International Journal of Mass Spectrometry 461 (2021) 116495

Contents lists available at ScienceDirect

International Journal of Mass Spectrometry

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

A general new method for calculating the molecular nonpolar surface for analysis of LC-MS data

Rabin Dhakal [a], Reed Nieman [b, **, 1], Daniel C.A. Valente [c], Thiago M. Cardozo [c], Bhumika Jayee [b], Amna Aqdas [b], Wenjing Peng [b], Adelia J.A. Aquino [a, ***], Yehia Mechref [b], Hans Lischka [b, *], Hanna Moussa [a]

[a] Department of Mechanical Engineering, Texas Tech University, Lubbock, TX, 79409, USA
[b] Department of Chemistry and Biochemistry, Texas Tech University, Lubbock, TX, 79409-1061, USA
[c] Instituto de Química, Universidade Federal Do Rio de Janeiro, Rio de Janeiro – RJ, 21941-901, Brazil

ARTICLE INFO

Article history:
Received 21 September 2020
Received in revised form
4 December 2020
Accepted 7 December 2020
Available online 16 December 2020

Keywords:
Nonpolar surface area
Solvent accessible surface
Continuum solvation
Analysis of LC-MS data

ABSTRACT

The accurate determination of the nonpolar surface area of glycans is vital when utilizing liquid chromatography/mass spectrometry (LC-MS) for structural characterization. A new approach for defining and computing nonpolar surface areas based on continuum solvation models (CS-NPSA) is presented. It is based on the classification of individual surface elements representing the solvent accessible surface used for the description of the polarized charge density elements in the CS models. Each element can be classified as polar or nonpolar according to a threshold value. The summation of the nonpolar elements then results in the NPSA leading to a very fine resolution of this surface. The further advantage of the CS-NPSA approach is the straightforward connection to standard quantum chemical methods and program packages. The method has been analyzed in terms of the contributions of different atoms to the NPSA. The analysis showed that not only atoms normally classified as nonpolar contributed to the NPSA, but at least partially also atoms next to polar atoms or N atoms. By virtue of the construction of the solvent accessible surface, atoms in the inner regions of a molecule can be automatically identified as not contributing to the NPSA. The method has been applied to a variety of examples such as the phenylbutanehydrazide series, model dextrans consisting of glucose units and biantennary glycans. Linear correlation of the CS-NPSA values with retention times obtained from liquid chromatographic separations measurements in the mentioned cases give excellent results and promise for more extended applications on a larger variety of compounds.

1. Introduction

Glycosylation is a post-transcriptional modification of proteins and is critical in many biological processes such as intracellular processes, cell signaling, and protein localization, folding, and transport[1,2]. The diverse functionality found in glycans results from the vastly complex stereochemical information exhibited by the structures. This can be attributed to several types of structural characteristics such as a) composition of monosaccharide residues, b) residue linkage sites with potential α- or β-stereochemistry, c) different potential branching features, and d) additional functional group modification of glycans with e.g., methyl, phosphate, or sulfate groups. As aberrant glycosylation is well known to be linked to certain diseases[3–5], cancers[6], inflammation[7], and immune disorders[8], this has led to increased attention for quantitative glycomics and glycoproteomics studies requiring reliable methods of identification and quantification of glycans and glycoproteins [9,10]. Analysis of glycans and glycoproteins is challenging because of the complications associated with the high complexity and microheterogeneity resulting primarily from multiple isomers. Efforts to characterize and separate isomers, therefore, provides essential insights into glycosylation processes[11–13].

* Corresponding author.
** Corresponding author.
*** Corresponding author.
E-mail addresses: reed.nieman@ttu.edu (R. Nieman), adelia.aquino@univie.ac.at (A.J.A. Aquino), hans.lischka@univie.ac.at (H. Lischka).
[1] Current address: Department of Chemistry and Chemical Biology, University of New Mexico, Albuquerque, NM, 87313–0001, USA.

https://doi.org/10.1016/j.ijms.2020.116495
1387-3806/

The biological importance and significance of proper characterization of glycan isomers has been previously investigated in several specific cases[14,15]. It was found that there is a general trend in blood serum samples of state IV breast cancer patients of increased amount of acidic glycans with α-2,6 linked sialic acid residues highlighting the role N-linked glycans can play in diagnostic tests[14]. In another study, it was found that when the blood serum samples from cancer patients treated with a chimeric mouse−human IgG1 monoclonal antibody were screened for Immunoglobulin E (IgE) antibodies that the majority of subjects with hypersensitive reactions had the specific preexisting IgE antibodies for galactose-α-1,3-α-galactose[15].

Mass spectrometry (MS) is an ideal method for providing accurate and comprehensive glycan structural information which is highly useful for glycoconjugate analysis[16−18]. A complete characterization of glycan isomer structure is not possible utilizing only MS and tandem MS; while high order tandem MS can provide this data in some cases, unattainably high concentrations are necessary from biological systems. In light of this complication, various separation methods must be used in conjunction with MS to obtain unambiguous isomeric glycan structure characterization. Liquid chromatography (LC) separation techniques such as hydrophilic interaction liquid chromatography (HILIC)[19−21], porous graphitized carbon (PGC) chromatography[22−25], and reverse-phase (RP) chromatography [26] are commonly employed methods to investigate glycans.

LC techniques have two distinct phases: the mobile phase and the stationary phase. The mobile, liquid phase contains the dissolved mixture of analytes in solution, while the stationary phase consists of a material coating the interior of the column. The analytes in the mobile phase are then separated as they pass through the column based on the degree of interaction with the stationary phase. For nonpolar stationary phases composed of e.g., C18, analytes that are more nonpolar will remain in the column longer resulting in a longer measured retention time than more polar ones that interact less with the stationary phase and pass through the column more quickly. Thus, nonpolar forces and interactions have been previously studied in relation to protein folding and binding reactions[27,28], and hydrophobic tagging in glycans[29,30].

