Carbohydrate Research 344 (2009) 541–544

Contents lists available at ScienceDirect

Carbohydrate Research

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

Note

Systematic probing of an atomic charge set of sialic acid disaccharides for the rational molecular modeling of avian influenza virus based on molecular dynamics simulations

Youngjin Choi [a], Hyunmyung Kim [b], Kum Won Cho [c], Seung R. Paik [d], Hyun-Won Kim [e], Karpjoo Jeong [f,*], Seunho Jung [b,*]

[a] Biochip Research Center, Hoseo University, Asan 336-795, Republic of Korea
[b] Department of Bioscience and Biotechnology, Konkuk University, 1 Hwayang-dong Gwangjin-gu, Seoul 143-701, Republic of Korea
[c] Korea Institute of Science and Technology Information, Daejeon 305-806, Republic of Korea
[d] School of Chemical and Biological Engineering, College of Engineering, Seoul National University, Seoul 151-744, Republic of Korea
[e] Department of Biochemistry and Institute of Basic Medical Science, Wonju College of Medicine, Yonsei University, Wonju 220-701, Republic of Korea
[f] College of Information and Communication & Department of Advanced Technology Fusion, Konkuk University, 1 Hwayang-dong Gwangjin-gu, Seoul 143-701, Republic of Korea

ARTICLE INFO

Article history:
Received 7 October 2008
Received in revised form 19 December 2008
Accepted 23 December 2008
Available online 31 December 2008

Keywords:
Avian flu
Conformational analysis
Molecular dynamics simulations
Partial charge
Sialic acid

ABSTRACT

A systematic searching approach for an atomic charge set through molecular dynamics simulations is introduced to calculate a reasonable sialic acid carbohydrate conformation with respect to the experimentally observed structures. The present molecular dynamics simulation study demonstrated that B3LYP/6-31G is the most suitable basis set for the sialic acid disaccharides, attaining good agreement with experimental data.

Carbohydrates are the most abundant biological molecules in nature and have many important biological functions for cell–cell communication as well as energy storage.[1] However, our knowledge about these molecules remains very limited due to their inherent diversity and flexibility via glycosidic linkages.[2] As such, it is widely considered as a challenging task for structural scientists to obtain the accurate solution conformation of carbohydrates.

Computational approaches to establish structural features of carbohydrates have long been regarded as good alternatives to experimental methods such as X-ray crystallography and NMR spectroscopy. In many cases, accurate molecular parameters of the partial charge sets are essential for computer simulations of biological carbohydrates in order to describe atomic interactions within the molecules.[3,4]

In this report, a systematic searching approach for an atomic charge set was introduced to calculate a reasonable sialic acid carbohydrate conformation with respect to the experimentally observed structures. Sequential molecular dynamics (MD) simulations were performed for biological sialic acid compounds with various kinds of partial charge sets.

Sialic acid carbohydrates are commonly occurring residues in various cell-surface proteins, and they function as key molecular determinants, especially for influenza A-type virus.[5] This residue is normally linked to the terminal galactose residue via an α-(2→3)- or α-(2→6)-glycosidic linkage, where the specific linkage type determines species preference between bird and human.[6,7] Figure 1 is a 2D representation of each α-(2→3)-linked and α-(2→6)-linked sialic acid disaccharide. Pandemic avian influenza viruses exert a fatal effect upon poultry, and even upon humans in some cases.[5] Extensive surveys of the viruses isolated from avian, swine, equine, and human sources show a correlation of receptor-binding preference with species of origin. Human isolates prefer the terminal sialic acids of glycoprotein and glycolipid receptors in the α-(2→6)-linkage to galactose, avian and equine ones prefer α-(2→3)-linkages, and viruses from swine bind to both.[8] A switch in receptor specificity from sialic acid disaccharides in the α-(2→3)-linkage (avian) to the α-(2→6)-linkage (human) is a major obstacle for influenza A virus to cross the species barrier and to adapt to a new host.[5] From this viewpoint, establishing the optimal molecular parameter set for the sialic acid disaccha-

* Corresponding authors. Tel.: +82 2 450 3520; fax: +82 2 452 3611 (S.J.).
E-mail addresses: jeongk@konkuk.ac.kr (K. Jeong), shjung@konkuk.ac.kr (S. Jung).

