Appl Biochem Biotechnol (2009) 159:343–354
DOI 10.1007/s12010-008-8518-3

Exploring the C–H…O Interactions in Glycoproteins

Sudha Anand · Anand Anbarasu · Rao Sethumadhavan

Received: 28 August 2008 / Accepted: 29 December 2008 /
Published online: 21 January 2009

Abstract Glycoproteins are an important class of proteins that play a significant role in many cellular events. In the present study, we analyze the influence of C–H…O interactions in relation to other environmental preferences in glycoproteins. CH…O interactions are now accepted as a genuine hydrogen bond. Main chain–main chain interactions are predominant. Proline residues stabilize strands by C–H…O interactions in glycoproteins. Majority of the C–H…O interacting residues were conserved and had one or more stabilization centers. CH…O interactions might be responsible for the global conformational stability, since long-range CH…O contacts were predominant. The results presented in this study might be useful for structural stability studies in glycoproteins.

Keywords Glycoproteins · C–H…O interactions · Stabilization centers · Conservation score

Introduction

Glycoproteins are increasingly being discovered to play salient roles in biological carbohydrate–protein recognition systems [1]. Glycoproteins are involved in various cellular events such as cell–cell recognition, adhesion, growth regulation, inter- and intracellular routing, neurodevelopment, protection against neurodegeneration, and synaptic plasticity [2–5]. Misfolding of normal cellular glycoproteins to pathogenic and protease resistant isoforms is the hallmark of prion diseases [6]. Glycoproteins play a major role in primary and secondary homeostasis [7, 8]. Glycoproteins are considered as very sensitive antigenic markers for the immunodiagnostic tools such as enzyme linked immunosorbent assays [9]. Antigenic variation, a very important immune evasion strategy among microorganisms, is brought about by changes in the expression of surface glycoproteins [10, 11]. Glycoproteins are also involved in limiting the entry of potentially harmful

S. Anand · A. Anbarasu · R. Sethumadhavan (✉)
Bioinformatics & Computational Biology Division, School of Biotechnology,
Chemical & Biomedical Engineering, Vellore Institute of Technology, Vellore 632014, India
e-mail: aanand@vit.ac.in

substances into the brain [12]. There a number of very significant biological events involving glycoproteins.

CH…O interactions are now accepted as a genuine hydrogen bond [13, 14]. C–H…O interactions play a significant role in the stability of a large number of biomolecular complexes and crystal structures [15–17]. C–H…O interactions are now understood to be present in many biological systems, including proteins [18–22], nucleic acids [23–29], and carbohydrates [30]. C–H…O interactions have also been implicated in the interaction of nucleic acids with proteins [31, 32], drug binding [33–35], the stability and specificity of trans-membrane helices [36], the mechanism of serine protease catalysis [37], crystal engineering [38–40], and super-molecular assemblies [16].

Even though there are a lot of information available in the literature regarding the functional, microbiological, and molecular biology aspects of glycoproteins, there are no reports on computational aspects of glycoproteins in particular C–H…O interactions. Our group has recently reported the influence of non-canonical interactions in the structural stability of RNA-binding proteins [41–43], interleukins [44–46], and comparisons on the role of cation–π interactions in RNA binding, lipid binding, and glycoproteins [47]. We think that this report will be a good continuation and supplement to our earlier articles on non-canonical interactions. Ours is probably the first such attempt on computational analysis of C–H…O interactions in glycoproteins, and our results will be useful for further investigations on this class of proteins, which is diverse in its functions.

Materials and Methods

Data Set

We have selected all available non-homologous crystal structures of glycoproteins from the protein data bank (PDB) [48] for our investigation. The sequence identity among the proteins in the dataset was less than 40%, and the three dimensional structures of these proteins have been solved with ≤2.5 A° resolution. We have excluded homologous glycoproteins from the study to avoid redundancy. The PDB IDs of the glycoproteins investigated in the present study are as follows: 1DYL-A, 1G9M-G, 1HNG-A, 1JND-A, 1UZG-A, 1WCQ-A, 1X3W-A, 1YUK-B, 1Z8Y-A, 2F40-A, 2F4M-A and 2VSG-A.

C–H…O H-bonds

We used a stand-alone program to identify the C–H…O interactions [49]. This program uses four different geometric criteria that are defined based on distances and angles of the atoms under consideration, which is depicted in Fig. 1. The default distance parameters between C…O=C and H…O=C are ≤3.8 and ≤3.3 A°, respectively. The default parameters for the angles C–H…O and H…O=C are ≥120 and ≥903 A°, respectively. The program uses ‘REDUCE’ [50] to fix hydrogen atoms and ‘matrix2png’ [51] to create color coded interaction matrix. The coding for the program is done in PERL and the web interface is by CGI-PERL.