A useful quantity describing these interactions has been found in the form of the nonpolar surface area (NPSA), which is defined as the sum over all nonpolar surface components of the solvent accessible surface of a compound. In previous work, the NPSA of a compound was computed using an atomistic method described by Muddiman and coworkers [29,30] based upon summing the solvent accessible surface of nonpolar atoms (referred to here as the nonpolar atom (NPA) method). The choice of polar atoms was determined based largely upon chemical intuition. The NPA model will be discussed in greater detail in the methods section. While this was successfully utilized to show how increased NPSA correlates to the retention time and greater ion abundance in electrospray ionization experiments, the method lacks flexibility in the determination of polar/nonpolar atoms as a result of its electronic environment and does not account for the complex 3D structural configuration. Later work by Hu et al. [10] improved the NPA method via an *ad hoc* correction to remove the internal contributions to the solvent surface. The estimated NPSA was then correlated to the experimentally determined retention times of permethylated N-glycans [10] aiding structural identification without the necessity of obtaining costly MS/MS data. A positive correlation was found between the NPSA and the retention time such that larger NPSA values result in longer retention times, indicating a greater degree of nonpolar character and more interaction with the nonpolar stationary phase.

Structural microheterogeneity of glycan isomers still presents a challenge for correct identification, even with combinations of techniques, such as LC-MS[9,31]. Hence, accurate theoretical estimations of the NPSA through a fast and reliable protocol would provide an invaluable tool in biological samples analysis. The purpose of this work is to develop a flexible method based on quantum chemical tools which should be applicable for reliable and automatic determination of the NPSA. The new scheme for calculating the NPSA is based on the following points:

- The NPSA uses the analysis of surface charge distributions computed with continuum solvation (CS) models in order to obtain an intrinsic and flexible criterion for the assignment of polar and nonpolar segments of the molecule's solvent accessible surface
- The definition of the NPSA should refer to the solvent accessible surface and should automatically exclude interior surface areas derived from interchain interaction e.g., in bi- and triantennary glycans

In contrast, the previously utilized NPA-based NPSA model [10,29,30] requires an *a priori* definition of polar and nonpolar atomic contributions and does not account automatically for the identification of the atoms at the solvent accessible surface.

2. Methods

2.1. Nonpolar surface area determination: The CS-NPSA method

The apparent surface charge densities available in continuum solvation models such as the polarized continuum model (PCM) [32,33] or the conductor-like screening model (COSMO) [34] can be used to analyze polar and nonpolar regions of a molecule. These surface charge densities represent the dielectric polarization of the solvent continuum by the solute. They are located on the solvent accessible surface, which is constructed from small surface segments $s_i, i = 1 \ldots N_{seg}$, with N_{seg} being the number of surface segments. Each of these segments is connected with a polarization charge density σ_i. These polarization charges constitute the counter-image in response to the actual solute charge distribution [35] and can be used as a measure for the van der Waals interaction with a surrounding environment. They should, therefore, be well suited to indicate the nonpolar regions of a molecule. In Fig. 1a the polarization charge density for the molecule phenylbutanehydrazide (GPN1, Scheme 1) using COSMO with acetonitrile as solvent is visualized, which will be used further below for comparison with previous work on NPSA [30]. This figure shows that most of the surface has polarization charge densities in the interval of approximately ±0.01 a.u./Å^2 which is associated with the nonpolar regions of the benzene ring and the two C atoms. On the other hand, the pronounced red and blue colors of the positive and negative polarization charges, respectively, are derived from the carbonyl O atom and NH and NH_2 groups. The histogram of the polarization charge density elements (denoted as σ profile) [35] can be found in Fig. 2 for acetonitrile. The σ profile for GPN1 using several other solvents of varying polarity have also been presented for comparison in the Supporting Information (SI), Fig. S1. Only small apparent changes are observed upon changing the solvent environment. For comparison, the σ profile of a nonpolar model compound composed of a linear chain of nine cyclohexane rings is shown in Fig. S2. For this type of nonpolar molecule, the NPSA should be practically equal to the total solvent surface area. From this figure it can be seen that using a threshold of ±0.01 a.u./Å^2 will encompass practically all of the entire polarization charge density. This limit also fits well for the definition of the NPSA of GPN and of the glycans discussed below. Thus, we decided to use it further on

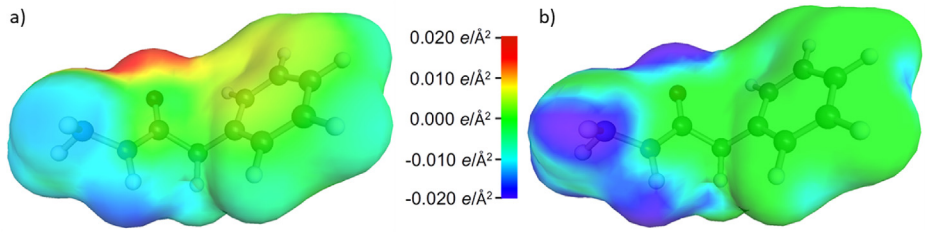
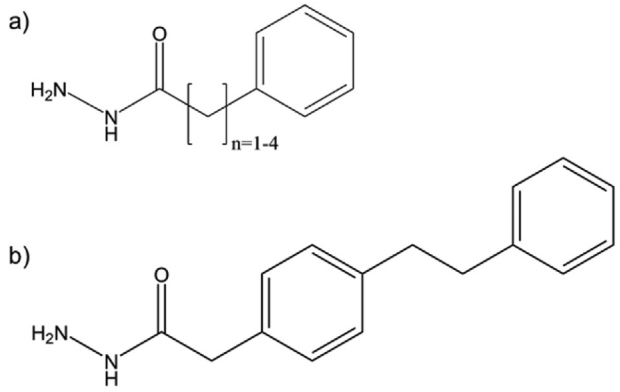


Fig. 1. a) visualization of the polarization charge density of GPN1 from COSMO calculations using acetonitrile solvent and b) nonpolar (green) and polar surfaces (blue). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)



Scheme 1. Structures of a) GPNn for n = 1–4, and b) phenyl2-GPN.

