, formula (3) uses molecular vibrations of TSs, as well as, then, it becomes possible to distinguish accurately among enantiomers determining their D_{QC} -values. However, in context exact predictive power of first-order saddle point vibrational properties (**Table S8**,) there should be highlighted that a direct comparison between experiment and theoretical MD results is often difficult, due to inaccuracy of so-called *interatomic potentials*. Thus, there are examined both anharmonic and harmonic parts of potential energy surfaces (PES) [64]. Computation of PES of TSs (**Figure 12**) enable us to detail accurately on thermal motions and their fluctuations within the framework of MD analysis, thus distinguishing between subtle electronic effects of enantiomers in their TSs. Highly accuracy data on PES can be obtained by means of *ab initio* and DFT methods [65]. Owing to the fact that theoretical examining of CBs, generally, represents a complex research task, due to reasons mentioned in preceding (sub-)sections, obtained, herein, excellent $|r|$ -data on relation $D''_{SD}=f(D_{QC})$ within the harmonic approximation and DFT method only highlight a great prospect of equation (2) for a broad interdisciplinary studies devoted to very complex molecular objects such as, for example, glycans.

[63] R. Kaur, Vikasa, J. Chem. Phys. 142 (2015) 074307.

[64] A. Rohskopf, Sp. Wyant, K. Gordiz, H. Seyf, M. Muraleedharand, A. Henry, Comput. Mater. Sci. 184 (2020) 109884.

[65] Z. Morrow, H. Kwon, C. Kelley, E. Jakubikova, J. Chem. Theory Comput. 17 (2021) 5673–5683.

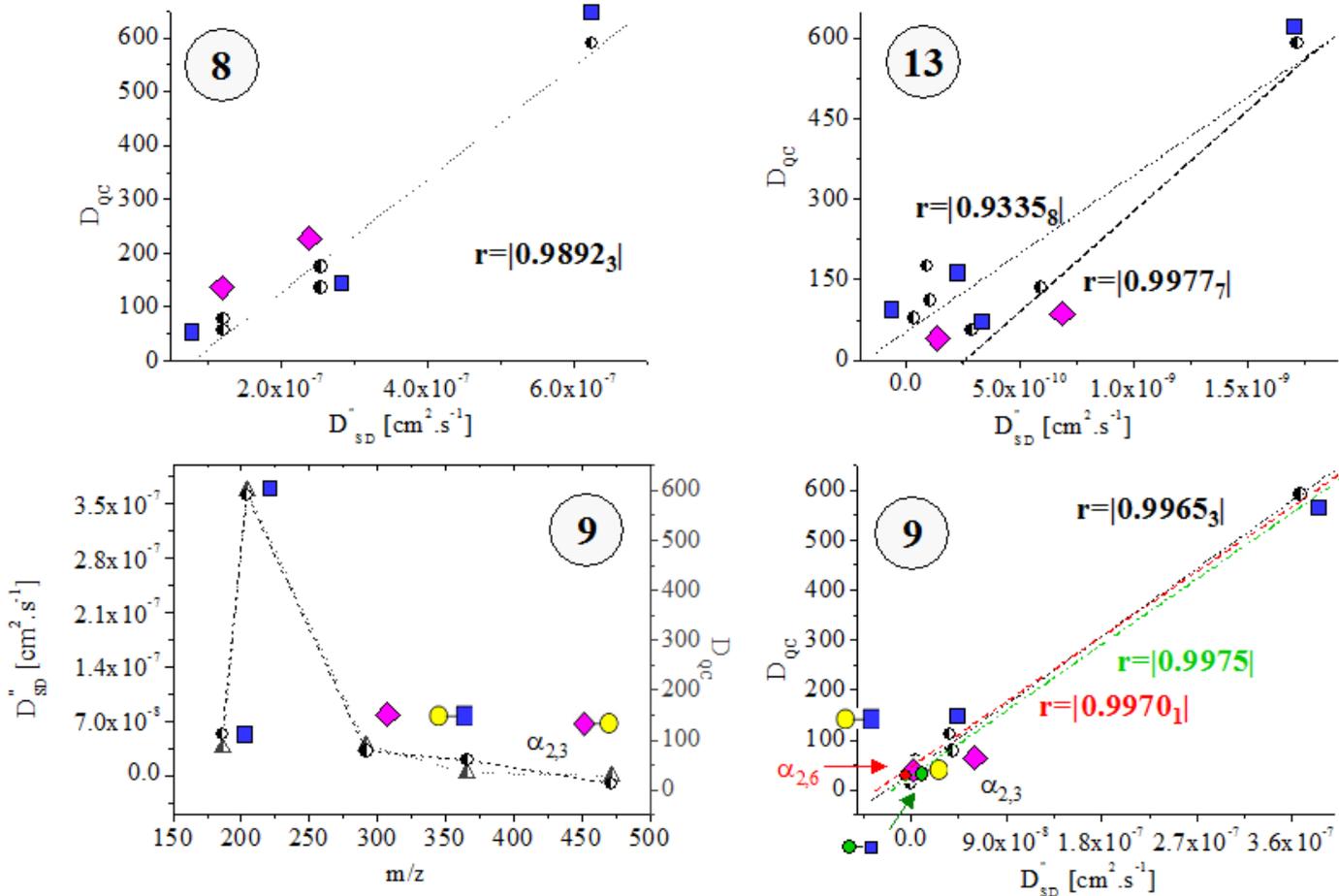


Figure 14. Correlation between D''_{SD} and D_{QC} data on formulas (2) and (3) of glycans; symbols of carbohydrate units; chemometrics.

SUPPORTING INFORMATION

Supporting information data on the study can be found free of charge [66].

[66] [<https://doi.org/10.5281/zenodo.6877054>].

ECMC
2022

The 8th International Electronic
Conference on Medicinal Chemistry
01-30 NOVEMBER 2022 | ONLINE

Conclusions

We began this presentation with a proposition that the problem of 3D structural determination of *glycans* in mixtures can be solved unambiguously from perspective of *chemometrics* employing *soft ionization mass spectrometry* and relationship between equations (2) and (3). In this scheme we have exploited our innovative formula (2), quantifying MS measurable variable *intensity* of analyte MS peaks [26-31]. Our proposition seems to be confirmed by this empirical research, showing that function $D''_{SD}=f(D_{QC})$ of *thirteen glycans of fetal bovine serum* yields to excellent correlation coefficients between theory and experiment ($|r|=0.9892-0.9975_{1.}$) Owing to very serious difficulty, facing 3D structural analysis of glycans, due to random variation of non-template-driven processes of their glycosylation and fucosylation, in addition to, a lack of regioselective derivatization leading to mixtures of polydisperse analytes toward length and skeletal modifications; isomers of oligomers and polymers; including linear and branches molecular structures; as well as, great conformational flexibility; intramolecular proton and charge transfer effects; and rearrangement, and cyclization reactions, the obtained $|r|$ -quantities highlight our proposal as prospective new approach to determine glycans 3D molecular structurally with broad interdisciplinary application.

Acknowledgments

DAAD

Deutscher Akademischer Austausch Dienst
German Academic Exchange Service



Stabilitätspakt für Südosteuropa
Gefördert durch Deutschland
Stability Pact for South Eastern Europe
Sponsored by Germany

DFG

Deutsche
Forschungsgemeinschaft



**ECMC
2022**

**The 8th International Electronic
Conference on Medicinal Chemistry**

01–30 NOVEMBER 2022 | ONLINE