0008-6215/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.carres.2008.12.018

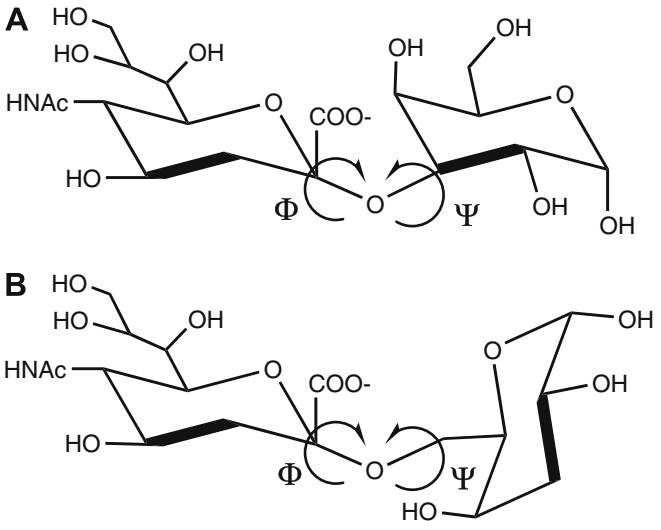

Figure 1. 2-D structural representations of (A) Sialyl-α-(2→3)-Gal and (B) Sialyl-α-(2→6)-Gal.

rides would be a valuable undertaking for carrying out further computational research to resolve key issues related with the avian flu virus.

The present MD simulation studies were conducted with the CHARMM program[9] with 10 different sets of partial charges. The high conformational variations reflected an importance for the use of a correct set of atomic partial charges. For our sialic acid disaccharide systems, MD simulations with Mulliken set of charges, derived from DFT B3LYP[10,11] (Becke three-parameter Lee-Yang-Parr)/6-31 calculations, were demonstrated to correctly reproduce the experimental data.

Table 1 summarizes the results for the partial charge values of representative oxygen atoms in sialic acid disaccharides depending on the approximation level and basis set. As may be expected from the available results in the literature, the Mulliken charges of oxygen atoms cover a very broad range of values from −0.30 at the low side to −0.95 at the high side at the Hartree–Fock level.[12,13] One can observe a steep increase in the negative charge for oxygen when increasing the basis set size from STO-3G to 3-21G, but a steady increase from 3-21G to 6-311G basis sets. This was also true for the case of Becke3LYP level, whereas semi-empirical methods such as AM1 and PM3 showed lower negative charge values for glycosidic oxygen atoms. The B3LYP level predicts uniformly smaller oxygen charges compared to the HF level. Furthermore, partial charge values from B3LYP were rather equally distributed through the carboxylic and glycosidic oxygens compared to those of HF. In particular, the B3LYP/6-31G conditions showed a distribution of −0.5497 for O1, −0.5946 for O11, and −0.5227 for O2, respectively. B3LYP has been successfully applied for the calculation of the

Table 1
Partial charge (q_0) of carboxylic and glycosidic oxygen as calculated from different ab initio parameters

Method	$O_{coo-}1$	$O_{coo-}11$	$O_{glycosidic}$	$O_{glycopyranosic}$
AM1	−0.5270	−0.5672	−0.2795	−0.3550
PM3	−0.5484	−0.6291	−0.2666	−0.3090
HF/STO-3G	−0.4603	−0.4547	−0.2763	−0.2650
HF/3-21G	−0.7463	−0.7747	−0.6632	−0.6370
HF/6-31G	−0.6963	−0.7979	−0.7328	−0.6250
HF/6-311G	−0.6420	−0.7464	−0.6660	−0.5770
B3LYP/STO-3G	−0.3151	−0.3149	−0.2201	−0.2360
B3LYP/3-21G	−0.5826	−0.5998	−0.4897	−0.4960
B3LYP/6-31G	−0.5497	−0.5946	−0.5227	−0.4780
B3LYP/6-311G	−0.5063	−0.5625	−0.4950	−0.4410

geometries and energy surfaces of some monosaccharides.[14,15] Csonka found that acceptability of the carbohydrate structure heavily depended on the kind of basis set used.[11] The conformations of α-(2→3)-linked and α-(2→6)-linked sialic acid disaccharides were traced with MD simulations using these different partial charge values.