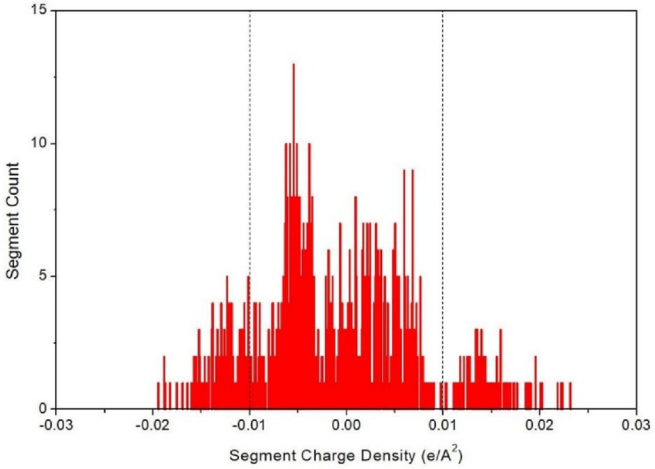


Fig. 2. σ profile of GPN1 in acetonitrile.

in our investigations. Consequently, the CS-NPSA is calculated as the sum of the area elements s_i for all segments with the norm of the charge/area smaller than the chosen threshold of 0.010 a.u./Angs^2. The remaining surface area is denoted as polar surface. Based on these criteria, the nonpolar CS-NPSA (green regions) and polar (blue regions) surfaces are shown in Fig. 1b for GPN1 using the values of the polarization charge density presented in Fig. 1a.

The CS-NPSA method presented above is compared to the NPA model [29,30] (Fig. S3), where the total surface area of all nonpolar atoms (carbon and hydrogen atoms, but excluding polar atoms such as oxygen and nitrogen and all atoms connected to them) is calculated as spheres with the van der Waals radii. From this total surface area value, overlapping regions corresponding to bonds are subtracted to give the NPSA, using standard bond lengths.

Two particular deficiencies of the NPA method are the definitions of which atoms contribute to the NPSA based upon an *a priori* determination of atomic polarity derived from chemical intuition, and the inability to easily reduce the calculated NPSA by the solvent accessible surface area coming from atoms buried in the interior regions of the molecule by folding. A consequence of these issues is that the NPSA of isomers determined utilizing the NPA model would have the same NPSA, thus this would be ineffective in the differentiation and characterization of isomers present in an LC chromatogram.

In the case of the NPA model, the NPSA is calculated from a yes/no decision based on the atom type and the bonding environment. If an atom is determined to contribute, then the entire contribution of the solvent accessible surface of that atom is added to the NPSA. On the contrary, this yes/no decision in the CS-NPSA (Fig. 3a) is made in a more flexible way since each small surface element of the solvent accessible surface is evaluated according to the respective polarization charge density. Fig. 3b shows the actual nonpolar and polar surfaces determined in that way. Moreover, this charge density is not determined *a priori* but computed by means of the quantum chemical formalism of the continuum solvation method. Additionally, the general problem of the NPA method is that it does not account for the amount of inner surface that is lost to the interior of large polymer molecules and is not solvent-accessible. This problem is solved in the CS-NPSA method consistently since only the solvent-accessible surface is constructed by means of the COSMO approach. There is no need to make additional *ad hoc* corrections to remove internal surface areas as already mentioned in the Introduction[10]. Moreover, the CS-NPSA method can differentiate isomers based upon their NPSA, whereas the NPA method would find that they are identical. To show this, two pairs of isomers are investigated here based upon the bonding location of either fucose (core or branch) or sialic acid (2,3- or 2,6-connected).

2.2. Quantum chemistry calculations

When performing the quantum chemical calculations, starting geometries for the glycans were generated using the GLYCAM-Web Carbohydrate Builder tool (http://www.glycam.org)[36], where the dihedral angles of the glycosidic bonds (rotamers) were chosen to create a series of conformers with consistent structural characteristics and were then optimized using the Glycam-parameterized AMBER force field. The choice of rotamers is further discussed in section 3.2. It should be noted that only a consistent set of structures was constructed without aiming for global energy optimizations. All glycan structures were permethylated to maximize the nonpolar functionality and coordinate with experimental findings. Each glycan structure was optimized further at the density functional theory (DFT) level utilizing the Perdew, Becke, and Ernzerhof (PBE) [37] and the split valence (SV) basis set[38]. The nonpolar cyclohexane chain model utilized this optimization method as well

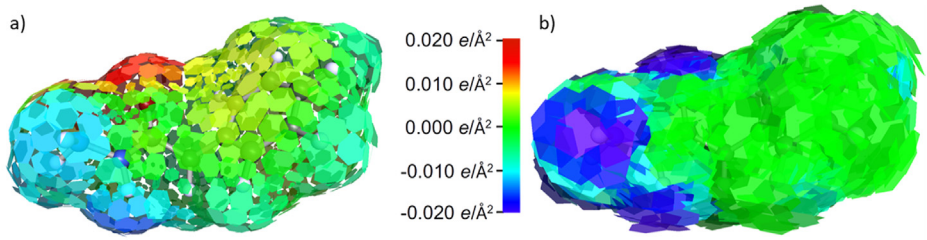


Fig. 3. a) Polarization charge density of individual segments of GPN1 and b) resulting nonpolar surface (green) and polar surface (blue). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

due to its computational efficiency, while the simpler GPN structures were calculated with the Becke three-parameter Lee, Yang, and Parr (B3LYP) [39] density functional with the SVP and def2-TZVP [40] basis sets. The D3 dispersion correction [41] was utilized for nonbonding interactions. To aid computational efficiency, the resolution of the identity (RI) approximation was employed with the multipole accelerated RI method for the Coulomb integrals (MARI-J)[42]. The effect of the solvent was taken into account using the conductor-like screening model (COSMO) method using acetonitrile as the solvent ($\varepsilon = 35.69$). While DFT calculations are utilized here to calculate the NPSA with the CS-NPSA method, other methods could be utilized so long as they use the continuum solvation method where the numerical representation of the tessellated van der Waals surface and polarization charges are available.

The TURBOMOLE electronic structure program package [43] was utilized for all quantum chemistry calculations. The visualization of the COSMO polarization charge density and the CS-NPSA has been performed by means of COSMOview, part of TmoleX [44,45]. Jmol, an open-source Java viewer for chemical structures in 3D[46], has been used for 3D visualization of molecular structures.

3. Results and discussion

3.1. Comparison of the CS-NPSA and NPA methods

The analysis shown in Fig. 4 demonstrates the way individual atoms can contribute partially to the nonpolar as well as the polar surfaces. In Fig. 4a, the polarization charges are shown for GPN3. Fig. 4b is created by switching to a per-atom basis based on the surface segment sections assigned to each atom in the course of the construction of the solvent accessible surface by the COSMO calculation. In the coloring scheme of Fig. 4b an atom is considered nonpolar (colored blue) if it contributes at least 50% of its associated area to the NPSA. The association of the charge polarization segments is directly available from the tessellation process of the scaled van der Waals surface for each atom[34]. Analysis of the charge density polarization of segments attributed to these atoms finds that only the hydrogen atoms bonded to the nitrogen atoms and the oxygen atom are sufficiently polarized to be considered completely polar. The surface contributing to the NPSA is found colored green in Fig. 4c and correlates well to the nonpolar atoms in Fig. 4b.

