ELSEVIER

Carbohydrate Research 284 (1996) 25–34

CARBOHYDRATE RESEARCH

# Conformational analysis of two glycoproteins: a Monte Carlo simulated annealing approach using a soft-sphere potential

Hongyu Zhang, Yuting Yang, Luhua Lai *, Yougi Tang

*Institute of Physical Chemistry, Peking University, Beijing 100871, People's Republic of China*

Received 2 August 1995; accepted 20 December 1995

**Abstract**

The Monte Carlo simulated annealing method was effectively used to predict the three-dimensional structure of the carbohydrate part of two glycoproteins: 1vsg and 2fbj from a protein data bank, utilizing a soft-sphere potential. The result was compared both to the crystal structure and to the structure of the corresponding isolated oligosaccharide structure simulated using an ECEPP/2 force field. A good agreement with crystal structure was reached. The interaction with the protein environment was found to significantly influence the structure of the carbohydrate moiety.

*Keywords:* Glycoprotein; Monte Carlo; Simulated annealing; Conformation search

## 1. Introduction

There are two main themes dominating the study of structural glycobiology: flexibility and diversity, which endow oligosaccharides with the properties that define some unique functional niches. The biological activity of glycoproteins is often expressed by the oligosaccharide part. For example, the binding of the envelope glycoprotein from the human immunodeficiency virus (HIV) to the cell-differentiation-marker-4-receptor (CD4) is dependent on the oligosaccharide chains of the 120 kDa envelope glycoprotein [1]. These two characteristics — flexibility and diversity — however, make it difficult for experimental work to give a stable image of a three-dimensional (3D) structure, even for very small oligosaccharides, although powerful techniques such as X-ray crystallography and NMR spectroscopy have made great advances to date. Thus, one is forced to rely

* Corresponding author.


*SSDI* S0008-6215(96)00003-1

heavily on computational methods in order to arrive at some understanding of the structure and ultimately the dynamics of oligosaccharides [2].

Knowledge-based modeling and force-field computing can provide a large amount of information on the prediction of the 3D structure of proteins, while there is little experimental data in the literature that describes the 3D structures of glycoproteins [3–8], and even less of theoretical treatment compared to those of either the protein or the carbohydrate components.

The traditional force field can be used to establish potential energy surfaces for carbohydrates, provided that an appropriate parameterization has been made, as is the case with CHARMm, etc. [9]. Some other force fields have also been developed especially for carbohydrates, such as HSEA [10], yet there are still some uncertain factors making it difficult for them to correctly describe the detailed contour of a potential surface, especially when considering long-range forces and solvent effects. For example, how to adopt a proper value for the dielectric constant of carbohydrate chains, which in most cases is located on the surface of biopolymers and is often surrounded by both solute (dielectric constant: $\sim 4$–$6$) and solvent (dielectric constant: $\sim 80$) atoms, is a major problem. Actually, some workers have neglected the electrostatic interaction during modeling and have obtained reasonable results [11].

Prohibited by lack of computer power, most work hitherto has considered only a simple glycopeptide system, which can only mimic the bulk system of the glycoprotein as a more or less vague outline. Stuike-Prill and Meyer [11] studied the energetically favored conformation of a glycopeptide that is a part of the Fc fragment of $IgG_1$ utilizing a combination of a force field of HSEA for the carbohydrate moiety and ECEPP/2 [12] for the peptide part. The calculation started from the combination of the preferred conformations of the individual components, then the total structures were optimized to get the final conformations. They found that the peptide and oligosaccharide part of the glycopeptide influenced each other's conformation, but no 3D structures of the carbohydrate part were observed that differed in their overall shape from the conformations calculated for the isolated oligosaccharide. Because no global conformation search was performed, it is not possible to tell from their results what the final global energy minimum would be. Also, if we wanted to predict the structure of the saccharide chain in a glycoprotein without any prior knowledge about it, such as that available in X-ray or NMR data, more realistic modeling methods would have to be developed.