C–H…O interaction types are represented by a two-letter code in which the first letter indicates the donor atom and the second the acceptor: M and S represent the main- and side-chain atoms, respectively. C–H…O interactions are classified into four types, namely,

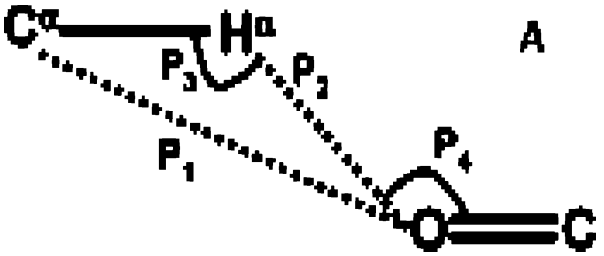

Fig. 1 Geometric criteria used to identify C–H…O interactions [49]

main- to main-chain C–H…O interactions, main- to side-chain C–H…O interactions, side- to main-chain C–H…O interactions, and side- to side-chain C–H…O interactions, which are abbreviated as MM C–H…O, MS C–H...O, SM C–H…O, and SS C–H…O, respectively.

Secondary Structure and Solvent Accessibility

The structure and function of proteins is determined by two important intermediate factors. They are secondary structure preference and solvent accessibility patterns. In order to obtain the preference and pattern of each C–H…O interaction-forming residue in glycoproteins, we conducted a systematic and careful analysis based on their location in different secondary structures and their solvent accessibility. We obtained the information about secondary structures and solvent accessibility of the proteins using the program DSSP [52]. Solvent accessibility was divided into three classes: buried, partially exposed, and exposed indicating respectively the least, moderate, and high accessibility of the amino acid residues to the solvent [53, 54].

Sequential Distance

It will be very useful to evaluate the contribution of short, medium, and long-range contacts in the formation of C–H…O interactions. The C–H…O interacting residues coming within a sphere of 8 A° was computed as described earlier [55–57]. For a given residue, the comparison of the surrounding residue is analyzed in terms of the location at the sequence level. The contribution from $<\pm 4$ are treated as short-range contacts, $>\pm 4$ to $<\pm 20$ as medium-range contacts and $>\pm 20$ are treated as long range contacts [58].

Conservation Score

Conservation score is a useful parameter for the identification of conserved residues in a protein sequence. We computed the conservation score of C–H…O interacting amino acid residues in each glycoprotein in the data set using the ConSurf program [59].

Stabilization Centers

Stabilization centers are clusters of residues that are involved in medium or long-range interactions [60]. Residues can be considered part of stabilization centers if they are involved in medium or long-range interactions and if two supporting residues can be

selected from both of their flanking tetra peptides, which, together with the central residues, form at least seven out of the nine possible contacts [61].

Results and Discussion

C–H…O H-bonds

There are four types of C–H…O interactions as described above [49]. The C–H…O interactions in glycoproteins are presented in Fig. 2. We found that 80% of the interactions were MM C–H…O interactions, 7% of the interactions were MS C–H…O interactions, 9% of the interactions were SM C–H…O interactions, and the remaining 4% interactions were SS C–H…O interactions. The major contribution to C-H...O interactions was mainly from MM C–H…O interactions and the contribution of MS and SS C–H…O interactions was minimal. From these observations, we consider the contribution of MM C–H…O interactions to be significant and hence may play an important role in determining the structural stability of glycoproteins. The MM C–H…O interaction between Asn 392 and Phe 361 in glycoprotein PDB ID 1G9M-G is shown in Fig. 3. There was an average of 70 C–H…O interactions per protein and an average of one significant C–H…O interaction for every five residues in the glycoproteins studied. The donor amino acid residues involved in C–H…O interactions are presented in Fig. 4. The contribution from aliphatic residues was predominant in MM C–H…O interaction. It might be due to the fact that apart from their higher natural occurrences, the $C_{\alpha}H$ group of aliphatic residues is directly adjacent to a pair of electronegative groups and hence serves as a relatively stronger proton donor compared to other C (sp^3) H groups that lack any substitution from electronegative groups. It is among the most prevalent C–H group involving the C–H…O interaction, and therefore, many investigations of C–H…O interactions focus on the C–H groups as donors [62]. In MS C–H…O interactions the donor atom contribution was mainly from Gly. In the case of SM C–H…O and SSC–H…O interactions the donor atom contribution was from Pro, Lys, and His. The increased contribution of Lys, Pro, and His residues in SS and SM C–H…O interactions might be due to the fact that the $C_{\varepsilon}H$ of Lys, $C_{\delta}H$ of Pro, and the $C_{\varepsilon}H$ of His participating in side-chain C–H…O H-bonding interactions [62].