It is noted from Fig. 4b that even the two nitrogen atoms and the carbonyl carbon atom are included in the list of nonpolar atoms. They would have been ignored by the previous NPA scheme. Fig. 5 shows the nonpolar surface of Fig. 4c in more detail from different angles to better indicate the nonpolar area (red) of each of these three atoms. The two nitrogen atoms do not contribute equally to the NPSA, though. Indeed only 53% of the polarization segments attributed to N10 are counted as nonpolar (Table 1, Fig. 5b), but 93% in case of N9 (Fig. 5a). The surface area assigned to the carbonyl carbon, C8 (Fig. 5c), is found to be highly nonpolar despite bonding to both oxygen and nitrogen atoms. A statistical overview of the nonpolar areas and atomic contributions is given in Table S1.

Another example is found in Fig. 6 for a model dextran, Glc12, consisting of 12 glucose residues. The NPA method would neglect the NPSA contributions of all oxygen atoms and carbon atoms, since each carbon atom is bonded to at least one oxygen atom and thus judged too polar. With the NPA scheme, only the hydrogen atoms would contribute to the NPSA. The total COSMO surface is shown in Fig. 6a, from which the nonpolar atoms in blue (Fig. 6b) are determined using the CS-NPSA method. A statistical overview of the nonpolar areas and atomic contributions is given in Table S1. Almost all of the methyl carbon atoms contribute to the NPSA, while the carbon atoms excluded from the NPSA, colored in yellow, are found to be those buried in the interior of the surface contributing no surface segments. The σ profile for GLC 12 is shown in Fig. S4 and is similar to that shown for the cyclohexane chain model system (Fig. S2), though a second much smaller peak of positively charged surface segments appears.

3.2. Applications

Increased NPSA has been shown previously [10,29,30] to correlate to increased retention times in liquid chromatographic separation columns. In the case of the commonly used C18 column, the increase in retention time is supposed to result from the enhanced nonpolar functionalization allowing for greater

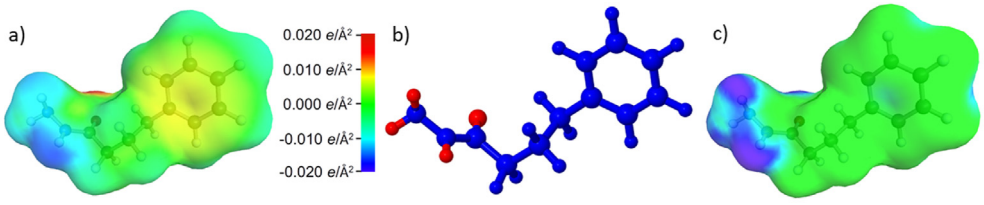


Fig. 4. GPN3 optimized using the B3LYP-D3/SVP method in the COSMO/acetonitrile environment. Representation a) shows the polarization charges, b) the polar atoms in red, nonpolar atoms, in blue, contributing at least half of their total surface area to the NPSA, and c) the nonpolar surface in green and polar surface in blue. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

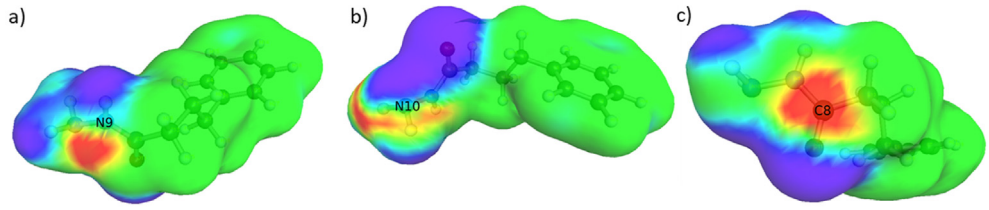


Fig. 5. Charge segment analysis of GPN3 for atoms: a) N9, b) N10, and c) C8 in acetonitrile. The images show the modified CP-NPSA surface where the red regions correspond to the nonpolar charge polarization segments for the atom of interest, the blue show the polar surface, and the green show the remaining nonpolar surface. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 1
Total surface area values and percentage of the total surface which contributes to the CS-NPSA value for C8, N9, and N10 of GPN3 in acetonitrile.

NPSA for atom	Total (Å^2)	Percentage nonpolar
C8	7.0	99
N9	7.0	93
N10	15.1	53

Table 2
Retention times [10] and CS-NPSA values for the dextran linear glycans, Glc9-Glc14, in gg conformation.

Glycan	CS-NPSA (Å^2)	Ret. Time (min.)
Glc 9	1270.8	28.5
Glc 10	1375.9	29.8
Glc 11	1520.0	31.0
Glc 12	1649.7	32.1
Glc 13	1772.9	33.2
Glc 14	1878.7	34.3

interaction with the nonpolar stationary phase, meaning that the analyte remains in the column for a longer time. An illustrative example comes from a series of extending linear glycan chains, dextrans, Glc9-Glc14 (Table 2). With each successive member of the series, only an additional glucose residue is added to the chain, where each member of the series forms an elongated helical structure as shown in Fig. 6. The connection of glycosidic bonds was always gauche/gauche (gg). A positive and strongly linear relationship (Fig. 7a, $R^2 = 0.999$) is demonstrated between the calculated CS-NPSA value and the experimentally determined retention time[10]. A similar linear relationship is demonstrated for the other two sets of conformers (trans/gauche (tg) and gt, Fig. 7b and c). Each series of conformers has an R^2 value greater than 0.9, but shows the strongest correlation for the gg case. Such a linear relationship can also be found for the series of extending GPN molecules as well (Table S2 and Fig. 8). In this case the above-mentioned relationship of NPSA and retention time is reversed because the retention time measurements were made in this case with the hydrophobic interaction liquid chromatograph (HILIC) technique.