From the MD simulations, we found that B3LYP/6-31 was the most suitable basis set for the structural calculation of sialic acid disaccharides. Figure 2 presents probability density maps for conformational status of the α-(2→3)-linked and α-(2→6)-linked sialic acid disaccharides derived during MD simulations with different partial charge sets. As a reference, the experimentally observed structural range, taken from the *GlyTorsion* analysis results of *Glycosciences.de* web database (http://www.glycosciences.de), is delineated by gray squares.[16] There were 77 items of structural information for the α-(2→3)-linked and 16 α-(2→6)-linked sialoside fragment in the database, respectively. The MD-calculated conformational map for each sialic acid disaccharide was compared with the conformational spaces of glycosidic linkages in the database. The glycosidic dihedral angle of both disaccharides from MD calculations reasonably reflected experimentally observed values when the partial charge set from B3LYP/6-31G was used (Fig. 2E and F). However, the same basis set with the HF level showed incorrect conformations for both α-(2→3)-linked and α-(2→6)-linked sialic acid disaccharides (Fig. 2A and B). None of the calculated conformational space of the sialic acid disaccharides overlapped with the experimentally determined conformational status. For a different basis set with the same level, B3LYP/STO-3G showed the correct conformation for the α-(2→6)-linked sialoside but an incorrect conformation for the α-(2→3)-linked sialoside (Fig. 2C and D). In contrast with B3LYP/6-31G, the MD simulations with other charge sets obtained from different basis sets could not reproduce all the sialic acid conformations (data not shown). Therefore, we conclude that the Mulliken atomic partial charges, extracted from the B3LYP/6-31G DFT calculation, are suitable for MD simulations of disaccharides having a negative charge.

The conformational complexity of the carbohydrates likely originates from a combination of anomeric effects, dipole–dipole interactions, and intra- and/or intermolecular hydrogen bonding[14,15], of which the contribution of intramolecular hydrogen bonding is believed to be the most deterministic factor. In order to describe the conformation of biological molecules accurately, a proper atomic charge set must be formulated for effective electrostatic interactions.[15,17] These electrostatic effects are very important for determining hydrogen bonds between carbohydrate residues, especially for electronically charged molecules. In this study, we used Mulliken charge sets to obtain a point charge of sialo-disaccharides for the MD simulations. In the case of a Mulliken charge, electronic density is shared equally between two atoms while the electrostatic potential (ESP) charge is evaluated at four different shells around the molecular surface.[18,19] Although the ESP or RESP charge model is known to successfully describe molecular dipole and interaction energy in quantum mechanics, this Mulliken charge set also provided reasonable molecular parameters for the MD simulations. Our MD simulations demonstrated that a systematic charge value approach would provide valuable information to research fields such as structural glyco-virology and/or structure-based anti-influenza drug design.

1. Experimental

1.1. Computational methods

All ab initio charge calculations of sialic acid disaccharides were carried out with the GAUSSIAN03 program. The semi-empirical (AM1, PM3), Hartree–Forck (HF), and DFT Becke three-parameter