The acceptor amino acid residues involved in C–H…O interactions are presented in Fig. 5. The acceptor atom contribution for MM C–H…O interactions was predominantly

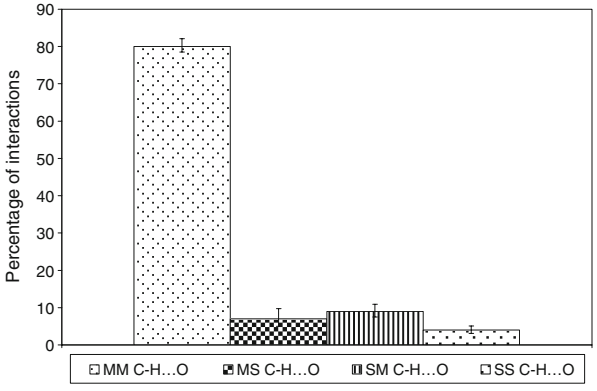

Fig. 2 C–H…O interaction types in glycoproteins

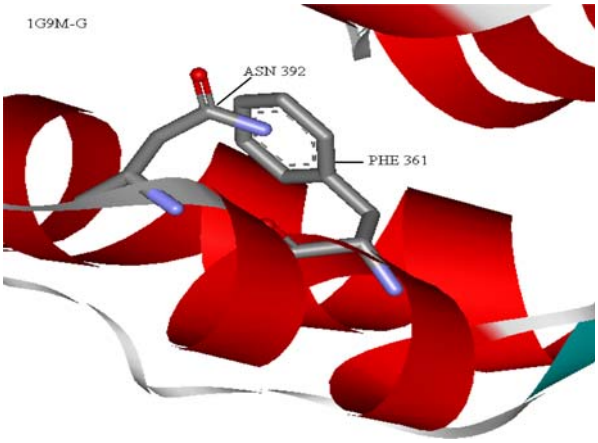

Fig. 3 PyMol view of MM C–H…O interaction in glycoprotein PDB ID 1G9M-G

from Val, Leu, Ser, Ile, Ala, and Gly residues, even though all naturally occurring amino acids were observed as acceptor residues. The contribution of acceptor atoms for MS C–H…O interactions were from Glu, Asp, Asn, and Gln residues. All the 20 amino acids contributed acceptor atoms in SM C–H…O interactions. The acceptor atom contribution for SS C–H…O interactions was only from Asp, Glu, Asn, and Gln.

All the naturally occurring amino acids had donor and acceptor atoms that were involved in C–H…O interactions. The contribution of MM C–H…O mediated interactions was higher in C–H…O interactions as compared to MS C–H…O, SM C–H…O, and SS C–H…O interactions irrespective of the amino acids involved. In terms of total donor atom