The threshold limits for the determination of the CS-NPSA calculation were varied to either 0.008 or 0.012 e/Å^2 for the GPN and linear glycan molecules (Tables S3-S4 and Fig. S5-S6). The computed CS-NPSA values decrease somewhat with respect to the smaller threshold as fewer segments are contributing, while the larger threshold increases the CS-NPSA slightly as more segments qualify to contribute. The quality of the linear regression with respect to the retention times (Fig. S5-S6) remains practically unchanged. The coefficient of determination (R^2) value increases only slightly in either situation. Another test was performed concerning basis set sensitivity of the CS-NPSA values. Increasing the basis set size for the GPN series changes the CS-NPSA values only slightly (Table S2). The R^2 values (Fig. S7) increase also somewhat. Variation of the tessellation parameters beyond the standard values did not show noteworthy changes in the NPSA values.

Two sets of standard biantennary glycan isomers were analyzed: one set differed in either the 2,3 or 2,6 linkage of the sialic acid residue, and a second set differed in the placement of the fucose residue, either at the N-Acetylglucosamine (GlcNAc) core or on a branch. The core- and branch-fucosylated glycan isomers are shown in Fig. 9. Since the linear glycan conformers corresponding to the choice of gg conformation showed the strongest correlation between the CP-NPSA values and retention times, this rotamer was chosen for the biantennary structures as well. The σ profiles for each of the four biantennary glycans (Fig. S8) are qualitatively similar to that of the GLC 12 linear glycan (Fig. S4). A greater amount of total COSMO surface is found in Fig. 9c (1350.3 Å^2) for the branch-fucosylated structure compared to the core-fucosylated structure in Fig. 9a (1295.6 Å^2). Looking at the atomistic representation of the CS-NPSA, the branch-fucosylated structure (Fig. 9d) is less compact than the core-fucosylated structure (Fig. 9b) with more nonpolar atoms contributing to the CS-NPSA (blue atoms). A statistical overview of the nonpolar areas and atomic contributions for the latter molecule is given in Table S1.

The correlation is shown for all four biantennary glycan standards in Fig. 10 and Table S5. Within each set of isomers, the elution order correctly correlates with the calculated CS-NPSA values: the structures with the smaller retention time (core-fucosylated and 2,6-sialylated) have smaller CS-NPSA values than their isomer (branch-fucosylated and 2,3-sialylated). An excellent linear correlation is demonstrated with a R^2 of 0.994 between the retention times and CS-NPSA values. This further justifies the choice of gg

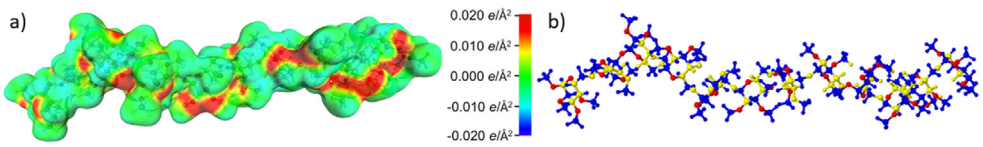


Fig. 6. Dextran linear glycan, Glc12, in the gg conformation optimized using the PBE-D3/SV method in the COSMO/acetonitrile environment. Representation a) shows the polarization charges, and b) the polar atoms in red, nonpolar atoms contributing to the NPSA in blue, and buried nonpolar atoms that do not contribute to the NPSA in yellow. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

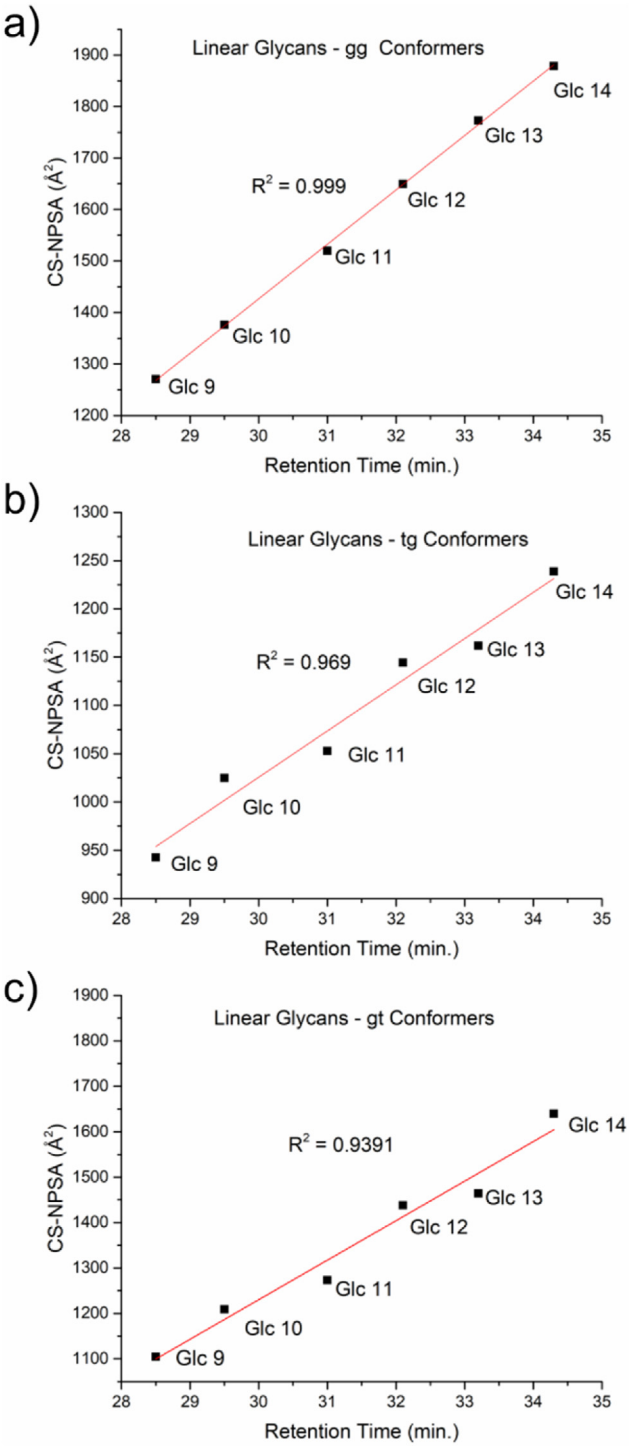


Fig. 7. Linear regression correlating experimentally determined retention times [10] with CS-NPSA values for the series of linear glycans (Table 2) using the PBE-D3/SV approach. The results for the three investigated conformers are shown with a) gg, b) tg, and c) gt conformations.

conformers for the biantennary glycans. It is further noted that varying the threshold values for the biantennary glycans from the standard value of 0.010 to 0.008 or 0.012 $e/Å^2$ (Table S5 and Fig. S9) shows only to slightly increases the R^2 value (0.995 and 0.996, respectively).