Early calculations employing only the hard-sphere potential in which there is no attractive term and only an infinite repulsion when two atoms approach each other within a distance less than or equal to the sum of their van der Waals radii have been able to provide very useful, although approximate, insights into the structure of the oligopeptide chain [13]. Here we designed a so-called soft-sphere potential evaluating the steric interaction between nonbonded atom pairs by considering only the van der Waals volume of atoms, which also omitted the attractive term and static electric term. This is done to avoid artificial attractive forces between the atoms that could, in the absence of solvent molecules, lead to a biasing of conformations towards those that have internal van der Waals type attractions, i.e., conformations that show artificially high internal interactions. As it has a simple form, a great deal of computational effort could be saved in calculation, which made it possible to be used in the detailed conformation

search of a carbohydrate structure in a glycoprotein. We have compared the Monte Carlo simulated annealing (MCSA) result both to the crystal structure and to the calculated conformation of the isolated oligosaccharide with the ECEPP/2 force field, and then we have discussed how the protein environment influences the conformation of the oligosaccharide.

## 2. Experimental procedures

The intra-action of the oligosaccharide and the interaction between the oligosaccharide and the protein part of the glycoprotein were expressed by an atom-to-atom soft-sphere potential

$$E = \begin{cases} k(d_0^2 - d^2), & d_0 > d \\ 0, & d_0 < d \end{cases} \quad (1)$$

where $E$, $d$, $d_0$, and $k$ represent, respectively, the interaction energy between two atoms, the distance between them, the standard van der Waals distance between them (which is equal to the sum of the standard van der Waals radii according to the specific atomic type [14], and the force coefficient which can be properly set by the user. (It was set to be 100 kcal/(mol atom $\mathring{A}^2$) in this paper.) The interaction for sequential atoms, i.e., 1–2 and 1–3 interactions were omitted, and there is no difference in interaction between 1–4 and 1–5 or the longer ones. This type of potential will neglect the interaction between atoms when their van der Waals volumes do not overlap. Clashing is allowed, but the closer they are, the more repulsive the interaction is. In contrast to the hard-sphere model with absolute exclusion, this model is soft and can easily be used by a conformational searching method, such as Monte Carlo or molecular dynamics, to scan a large conformational space and ultimately to find the global energy minimum.

For those protein atoms having distances to the connection point that connects the carbohydrate moiety with the protein part longer than the stretch length of the carbohydrate chain, clashing with the carbohydrate moiety is impossible. Therefore, only those with shorter distances were taken to be in the environment, and consequently, a large amount of computing time was saved. As is well known, generally the limiting factor in a Monte Carlo calculation is the evaluation of potential energy. Thus this simple type of potential makes it possible to take the environment of the bulk protein into consideration in a computationally acceptable time, and then give a good estimate of the entire glycoprotein system.

Much work has been done in modeling isolated oligosaccharides [9,10,15]. Effective as the techniques are for short oligosaccharides, e.g., for disaccharides, their accuracy for longer oligosaccharides is in our experience not always clearly definitive and somewhat dependent both on the specific system and on the specific force field. As stated above, our interest is to find an acceptable potential and a suitable algorithm for treatment of a glycoprotein. For the present, we do not expect this kind of simplified potential to give more accurate information for an isolated saccharide than those that have been carefully done previously. Therefore, in this paper, the empirical energy

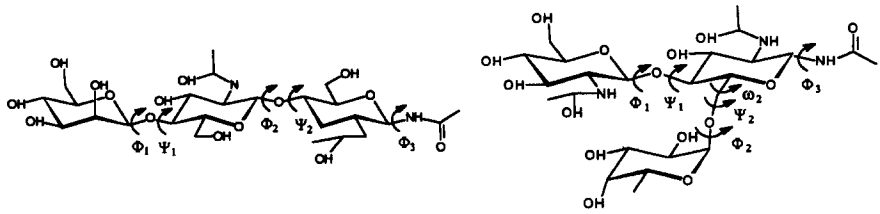
ECEPP/2 was used for modeling the isolated oligosaccharide, which was demonstrated to be able to give a better agreement of conformational properties with those derived from experiment than HSEA [15].