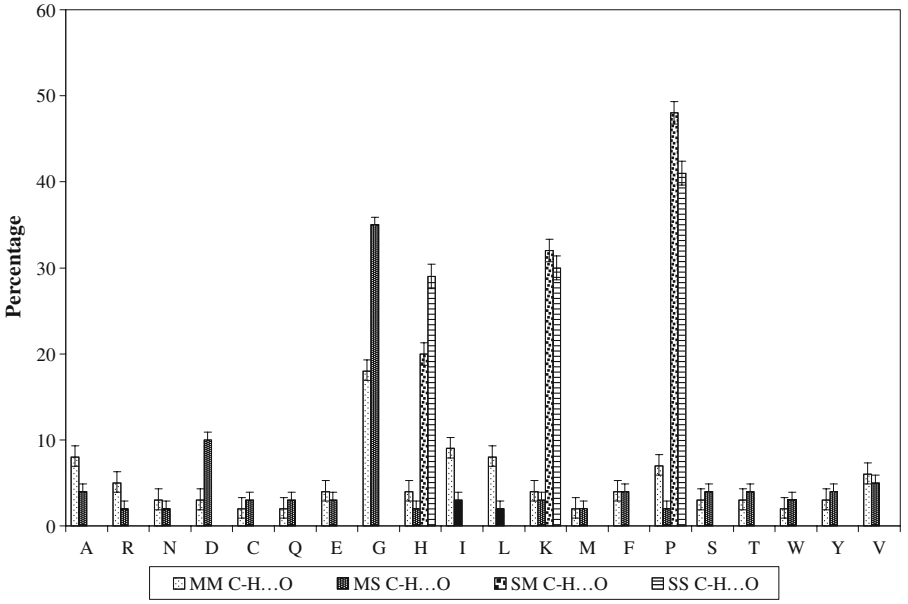

Fig. 4 Donor residues in C–H…O interactions

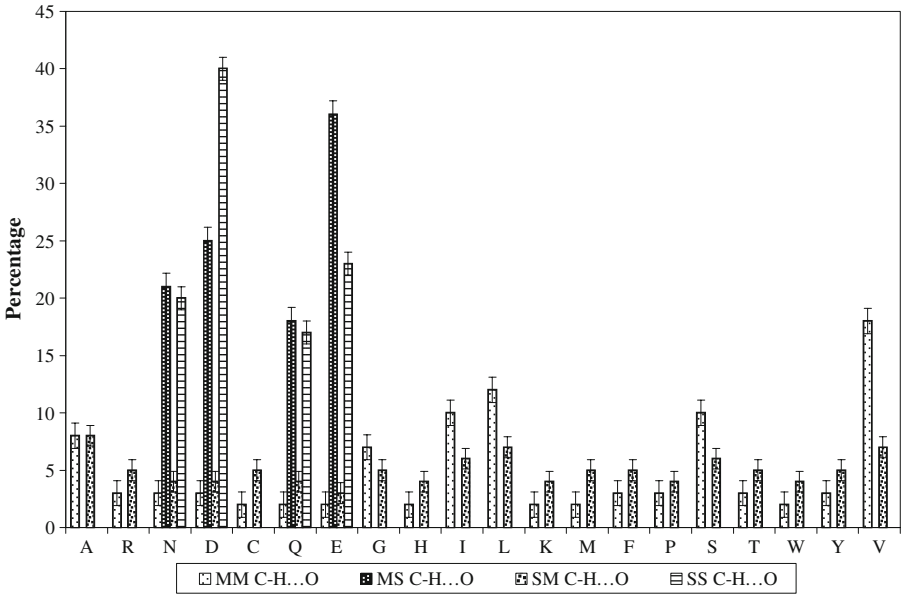

Fig. 5 Acceptor residues in C–H…O interactions

contribution to C–H…O interactions, the highest contribution of donor atoms was mainly from aliphatic residues. Pro, Lys, and His residues were involved as donor residues in all the four types of C–H…O interactions. In the case of acceptor atom contribution to C–H…O interactions, all the 20 naturally occurring amino acids were seen as acceptor residues in MM C–H…O and MS C–H…O interactions. Glu, Gln, Asp, and Asn were involved as acceptor residues in all the four types of C–H…O interactions in glycoproteins.

Secondary Structure Preferences

The occurrence of weak interactions has been observed at the terminus of the secondary structural units, in particular α-helix and β-sheets [63, 64]. These interactions play a definitive role in stabilizing the proteins. The propensity of the amino acid residues to favor a particular conformation has been well documented. Such conformational preference is not only dependent on the amino acid alone but is also dependent on the local amino acid sequence. We analyzed the secondary structure preference of each amino acid, which participated in all the four different types of C–H…O interactions. The secondary structure preferences of each of the amino acids involved in C–H…O interactions are depicted in Fig. 6. We found that, Lys, Glu, Ser, and Thr preferred to be in helix, while other amino acid residues preferred strand conformation. Pro residues preferred to be in strands and helices. Even though the secondary structure preferences of most of the amino acids involved in C–H…O interactions were consistent with the information available in literature, it was interesting to observe that a significant percentage of Pro residues favored strand and helix conformation. From this observation, we infer that Pro residues might stabilize strands and helices by C–H…O H-bonding in glycoproteins. Similar findings were also reported in globular and transmembrane proteins [20] as well as in other protein structures and complexes [65].

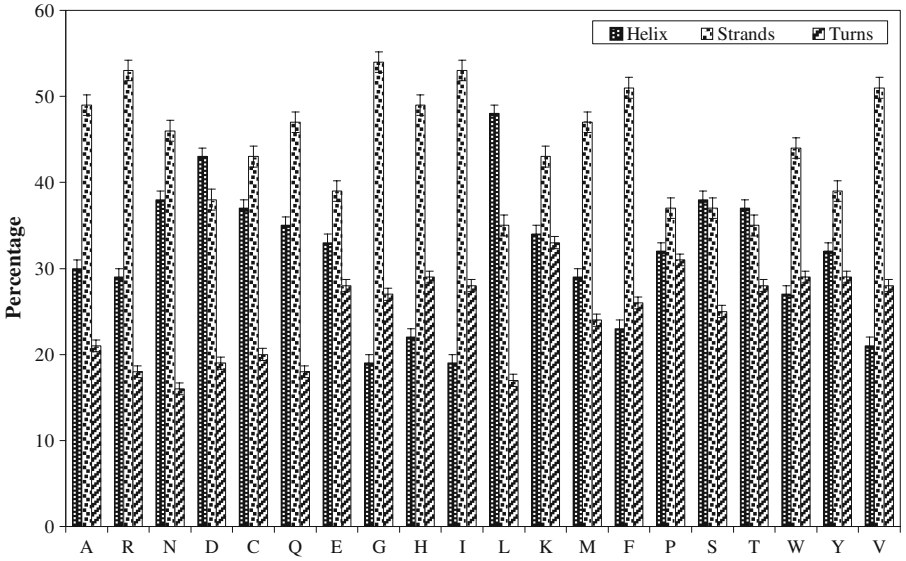

Fig. 6 Secondary structure preferences in glycoproteins

Solvent Accessibility

We estimated the solvent accessibility of the amino acid residues that are involved in C–H…O interactions. The relation between the amino acid residues in C–H…O interactions and solvent accessibility are depicted in Fig. 7. The solvent accessibility of amino acid residues has been categorized as buried, partially exposed, and exposed [53, 54]. We found that of the different amino acids that were involved in C–H…O interactions, Arg, Asn, Asp, Gln, Glu, His, Lys, Ser, Thr, Tyr, and Trp were in the exposed regions. Pro preferred to be in partially exposed region. Ala, Cys, Gly, Ile, Leu, Met, Phe, and Val residues were in the