The CS-NPSA values of Fig. 10 and the linear correlation with experimental retention times are compared with a corresponding

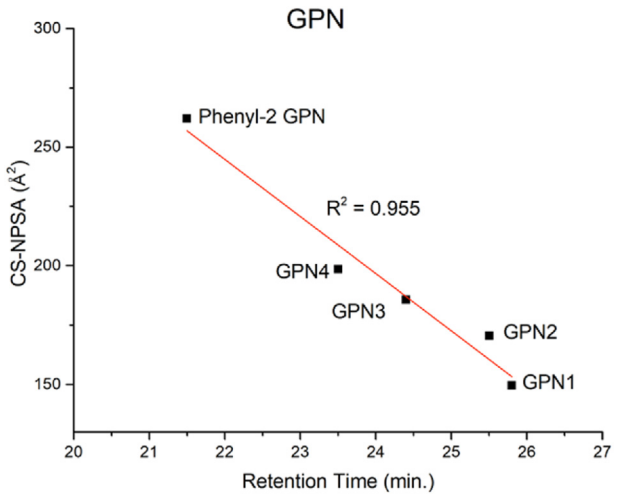


Fig. 8. Linear regression correlating experimentally determined retention [30] times with NPSA calculated with the CS-NPSA method.

plot obtained with the NPSA values of the NPA method (Fig. S10). The important difference is that with the NPA method the NPSA values for the isomer pairs are identical because structural differences are not accounted for in the latter method. Thus, it cannot be used to distinguish these isomers. Moreover, comparison of the NPSA values between the two methods shows much larger values in the range of 2000 $Å^2$ to 2450 $Å^2$ for NPA instead of 1100 $Å^2$ to 1400 $Å^2$ for CS-NPSA. This difference demonstrates the significant contribution of internal surface areas of the two arms in the biantennary glycans (Fig. 9b and d), which would not be available to adsorption interactions with the stationary phase of a LC column. A similar comparison for the more stretched GLC gg conformers shows a much smaller loss of about 200 $Å^2$ to 375 $Å^2$ (Table S5).

4. Conclusions

A new and flexible method for calculating the NPSA of a molecule has been proposed based on the polarization charge densities obtained from a COSMO calculation. The fact that these charge densities are assigned to small segments of the solvent accessible surface of a molecule allows an individual classification of segments as polar or nonpolar. The sum of all nonpolar segments defines the NPSA. From the known assignment of surface segments to the atoms in the course of the tessellation process, atoms can still be classified as contributing to the NPSA or to the polar surface or as not contributing at all to the solvent accessible surface since they are buried inside of the molecule. It is interesting to note that beyond atoms normally classified as nonpolar, others such as those neighboring polar atoms or nitrogen atoms contributed at least partially to the NPSA. All assignments can be performed automatically from the results of a PCM calculation or any convenient variant. In the present calculations the COSMO method had been used, but other continuum solvation methods could also be used as long as they provide the information about the tessellation structure of the solvent polarization charges. The CS-NPSA analysis program developed in this work is a straightforward follow-up program to a quantum chemical calculation providing this information and is available for download in the Internet [47].

The advantage of the current CS-NPSA method over the previous NPA method is, beyond the more detailed quantum chemical analysis of the polarization charges, the direct use of the solvent accessible surface, which allows the distinction of isomers — a feature of great practical importance. On the contrary, the NPA

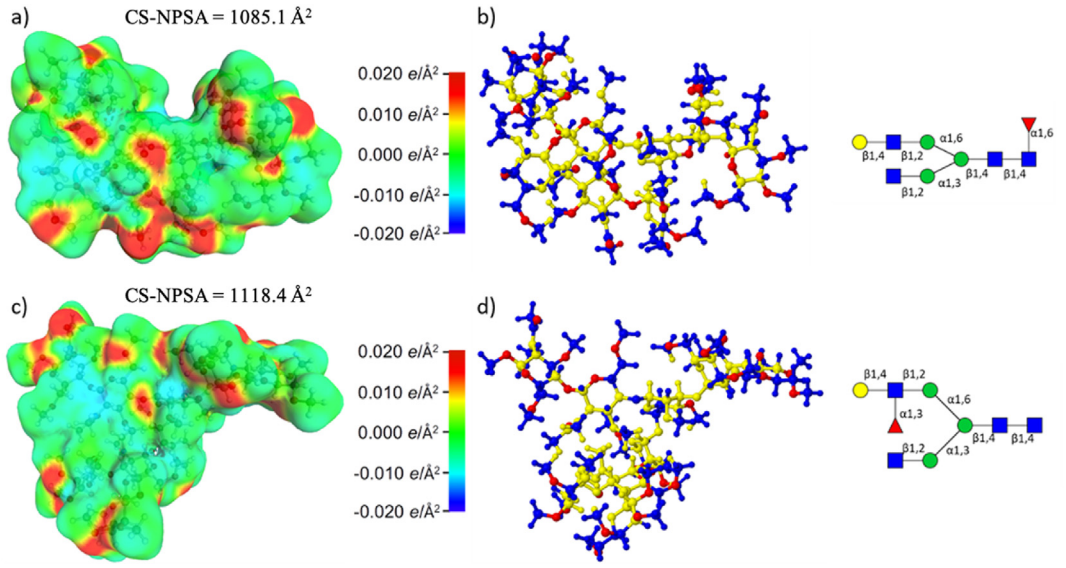


Fig. 9. Biantennary standard glycan structures in gg conformation optimized using the PBE-D3/SV method in the COSMO/acetonitrile environment. Representations a) and c) show the polarization charges and CS-NPSA values, and b) and d) the polar atoms in red, nonpolar atoms contributing to the NPSA in blue, and buried nonpolar atoms that do not contribute to the NPSA in yellow. The core-fucosylated structure is shown in a) and b), while the branch-fucolyslated structure is shown in c) and d). Symbolic representations: blue square – N-Acetylglucosamine (GlcNAc), green circle – mannose, yellow circle – galactose, red triangle – fucose. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

method will give identical NPSA values in this case. Moreover, since the NPSA values are derived directly from a solvent accessible surface, they will not contain, in contrast to the NPA method, contributions of inner surface segments of globular structures.