In a typical MCSA run, all the protein atoms were kept fixed in the same coordinates as in the crystal structure. The initial conformation of the saccharide had all the variable torsion angles, including glycosidic angles and side-chain angles, set to 180°, and the pyranose rings were kept rigid with the same conformation as that determined in the crystal structure. Conformations are sampled by randomly picking one of the coordinates and assigning a new value between $-180°$ and $+180°$, where the coordinates refer to all the variable torsion angles. The total energy of the new conformation, which consists of nonbonded soft-sphere intra-action within the saccharide moiety and nonbonded soft-sphere interaction between the saccharide moiety and the protein environment, was evaluated in each step and compared to the prior one, after which the new conformation was either accepted or rejected based on the Metropolis criteria [16]. A circle in the simulation was completed after a specified number of conformations were evaluated, where the specified conformation searching number was set to be 100 times the number of the variable torsion angles, and the accepting rate of each circle was equal to the number of accepted conformation divided by the conformation searching number. The simulation starts from a high temperature where a high accepted rate (80% in our work) could be achieved, and the temperature was lowered at the end of each circle by multiplying a scale factor of 0.83. When the temperature dropped to a value near zero, or, after a specified number of circles, the last circle of the simulation was executed. The uniform random number generator GGL [17], which is based on a linear congruential method, was adopted in our MCSA, and all the algorithm was implemented in $C^{++}$ code.

Two systems were selected to test our method: 1vsg and 2fbj from the protein data bank (PDB). Glycoprotein 1vsg is the variant surface glycoprotein from *Trypanosoma brucei* and has a resolution 2.9 Å. The crystal structure is a dimer and the carbohydrate of chain A was selected in the study, which has a linear trisaccharide *N*-linked to Asn263. Glycoprotein 2fbj is the Fab segment of the galactan-binding immunoglobulin J539 with a resolution of 1.95 Å, which has a branched trisaccharide *N*-linked to Asn156 of the heavy chain. The structure of the two carbohydrate chains are illustrated in Fig. 1.

In order to compare the crystal structure with the calculated result, the two crystal structures were optimized with 100 steps of ABNR methods using CHARMm (QUANTA 4.0) in order to remove bad contacts, during which time the protein part was kept fixed and the carbohydrate was constrained by a harmonic force of 5 kcal/(mol atom Å). The minimized crystal structures had RMSDs from their initial conformations equal to 0.04 and 0.11 Å for 1vsg and 2fbj, respectively.

For each system, 100 conformations were generated, and the results were analyzed by means of clustering. The procedures of clustering can be briefly described as follows: (i) generate the torsion angles' RMSDs (root-mean-square deviations) of each conformation to all the others, which can be considered to be the distances in internal coordinate space that is defined by those glycosidic torsion angles shown in Fig. 1; then, (ii) given a specified cut-off value, pick out the representative conformation which has the largest



β-D-Man-(1→4)-β-D-GlcNAc-(1→4)-β-D-GlcNAc-N-Asn

β-D-NAG-(1→4)
α-D-FUC-(1→6)
β-D-NAG-N-Asn

**a** **b**

Fig. 1. Chemical structure of carbohydrate chains for glycoproteins. (a) 1vsg; (b) 2fbj. The torsion angles of the oligosaccharide are defined as $\phi$ (C-2–C-1–O-4–C-4) and $\psi$ (C-1–O-4–C-4–C-5), or in the case of a (1 → 6)-glycosidic linkage, as $\phi$ (C-2–C-1–O-6–C-6), $\psi$ (C-1–O-6–C-6–C-5) and $\omega$ (O-6–C-6–C-5–C-4), and for glycosyl residue to Asn, as $\phi$ (C-2–C-1–ND–CG). Here, all atom names are the same as those in the PDB file. These definitions are, in certain respects, different from what are commonly used in other references, because there are no hydrogen atoms in this simulation.

number of neighbors in conformational space, and put it, as well as its neighbors, into the first cluster. After deleting the first cluster from the conformational space, the second step is repeated to find the rest of the clusters.

## 3. Results and discussion

*Glycoprotein 1vsg.*—The 100 resultant conformations of the glycoprotein were clustered according to RMSDs from each other, as listed in Table 1. One can see that, based on our simple type potential, the MCSA calculation almost always resulted in conformations close to the crystal structure. Actually, with a small cut-off of 7.0°, most samples (90%) fell into the main group and had low RMSDs (around 17°) from the crystal structure, as seen in Fig. 2a. Also, the standard deviation of glycoside torsion angles within a cluster is very small. This results, as we hoped, demonstrated the effectiveness of our potential.