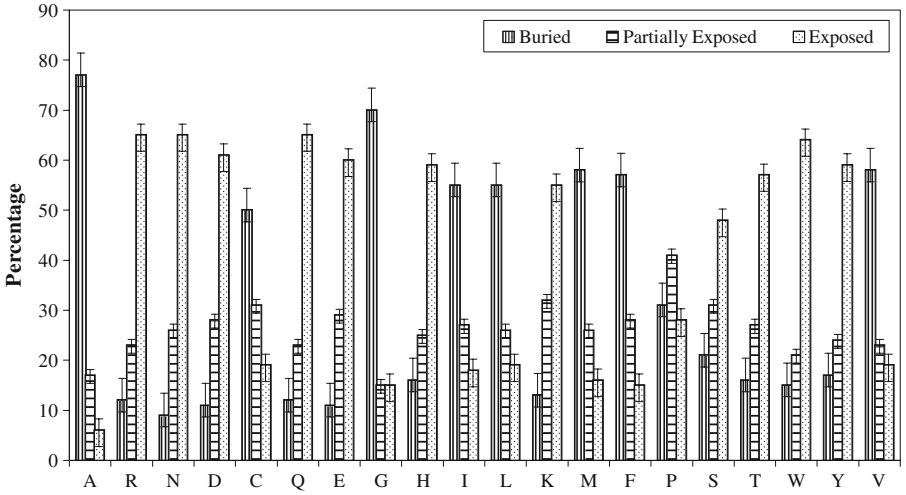

Fig. 7 Solvent accessibility in glycoproteins

buried regions. We found that most of the polar amino acid residues involved in C–H…O interactions were solvent exposed and most of the non-polar residues involved in C–H…O interactions were excluded from the solvent. Hence, the polar residues might contribute significantly to the stability of glycoproteins.

Sequential Separation

The contribution of C–H…O interactions in glycoproteins could define either the local or the global stability of the proteins. Therefore, there is a need to evaluate the contribution of C–H…O interactions. The sequential distance between residues that contributed donor and acceptor atoms to C–H…O interactions was calculated, and results are presented in Fig. 8. Of the C–H…O interactions, 69%, 17% and 14% were found to be long-, medium-, and short-range interactions. Long-range C–H…O interactions are the predominant type of interactions in glycoproteins. The contribution of short-range interactions is comparatively less in all the four types of C–H…O interactions studied. These results indicate that long-range C–H…O interactions might contribute significantly to the global conformational stability of glycoproteins.

Conservation Score

Conservation score was computed based on the comparison of the sequence of a PDB chain with the proteins deposited in Swiss-Prot [66] and finding the ones that are homologous to the PDB sequence. The number of PSI-BLAST iterations and the *E* value cutoff used in all similarity searches were 1 and 0.001, respectively. All the sequences that are evolutionarily related with each one of the proteins in the data set were used in the subsequent multiple alignments. Based on these protein sequence alignments the residues are classified into nine categories from highly variable to highly conserved. Residues with a score of 1 are considered highly variable and residues with a score of 9 are considered highly conserved. Conservation score of the amino acid residues involved in C–H…O H-bonding in glycoproteins were computed, and the scores are depicted in Fig. 9. Conservation score, ≥6 is the cutoff value used to identify the stabilizing residues. Conservation scores for the glycoprotein IDs 2F40-A and 2VSG-A could not be computed, as there were not many homologs. Majority of residues involved in C–H…O H-bonding had a higher conservation score, and so these residues might be conserved. From these observations, we were able to