The usefulness of the CS-NPSA approach has been demonstrated in connection with the correlation of the NPSA with retention times measured by liquid chromatography separation. Applying the method to selected examples of a series of GPN chains, linear glycan chains and biantennary glycans shows excellent correlation between NPSA values and measured retention times. Because of the versatility of the method, application in the field of glycan separation and identification, for peptides and glycopeptides and many more classes of compounds can be envisaged.

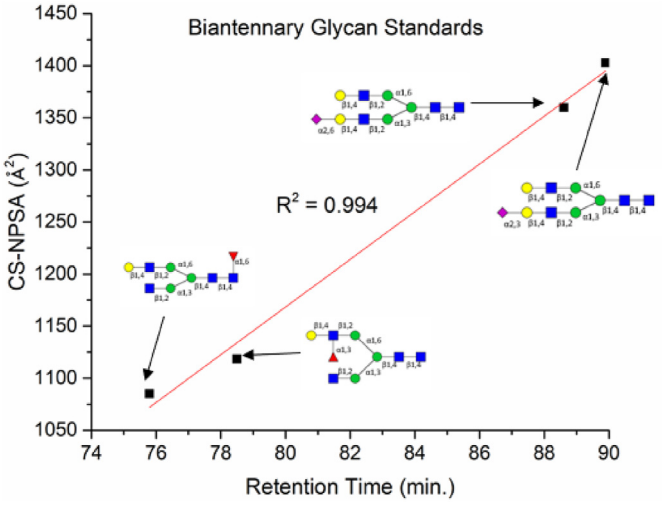


Fig. 10. Linear regression correlating experimentally determined retention times with CS-NPSA values for the core- and branch-fucosylated isomers and the 2,3- and 2,6-sialylated isomers in gg conformation using the PBE-D3/SV approach (Table S5).

Supporting information

σ profiles of GPN1 for various solvents, σ profile for the nonpolar model of nine cyclohexane rings, scheme for the NPA model, retention times for the GPN series and the biantennary glycans.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by grants from NIH (1R01GM112490-06 and 1R01GM130091-02). It was funded in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001. We thank the HPCC of Texas Tech University for computer time.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijms.2020.116495.

References

[1] A. Helenius, Aebi, markus, intracellular functions of N-linked glycans, Science 291 (2001) 2364–2369.
[2] K.W. Moremen, M. Tiemeyer, A.V. Nairn, Vertebrate protein glycosylation: diversity, synthesis and function, Nat. Rev. Mol. Cell Biol. 13 (2012) 448–462.
[3] M.L. Bermingham, M. Colombo, S.J. McGurnaghan, L.A.K. Blackbourn, F. Vučković, M. Pučić Baković, I. Trbojević-Akmačić, G. Lauc, F. Agakov, A.S. Agakova, C. Hayward, L. Klarić, C.N.A. Palmer, J.R. Petrie, J. Chalmers, A. Collier, F. Green, R.S. Lindsay, S. Macrury, J.A. McKnight, A.W. Patrick, S. Thekkepat, O. Gornik, P.M. McKeigue, H.M. Colhoun, N-glycan profile and kidney disease in type 1 diabetes, Diabetes Care 41 (2018) 79–87.
[4] A.V. Everest-Dass, E.S.X. Moh, C. Ashwood, A.M.M. Shathili, N.H. Packer, Human disease glycomics: technology advances enabling protein glycosylation analysis – part 1, Expert Rev. Proteomics 15 (2018) 165–182.
[5] A.V. Everest-Dass, E.S.X. Moh, C. Ashwood, A.M.M. Shathili, N.H. Packer,