For the isolated carbohydrate simulated with the ECEPP/2 force field, the resultant conformations could be grouped even better with a smaller cut-off; however, they deviated much more from the crystal structure than did the glycoprotein samples (all RMSDs around 88°). It can be seen both from Table 1 and from Fig. 2b that all the angles were in similar conformations to the crystal except the $\psi_2$ angle, which adopted a reverse conformation (rotating 180°) from the real angle. This indicated the difficulty in modeling the conformation of the saccharide moiety in a glycoprotein without including the protein environment.

*Glycoprotein 2fbj.*—Similar steps were taken to study the system of 2fbj. It can be seen from Table 2 that, in contrast to cases in 1vsg, conformations of both glycoprotein

Table 1
Cluster of 100 simulated conformations of glycoprotein 1vsg

| | Density [a] (%) | Representative conformation/standard deviation in a cluster (°) [b] | | | | | RMSD [c] (°) |
|---|---|---|---|---|---|---|---|
| | | $\phi_1$ | $\psi_1$ | $\phi_2$ | $\psi_2$ | $\phi_3$ | |
| Glycoprotein saccharide | 90 | 113.0/4.0 | −120.4/5.1 | 169.6/2.7 | −114.1/2.5 | 151.3/1.6 | 17.1 |
| (cut-off: 7.0°) | 6 | 109.8/5.5 | −107.9/5.0 | 160.6/3.8 | −118.1/3.4 | 150.0/2.1 | 22.3 |
| Isolated saccharide | 92 | 81.6/2.0 | −148.5/1.3 | 143.1/1.5 | 73.0/0.8 | | 88.2 |
| (cut-off: 4.0°) | 3 | 85.4/1.7 | −150.2/0.3 | 150.4/2.7 | 75.6/3.0 | | 86.8 |
| Crystal structure | | 98.2 | −155.0 | 166.9 | −113.2 | 155.6 | |

[a] Equal to the number of samples falling in a cluster divided by the total simulation number, which is 100 in our work.
[b] See Experimental procedures for definition of representative conformation and Fig. 1 for definition of torsion angles.
[c] Root-mean-square deviation of representative conformation from crystal structure.

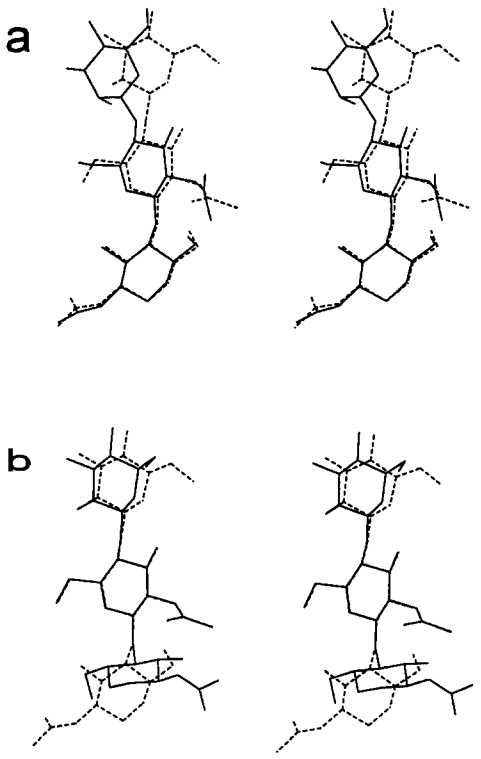



Fig. 2. Stereo plots of stick models of glycoprotein 1vsg. (a) Crystal structure and simulated glycoprotein saccharide; (b) crystal structure and simulated isolated saccharide; (---) crystal structure, (————) simulated structure.

and isolated oligosaccharide are diverse. The former were not able to group well until a large cut-off value of 35° was tested, and only a small group of samples (8%) were close to the crystal structure. The latter, although being able to group well with a cut-off 15°, deviated considerably from the crystal structure. Also, we can see that all of the clusters have larger standard deviations than those of 1vsg within clusters.