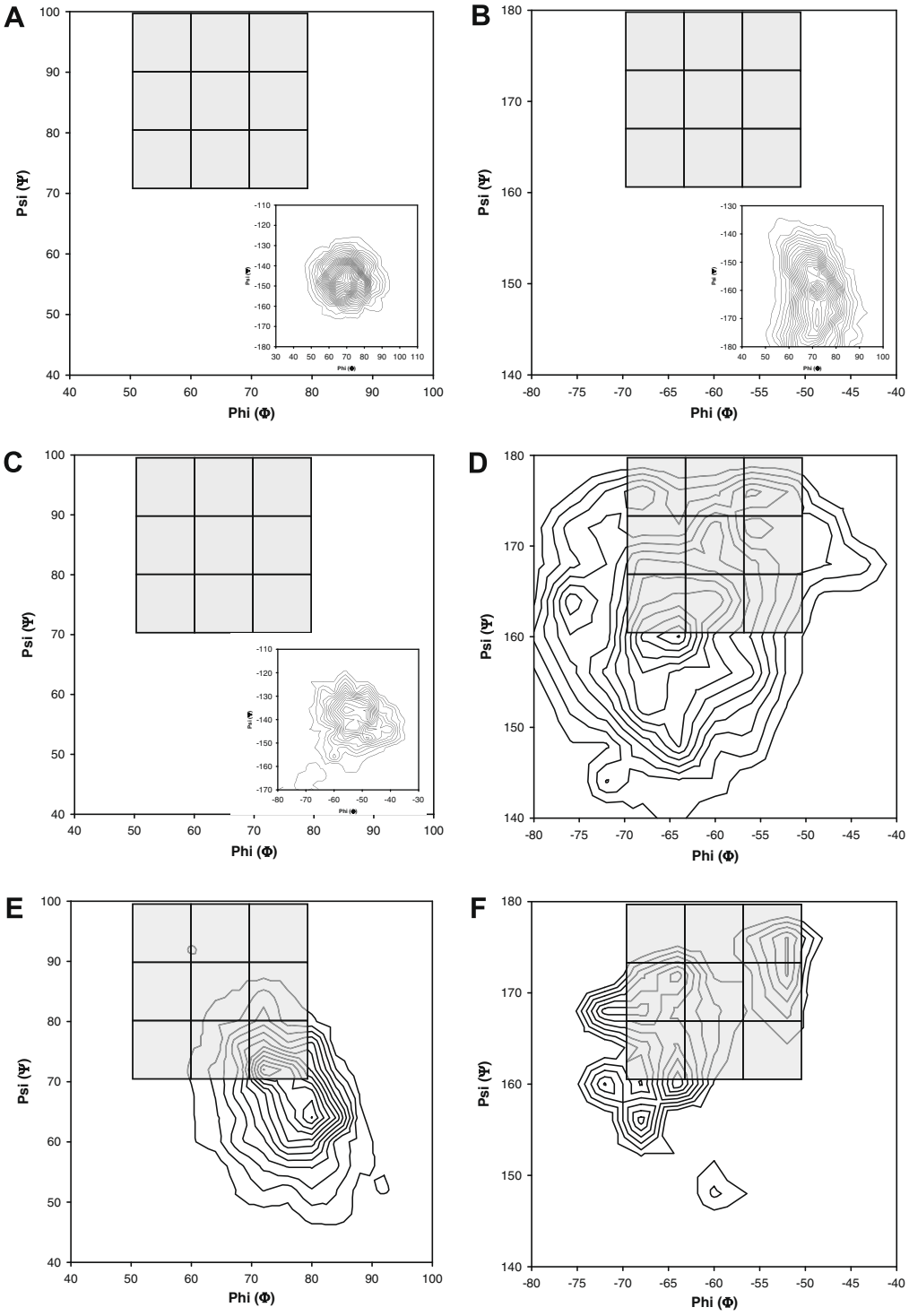

Figure 2. MD-conformational map for Sialyl-α-(2→3)-Gal (A, C, E) and Sialyl-α-(2→6)-Gal (B, D, F) from the MD simulations. Partial charge sets from HF/6-31G (A, B), B3LYP/STO-3G (C, D), and B3LYP/6-31G (E, F) were used to construct the MD-conformational map. Each gray square lattice indicates the experimentally observed range for the sialic acid disaccharides.

Lee–Yang–Parr (B3LYP) methods were used for geometry optimization with different basis sets including STO-3G, 3-21G, 6-31G, and 6-311G. The optimized structures were then used for single-point charge calculations.

A series of MD simulations of sialic acid disaccharides were performed using the CHARMM 30b1 program in the isothermal–isobaric ensemble (P = 1 bar, T = 300 K). All-atom CHARMM22 modified with a carbohydrate solution force field (CSFF)[17] was used to calculate dynamics and conformations of the carbohydrate molecules. MD simulations were performed at 300 K with a different atomic charge set from quantum calculations. A TIP3P three-site rigid water model[20] was used to solvate each α-(2→3)-linked or α-(2→6)-linked sialic acid disaccharide. Water molecules were removed if they were closer than 2.8 Å to any heavy atoms of the carbohydrate. Sodium counterion was added to maintain electric neutrality. In summary, each system was constructed using peri-

odic boundary conditions with a cubic box of dimensions 40 Å × 40 Å × 40 Å, consisting of the sialic acid disaccharides, counterions, and water molecules. The system was minimized by 1000 steps of conjugate gradient, followed by the Adopted Basis Newton–Raphson until the root-mean-square gradient was less than 0.001 kcal/mol.