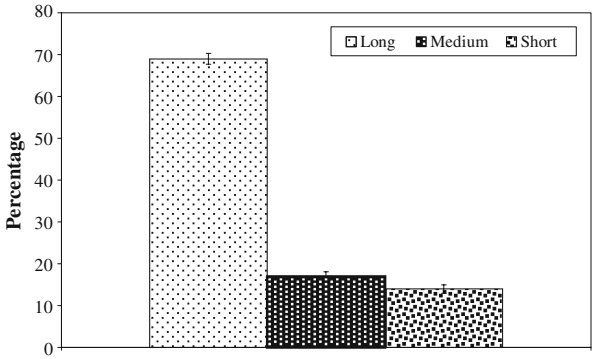

Fig. 8 Sequential separation of C–H…O interacting residues

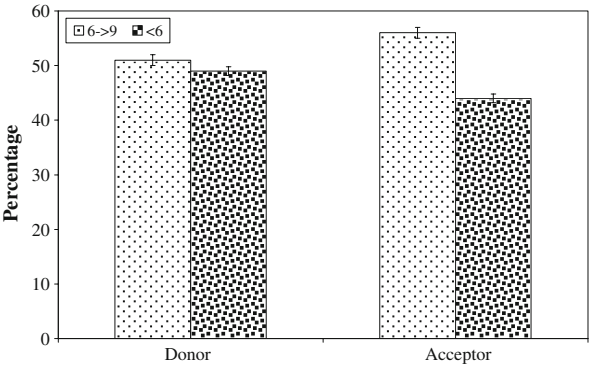

Fig. 9 Conservation patterns of C–H…O interacting residues

infer that most of the amino acid residues that are involved in C–H…O H-bonding might be conserved in glycoproteins.

Stabilization Centers

Stabilization centers in glycoproteins were computed, and the results are presented in Fig. 10. We found that significant percentage of amino acid residues that are involved in C–H…O H-bonding interactions had one or more stabilization centers in addition to their contribution to C–H…O interactions. From this observation, we infer that these residues might contribute additional stability to the glycoproteins in addition to their participation in C–H…O interactions.

Conclusion

In the present study, we have analyzed the roles played by C–H…O interactions in the structural stability of glycoproteins. There was an average of one significant C–H…O interaction for every five residues and an average of 70 C–H…O interactions per protein in the glycoproteins studied. We found that MM–CH…O interactions are the predominant type of interactions in glycoproteins. The higher contribution of aliphatic residues to C–H…O interactions might be due to the reason that the $C_{\alpha}H$ groups of aliphatic residues are directly adjacent to a pair of electronegative groups and hence serves as a relatively stronger

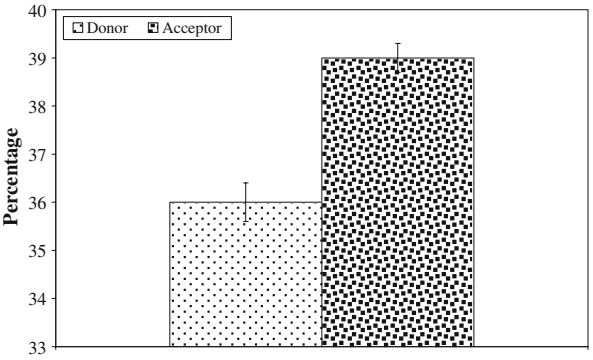

Fig. 10 Stabilization centers in glycoproteins

proton donor compared to other C–H groups that lack any substitution from electronegative groups. Similarly, the higher contribution of Lys, Pro, and His residues in SS and SM C–H…O interactions might be due to the ability of $C_{\varepsilon}H$ of Lys, $C_{\delta}H$ of Pro, and the $C_{\varepsilon}H$ of His to participate in side-chain C–H…O interactions. It was interesting to observe that Pro residues favored strand and helix conformation. It is noteworthy to mention that Pro residues might stabilize helices and strands by C–H…O H-bonding. We observed that most of the C–H…O interacting polar amino acid residues were solvent exposed, and most of the C–H…O interacting nonpolar residues were excluded from the solvent. Hence, the polar residues might contribute significantly to the stability of glycoproteins, as the contribution of the global energy is much greater in solvation. We found that long-range C–H…O interactions are the predominant type of interactions, and they might contribute significantly to the global conformational stability of glycoproteins. It was interesting to observe that most of C–H…O interacting residues had a higher conservation score and one or more stabilization centers. Since the C–H…O H-bonds occur frequently, cumulatively, these interactions might provide significant energy for protein stability. On the whole, we infer that our results on C–H…O H-bonds in glycoproteins might be helpful in understanding their role and importance in the structural stability of glycoproteins and as well as for further experimental studies in glycoproteins.

Acknowledgment The authors thank the management of VIT University for their support and encouragement to carry out this work and Dr. M. Madan Babu, Group leader, MRC Laboratory of Molecular Biology, Hills Road, Cambridge, CB22QH, UK for helping us with the NCI server.