By carefully examining Table 2, we found that, although the representative conformations of main groups of glycoprotein saccharide have large deviations from the crystal structure, it is very interesting that most torsion angles are, either close to the real angle, or 180° in the reverse to that of the crystal structure. For example, the conformations in the first two clusters have the angle of glycosyl residue to Asn, i.e., $\phi_3$ in Fig. 1, in the reverse conformation, which means that all the pyranose ring faces of the saccharide are rotated 180°. This phenomenon, however, was less obvious for the isolated saccharide except for the $\psi_1$ angle.

Since only a small population of correct conformations was obtained for 2fbj, we tried to find if it was possible to select these conformations from all the others using energy criteria. Fig. 3, however, does not show a good correspondence of energy vs. the 3D coordinates RMSD, and only shows a simple fact, i.e., that the conformations with

Table 2
Cluster of 100 simulated conformations of glycoprotein 2fbj

| | Density [a] (%) | Representative conformation/standard deviation in a cluster (°) [b] | | | | | | RMSD [c] (°) |
|---|---|---|---|---|---|---|---|---|
| | | $\phi_1$ | $\psi_1$ | $\phi_2$ | $\psi_2$ | $\omega_2$ | $\phi_3$ | |
| Glycoprotein saccharide | 77 | 112.2/19.0 | −136.0/14.3 | 96.0/9.0 | 173.8/152.3 | −114.7/22.8 | 115.5/18.7 | 69.4 |
| (cutt-off: 35.0°) | 10 | 148.4/9.0 | −127.6/9.1 | −94.8/3.1 | −111.0/17.6 | −106.6/5.9 | 87.2/5.8 | 99.2 |
| | 8 | 149.5/5.5 | −127.8/12.8 | 100.0/6.7 | −128.8/102.1 | −134.9/16.9 | −92.2/1.9 | 22.1 |
| Isolated saccharide | 48 | 26.0/19.6 | −134.4/6.9 | −6.2/7.5 | 102.6/3.0 | 44.6/2.15 | | 115.9 |
| (cutt-off: 15.0°) | 29 | 27.5/5.8 | −132.7/1.9 | 161.9/3.1 | −111.8/2.4 | 127.8/1.6 | | 93.5 |
| | 18 | 26.2/12.4 | −131.8/4.6 | −74.8/1.8 | −81.0/4.6 | 95.9/1.9 | | 130.2 |
| Crystal structure | | 173.4 | −135.2 | 110.3 | −162.9 | −103.2 | −90.1 | |

[a] Equals to the number of samples falling in a cluster divided by the total simulation number, which is 100 in our work.
[b] See Experimental procedures for definition of representative conformation and Fig. 1 for definition of torsion angles.
[c] Root-mean-square deviation of representative conformation from crystal structure.

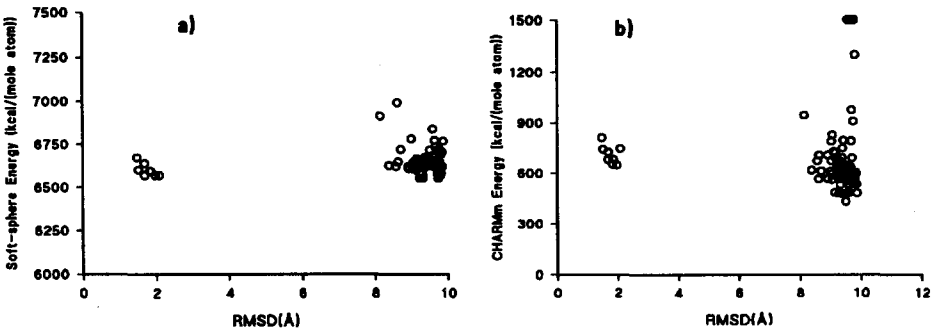



Fig. 3. Conformation energy vs. 3D coordinates RMSD from crystal structure of glycoprotein 2fbj. (a) Soft-sphere energy vs. RMSD; (b) CHARMm energy vs. RMSD.