The particle mesh Ewald summation[21] method was used to treat the long-range electrostatic interactions. The number of grid points for charge mesh and the kappa value was chosen to be 48 and 0.347, respectively. The bond lengths of water and each disaccharide were constrained with the SHAKE algorithm.[22] The time step was 2.0 fs, and the non-bonded pair list was updated every 50 fs. The non-bonded interactions were truncated with a 13-Å cutoff. A scaling factor of 1.0 was used for both 1–4 electrostatic and non-bonded interactions. The temperature and pressure of the system were regulated using the Langevin piston method in conjunction with Hoover's thermostat.[23] The system was gradually heated to 300 K for 80 ps and equilibrated for 500 ps at this temperature. The production MD trajectory with one snapshot per 5 ps was collected for 20 ns. For stable conformational sampling, the initial 4-ns of snapshots was discarded from the 20-ns of each MD trajectory judged by RMSD estimation.

All the dynamic simulations were performed on a computational grid system called MGrid (http://www.mgrid.or.kr).[24] Final MD trajectories were analyzed and processed using the Glyco-MGrid system in order to process a large number of MM calculations simultaneously. The MGrid system was designed to support remote execution, file sharing, and a standard web-interface to molecular simulation software to run successful MD simulations.

Acknowledgements

This study was supported by a Grant of the *e*-Science Project of KISTI (Korea Institute of Science and Technology Information) in MOST (Ministry of Science and Technology) and partly supported by KOSEF (R01-2008-000-12304-0) in 2008, SDG.

References

1. Varki, A.; Cummings, R.; Esko, J.; Freeze, H.; Harth, G.; Marth, J. *Essentials of Glycobiology*; Cold Spring Harbor Laboratory Press: New York, 1999.
2. Roseman, S. *J. Biol. Chem.* **2001**, *276*, 41527–41542.
3. Woods, R. J. *Glycoconjugate J.* **1998**, *15*, 209–216.
4. Imberty, A.; Perez, S. *Chem. Rev.* **2000**, *100*, 4567–4588.
5. Subbarao, K.; Joseph, T. *Nat. Rev. Immunol.* **2007**, *7*, 267–278.
6. Connor, R. J.; Kawaoka, Y.; Webster, R. G.; Paulson, J. C. *Virology* **1994**, *205*, 17–23.
7. Ito, T.; Couceiro, J. N.; Kelm, S.; Baum, L. G.; Krauss, S.; Castrucci, M. R.; Donatelli, I.; Kida, H.; Paulson, J. C.; Webster, R. G. *J. Virol.* **1998**, *72*, 7367–7373.
8. Ha, Y.; Stevens, D. J.; Skehel, J. J.; Wiley, D. C. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 11181–11186.
9. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.
10. Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.
11. Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
12. Basma, M.; Sundara, S.; Calgan, D.; Vernali, T.; Woods, R. J. *J. Comput. Chem.* **2001**, *22*, 1125–1137.
13. Martin, F.; Zipse, H. *J. Comput. Chem.* **2005**, *26*, 97–105.
14. Momany, F. A.; Appell, M.; Strati, G.; Willett, J. L. *Carbohydr. Res.* **2004**, *25*, 553–567.
15. Csonka, G. I. *J. Mol. Struct. THEOCHEM* **2002**, *584*, 1–4.
16. Toukach, P.; Joshi, H. J.; Ranzinger, R.; Knirel, Y.; von der Lieth, C. W. *Nucl. Acids Res.* **2007**, *35*, 280–286.
17. Naidoo, K. J.; Kuttel, M. *J. Comput. Chem.* **2001**, *22*, 445–456.
18. Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269–10280.
19. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. *J. Am. Chem. Soc.* **1993**, *115*, 9620–9631.
20. Jorgensen, W. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
21. Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
22. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.
23. Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. *J. Chem. Phys.* **1995**, *103*, 4613–4621.
24. Choi, Y.; Park, S.; Jeong, K.; Jung, S. *Bull. Korean Chem. Soc.* **2007**, *28*, 1811–1814.