References

1. Taylor, M. E., & Drickamer, K. (2003). *Introduction to glycobiology, chapter 9* (pp. 28–152). Oxford: Oxford University Press.
2. Hurtley, S., et al. (2001). *Science*, *29*, 2337. doi:10.1126/science.291.5512.2337.
3. Roucou, X., & LeBlanc, A. C. (2005). *Journal of Molecular Medicine (Berlin, Germany)*, *83*, 3–11. doi:10.1007/s00109-004-0605-5.
4. Graner, E., et al. (2000). *Brain Research. Molecular Brain Research*, *76*, 85–92. doi:10.1016/S0169-328X(99)00334-4.
5. Davies, G. A., et al. (2006). *Canadian Journal of Gastroenterology*, *20*, 18–24.
6. Pan, K., et al. (1993). *Proceedings of the National Academy of Sciences of the United States of America*, *90*, 10962–10966. doi:10.1073/pnas.90.23.10962.
7. Ruggeri, Z. M. (2004). *Haemostaseol*, *24*, 1–11. doi:10.1159/000075346.
8. Sugimoto, M., & Miyata, S. (2002). *International Journal of Hematology*, *75*, 19–24. doi:10.1007/BF02981974.
9. Dorny, P., et al. (2003). *Acta Tropica*, *87*, 79–86.
10. Barry, J. D., & McCulloch, R. (2001). *Advances in Parasitology*, *49*, 1–70. doi:10.1016/S0065-308X(01)49037-3.
11. Van der Woude, M. W., & Baumler, A. J. (2004). *Clinical Microbiology Reviews*, *17*, 581–611. doi:10.1128/CMR.17.3.581-611.2004.
12. Schinkel, A. H., et al. (1995). *The Journal of Clinical Investigation*, *96*(4), 1698–1705. doi:10.1172/JCI118214.
13. Wahl, M. C., & Sundaralingam, M. (1997). *Trends in Biochemical Sciences*, *22*, 97–102. doi:10.1016/S0968-0004(97)01004-9.
14. Desiraju, G. R., & Steiner, T. (1999). *The weak hydrogen bond in structural chemistry and biology*. New York: Oxford University press.
15. Li, Q., et al. (2007). *Journal of Molecular Structure. Theochem*, *847*, 68–74. doi:10.1016/j.theochem.2007.08.035.
16. Desiraju, G. R. (1996). *Accounts of Chemical Research*, *29*, 441–449. doi:10.1021/ar950135n.
17. Auffinger, P., & Westhof, E. (1997). *Journal of Molecular Biology*, *274*, 54–63. doi:10.1006/jmbi.1997.1370.