very high energies have a large deviation. We have also tried minimizing the resultant conformation using CHARMm, then redrawing the figure (data not shown here), but no better results were obtained. Since our soft-sphere potential is no worse than the more explicit force field in selecting efficiency, we reasoned that the small population of correct conformations may be caused by some specific factors in the different systems instead of being due to a problem in the force field. It is known that (1 → 6)-glycosidically linked branches have higher flexibility than other kinds, which, however, may not explain the big difference between the simulated results of these two proteins. Original PDB files were checked, and it was found that the temperature factor of the carbohydrate part of 2fbj was much larger (around 60 $Å^2$) than that of 1vsg (around 20 $Å^2$). As is well known, the larger the temperature factor, the more flexible the chain. Furthermore, it was found that there are water molecules around the saccharide residues in the crystal structure of 2fbj, and many of them form hydrogen bonds with the carbohydrate. In our simulation, all the solvent molecules were removed, and only solute atoms were considered, which may explain why a large number of simulated conformations of 2fbj failed to overlap with the crystal structure. Most of the oligosaccharide residues in 1vsg, however, pack against the surface of the protein in a pocket between two helices as described in the original ref. [8], which may account for their being well-ordered in the crystal.

So generally, in both systems, conformations of the isolated saccharide simulated using an ECEPP force field converged well, but failed to conform to that of the crystal structure, which may reflect the fact that, although progress has been made in recent years in saccharide simulation using an empirical force field, it is difficult to model the conformation of the glycoprotein saccharide without including the protein part.

It can be seen from our work that the conformational space of an oligosaccharide in a glycoprotein can be efficiently searched using a simplified potential combined with the MCSA algorithm. In fact, the correct conformations of both systems were finally determined. In the case of 2fbj there is a defect, and one is unable to distinguish the correct conformation from all the others, which may be attributed to the factors in the specific environment. Even when using a time-consuming, if not computationally

impossible, explicit force field in MCSA, it is probable that no better result can be obtained, a fact that has been implied in the selection failure using CHARMm.

## Acknowledgements

The authors would like to thank Prof. Kankaala of Tampere University of Technology of Finland for providing the random number generator code. This work was supported by the Chinese State Commission of Science and Technology and the State Commission of Education.

## References

[1] H. Lis and N. Sharon, *Annu. Rev. Biochem.*, 55 (1986) 35–67.
[2] S. Pérez. *Curr. Opin. Struct. Biol.*, 3 (1993) 675–680.
[3] J. Deisenhofer, *Biochemistry*, 20 (1981) 2361–2370.
[4] K.A. Wilson, J.J. Skchel, and D.C. Wiley, *Nature*, 289 (1981) 366–373.
[5] W. Bode, E. Meyer, and J.C. Powers, *Biochemistry*, 28 (1989) 1951–1963.
[6] B.J. Sutton and D.C. Philips, *Biochem. Soc. Trans.*, 11 (1983) 130–132.
[7] S.W. Suh, T.N. Bhat, M.A. Navia, G.H. Cohen, C.N. Rao, S. Rudikoff, and D.R. Davies, *Proteins: Struct. Func. Genet.*, 1 (1986) 74–80.
[8] D. Freymann, J. Down, M. Carrington, I. Roditi, M. Turner, and D. Wiley, *J. Mol. Biol.*, 216 (1990) 141–160.
[9] S.N. Ha, A. Gammona, M. Field, and J.W. Brady, *Carbohydr. Res.*, 180 (1988) 207–221.
[10] R.U. Lemieux, K. Bock, T.J. Delbaere, S. Koto, and V.S. Rao, *Can. J. Chem.* 58 (1980) 631–639.
[11] R. Stuike-Prill and B. Meyer, *Eur. J. Biochem.*, 194 (1990) 903–919.
[12] G. Nemethy, M.S. Potte, and L. Scheraga, *J. Phys. Chem.*, 87 (1983) 1883–1887.
[13] L. Scheraga, in K.B. Lipkowitz and D.B. Boyd (Eds.), *Reviews in Computational Chemistry*, Vol. 3, VCH, New York, 1992, pp. 73–142.
[14] A. Bondi, *J. Phys. Chem.*, 68 (1964) 441–451.
[15] A.J. Duben, M. Hricovini, and I. Tvaroska, *Carbohydr. Res.*, 247 (1993) 71–81.
[16] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E.J. Teller, *Chem. Phys.*, 21 (1953) 1087–1092.
[17] P. Lewis, A. Goodman, and J. Miller, *IBM Sys. J.*, 2 (1969) 136–159.