Human disease glycomics: technology advances enabling protein glycosylation analysis – part 2, Expert Rev. Proteomics 15 (2018) 341–352.
[6] M. Wang, J. Zhu, D.M. Lubman, C. Gao, Aberrant glycosylation and cancer biomarker discovery: a promising and thorny journey, Clin. Chem. Lab. Med. 57 (2019) 407–416.
[7] D.H. Dube, C.R. Bertozzi, Glycans in cancer and inflammation — potential for therapeutics and diagnostics, Nat. Rev. Drug Discov. 4 (2005) 477–488.
[8] P.M. Rudd, T. Elliott, P. Cresswell, I.A. Wilson, R.A. Dwek, Glycosylation and the immune system, Science 291 (2001) 2370–2376.
[9] X. Dong, Y. Huang, B.G. Cho, J. Zhong, S. Gautam, W. Peng, S.D. Williamson, A. Banazadeh, K.Y. Torres-Ulloa, Y. Mechref, Advances in mass spectrometry-based glycomics, Electrophoresis 39 (2018) 3063–3081.
[10] Y. Hu, T. Shihab, S. Zhou, K. Wooding, Y. Mechref, LC–MS/MS of permethylated N-glycans derived from model and human blood serum glycoproteins, Electrophoresis 37 (2016) 1498–1505.
[11] L. Veillon, Y. Huang, W. Peng, X. Dong, B.G. Cho, Y. Mechref, Characterization of isomeric glycan structures by LC-MS/MS, Electrophoresis 38 (2017) 2100–2114.
[12] S. Zhou, Y. Huang, X. Dong, W. Peng, L. Veillon, D.A.S. Kitagawa, A.J.A. Aquino, Y. Mechref, Isomeric separation of permethylated glycans by porous graphitic carbon (PGC)-LC-MS/MS at high temperatures, Anal. Chem. 89 (2017) 6590–6597.
[13] A. Banazadeh, R. Nieman, M. Goli, W. Peng, A. Hussein, E. Bursal, H. Lischka, Y. Mechref, Characterization of glycan isomers using magnetic carbon nanoparticles as a MALDI co-matrix, RSC Adv. 9 (2019) 20137–20148.
[14] W.R. Alley, M.V. Novotny, Glycomic analysis of sialic acid linkages in glycans derived from blood serum glycoproteins, J. Proteome Res. 9 (2010) 3062–3072.
[15] C.H. Chung, B. Mirakhur, E. Chan, Q.-T. Le, J. Berlin, M. Morse, B.A. Murphy, S.M. Satinover, J. Hosen, D. Mauro, R.J. Slebos, Q. Zhou, D. Gold, T. Hatley, D.J. Hicklin, T.A.E. Platts-Mills, Cetuximab-induced anaphylaxis and IgE specific for galactose-α-1,3-galactose, N. Engl. J. Med. 358 (2008) 1109–1117.
[16] A. Dell, H.R. Morris, Glycoprotein structure determination by mass spectrometry, Science 291 (2001) 2351–2356.
[17] Y. Mechref, M.V. Novotny, Structural investigations of glycoconjugates at high sensitivity, Chem. Rev. 102 (2002) 321–370.
[18] M.V. Novotny, Y. Mechref, New hyphenated methodologies in high-sensitivity glycoprotein analysis, J. Separ. Sci. 28 (2005) 1956–1968.
[19] A.J. Alpert, Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds, J. Chromatogr. A 499 (1990) 177–196.
[20] K.R. Anumula, Advances in fluorescence derivatization methods for high-performance liquid chromatographic analysis of glycoprotein carbohydrates, Anal. Biochem. 350 (2006) 1–23.
[21] M. Wuhrer, A.R. de Boer, A.M. Deelder, Structural glycomics using hydrophilic interaction chromatography (HILIC) with mass spectrometry, Mass Spectrom. Rev. 28 (2009) 192–206.
[22] L.R. Ruhaak, A.M. Deelder, M. Wuhrer, Oligosaccharide analysis by graphitized carbon liquid chromatography–mass spectrometry, Anal. Bioanal. Chem. 394 (2009) 163–174.
[23] K. Stavenhagen, D. Kolarich, M. Wuhrer, Clinical glycomics employing graphitized carbon liquid chromatography-mass spectrometry, Chromatographia 78 (2015) 307–320.
[24] D. Kolarich, M. Windwarder, K. Alagesan, F. Altmann, Isomer-specific analysis of released N-glycans by LC-ESI MS/MS with porous graphitized carbon. Glyco-Engineering, Springer, 2015, pp. 427–435.
[25] S. Zhou, X. Dong, L. Veillon, Y. Huang, Y. Mechref, LC-MS/MS analysis of permethylated N-glycans facilitating isomeric characterization, Anal. Bioanal. Chem. 409 (2017) 453–466.
[26] G.C. Vreeker, M. Wuhrer, Reversed-phase separation methods for glycan analysis, Anal. Bioanal. Chem. 409 (2017) 359–378.
[27] A. Nicholls, K.A. Sharp, B. Honig, Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons, Proteins: Struct., Funct., Bioinf. 11 (1991) 281–296.
[28] R.S. Spolar, J.R. Livingstone, M.T. Record, Use of liquid hydrocarbon and amide transfer data to estimate contributions to thermodynamic functions of protein folding from the removal of nonpolar and polar surface from water, Biochemistry 31 (1992) 3947–3955.
[29] S.H. Walker, B.N. Papas, D.L. Comins, D.C. Muddiman, Interplay of permanent charge and hydrophobicity in the electrospray ionization of glycans, Anal. Chem. 82 (2010) 6636–6642.
[30] S.H. Walker, L.M. Lilley, M.F. Enamorado, D.L. Comins, D.C. Muddiman, Hydrophobic derivatization of N-linked glycans for increased ion abundance in electrospray ionization mass spectrometry, J. Am. Soc. Mass Spectrom. 22 (2011) 1309–1317.
[31] A. Yu, J. Zhao, W. Peng, A. Banazadeh, S.D. Williamson, M. Goli, Y. Huang, Y. Mechref, Advances in mass spectrometry-based glycoproteomics, Electrophoresis 39 (2018) 3104–3122.
[32] S. Miertuš, E. Scrocco, J. Tomasi, Electrostatic interaction of a solute with a continuum. A direct utilizaion of AB initio molecular potentials for the prevision of solvent effects, Chem. Phys. 55 (1981) 117–129.
[33] C. Amovilli, V. Barone, R. Cammi, E. Cancès, M. Cossi, B. Mennucci, C.S. Pomelli, J. Tomasi, Recent advances in the description of solvent effects with the polarizable continuum model, in: P.-O. Löwdin (Ed.), Adv. Quantum Chem., Academic Press, 1998, pp. 227–261.
[34] A. Klamt, G. Schüürmann, COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient, J. Chem. Soc., Perkin Trans. 2 (1993) 799–805.
[35] A. Klamt, Conductor-like screening model for real solvents: a new approach to the quantitative calculation of solvation phenomena, J. Phys. Chem. 99 (1995) 2224–2235.
[36] K.N. Kirschner, A.B. Yongye, S.M. Tschampel, J. González-Outeiriño, C.R. Daniels, B.L. Foley, R.J. Woods, GLYCAM06: a generalizable biomolecular force field, Carbohydrates, J. Comput. Chem. 29 (2008) 622–655.
[37] J.P. Perdew, K. Burke, M. Ernzerhof, Generalized gradient approximation made simple, Phys. Rev. Lett. 77 (1996) 3865–3868.
[38] A. Schäfer, H. Horn, R. Ahlrichs, Fully optimized contracted Gaussian basis sets for atoms Li to Kr, J. Chem. Phys. 97 (1992) 2571–2577.
[39] A.D. Becke, Density-functional thermochemistry. III. The role of exact exchange, J. Chem. Phys. 98 (1993) 5648–5652.
[40] F. Weigend, R. Ahlrichs, Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: design and assessment of accuracy, Phys. Chem. Chem. Phys. 7 (2005) 3297–3305.
[41] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu, J. Chem. Phys. 132 (2010) 154104.
[42] C. Hättig, F. Weigend, CC2 excitation energy calculations on large molecules using the resolution of the identity approximation, J. Chem. Phys. 113 (2000) 5154–5161.
[43] R. Ahlrichs, M. Bär, M. Häser, H. Horn, C. Kölmel, Electronic structure calculations on workstation computers: the program system TURBOMOLE, Chem. Phys. Lett. 162 (1989) 165–169.
[44] C. Steffen, K. Thomas, U. Huniar, A. Hellweg, O. Rubner, A. Schroer, TmoleX—a graphical user interface for TURBOMOLE, J. Comput. Chem. 31 (2010) 2967–2970.
[45] BIOVIA - Dassault Systèmes, TmoleX:. https://www.3ds.com/products-services/biovia/products/molecular-modeling-simulation/solvation-chemistry/turbomoler/, 2020.
[46] Jmol: an open-source Java viewer for chemical structures in 3D.. http://www.jmol.org/.
[47] CS-NPSA - a package to analyze the non polar surface area. https://cs-npsa.sourceforge.io/.