18. Derewenda, Z. S., et al. (1995). *Journal of Molecular Biology*, *252*, 248–262. doi:10.1006/jmbi.1995.0492.
19. Bella, J., & Berman, H. M. (1996). *Journal of Molecular Biology*, *264*, 734–742. doi:10.1006/jmbi.1996.0673.
20. Chakrabarti, P., & Chakrabarti, S. (1998). *Journal of Molecular Biology*, *284*, 867–873. doi:10.1006/jmbi.1998.2199.
21. Brandl, M., et al. (2001). *Journal of Molecular Biology*, *307*, 357–377. doi:10.1006/jmbi.2000.4473.
22. Steiner, T., & Koellner, G. (2001). *Journal of Molecular Biology*, *305*, 535–557. doi:10.1006/jmbi.2000.4301.
23. Leonard, G. A., et al. (1995). *Acta Crystallographica. Section D, Biological Crystallography*, *51*, 136–139. doi:10.1107/S0907444994004713.
24. Wahl, M. C., et al. (1996). *Nature Structural Biology*, *3*, 24–31. doi:10.1038/nsb0196-24.
25. Berger, I., et al. (1996). *Proceedings of the National Academy of Sciences of the United States of America*, *93*, 12116–12121. doi:10.1073/pnas.93.22.12116.
26. Biswas, R., & Sundaralingam, M. (1997). *Journal of Molecular Biology*, *270*, 511–519. doi:10.1006/jmbi.1997.1118.
27. Auffinger, P., & Westhof, E. (1997). *Journal of Molecular Biology*, *274*, 54–63. doi:10.1006/jmbi.1997.1370.
28. Brandl, M., et al. (1999). *Theoretical Chemistry Accounts*, *101*, 103–113. doi:10.1007/s002140050415.
29. Ghosh, A., & Bansal, M. (1999). *Journal of Molecular Biology*, *294*, 1149–1158. doi:10.1006/jmbi.1999.3323.
30. Steiner, T., & Saenger, W. (1992). *Journal of the American Chemical Society*, *114*, 7123–7126. doi:10.1021/ja00044a025.
31. Mandel-Gutfreund, Y., et al. (1998). *Journal of Molecular Biology*, *277*, 1129–1140. doi:10.1006/jmbi.1998.1660.
32. Gray, N. S., et al. (1998). *Science*, *281*, 533–538. doi:10.1126/science.281.5376.533.
33. Glusker, J. P. (1995). *Acta Crystallographica. Section D, Biological Crystallography*, *51*, 418–427. doi:10.1107/S0907444995003313.
34. Pascard, C. (1995). *Acta Crystallographica. Section D, Biological Crystallography*, *51*, 407–417. doi:10.1107/S0907444994014526.
35. Takahara, P. M., et al. (1996). *Journal of the American Chemical Society*, *118*, 12309–12321. doi:10.1021/ja9625079.
36. Senes, A., et al. (2001). *Proceedings of the National Academy of Sciences of the United States of America*, *98*, 9056–9061. doi:10.1073/pnas.161280798.
37. Ash, E. L., et al. (2000). *Proceedings of the National Academy of Sciences of the United States of America*, *97*, 10371–10376. doi:10.1073/pnas.97.19.10371.
38. Desiraju, G. R. (1997). *Science*, *278*, 404–405. doi:10.1126/science.278.5337.404.
39. Harakas, G., et al. (1998). *Journal of the American Chemical Society*, *120*, 6405–6406. doi:10.1021/ja9803411.
40. Kuduva, S. S., et al. (1999). *Journal of the American Chemical Society*, *121*, 1936–1944. doi:10.1021/ja981967u.
41. Anbarasu, A., et al. (2007). *International Journal of Biological Macromolecules*, *40*, 479–483. doi:10.1016/j.ijbiomac.2006.11.008.
42. Anbarasu, A., et al. (2007). *International Journal of Biological Macromolecules*, *41*, 251–259. doi:10.1016/j.ijbiomac.2007.02.013.
43. Anbarasu, A., et al. (2007). *Biosystems*, *90*, 792–801. doi:10.1016/j.biosystems.2007.04.002.
44. Anbarasu, A., et al. (2006). *Cytokine*, *35*, 263–269. doi:10.1016/j.cyto.2006.09.005.
45. Anand, S., et al. (2008). *In Silico Biology*, *8*, 0022.
46. Anand, S., et al. (2008). *International Journal of Biological Macromolecules*, *43*, 468–473. doi:10.1016/j.ijbiomac.2008.09.004.
47. Anbarasu, A., & Sethumadhavan, R. (2007). *Journal of Theoretical Biology*, *247*, 346–353.
48. Berman, H. M., et al. (2000). *Nucleic Acids Research*, *28*, 235–242. doi:10.1093/nar/28.1.235.
49. Babu, M. M. (2003). *Nucleic Acids Research*, *31*, 3345–3348. doi:10.1093/nar/gkg528.
50. Word, J. M., et al. (1999). *Journal of Molecular Biology*, *285*, 1735–1747. doi:10.1006/jmbi.1998.2401.
51. Pavlidis, P., & Noble, W. S. (2003). *Bioinformatics (Oxford, England)*, *19*, 295–296. doi:10.1093/bioinformatics/19.2.295.
52. Kabsch, W., & Sander, C. (1983). *Biopolymers*, *22*, 2577–2637. doi:10.1002/bip.360221211.
53. Gilis, D., & Rooman, M. (1996). *Journal of Molecular Biology*, *257*, 1112–1126. doi:10.1006/jmbi.1996.0226.
54. Gilis, D., & Rooman, M. (1997). *Journal of Molecular Biology*, *272*, 276–290. doi:10.1006/jmbi.1997.1237.

55. Gromiha, M. M., & Selvaraj, S. (1997). *Journal of Biological Physics*, *23*, 151–162. doi:10.1023/A:1004981409616.
56. Gromiha, M. M., et al. (2004). *International Journal of Biological Macromolecules*, *34*, 203–211. doi:10.1016/j.ijbiomac.2004.04.003.
57. Selvaraj, S., & Gromiha, M. M. (2003). *Biophysical Journal*, *84*, 1919–1925.
58. Gromiha, M. M., & Selvaraj, S. (2004). *Progress in Biophysics and Molecular Biology*, *86*, 235–277. doi:10.1016/j.pbiomolbio.2003.09.003.
59. Glaser, F., et al. (2003). *Bioinformatics (Oxford, England)*, *19*, 163–164. doi:10.1093/bioinformatics/19.1.163.
60. Dosztanyi, Z., et al. (1997). *Journal of Molecular Biology*, *272*, 597–612. doi:10.1006/jmbi.1997.1242.
61. Dosztanyi, Z., et al. (2003). *Bioinformatics (Oxford, England)*, *19*, 899–900. doi:10.1093/bioinformatics/btg110.
62. Novoa, J. J., & Mota, F. (1997). *Chemical Physics Letters*, *266*, 23–30. doi:10.1016/S0009-2614(96)01514-X.
63. Fabiola, G. F., et al. (1997). *Acta Crystallographica. Section D, Biological Crystallography*, *53*, 316–320. doi:10.1107/S0907444997000383.
64. Babu, M. M., et al. (2002). *Journal of Molecular Biology*, *322*, 871–880. doi:10.1016/S0022-2836(02)00715-5.
65. Bhattacharyya, R., & Chakrabarti, P. (2003). *Journal of Molecular Biology*, *331*, 925–940. doi:10.1016/S0022-2836(03)00759-9.
66. Boeckman, B., et al. (2003). *Nucleic Acids Research*, *31*, 365–370. doi:10.1093/nar/gkg095.