

INSILICO IDENTIFICATION OF PREFERRED AMINO ACIDS FOR SELECTIVE BINDING OF D-NEUP5AC**MANDAGE RAJENDRA HARIBHAU*¹ AND WADNERKAR AMOL SHRIRAM¹**¹Centre for Advanced Life Sciences, Deogiri College, Aurangabad, India* *Corresponding author* rajendra.mandage@gmail.com**ABSTRACT**

A statistical analysis of the carbohydrate–protein interaction in the PDB was carried out to investigate sequential aspects of the amino acids that are preferred for the selective binding of N-acetyl-D-neuraminic acid (D-Neup5Ac) chain. 2264 amino acids from 106 protein sequences were analysed in the vicinity of 277 Neup5Ac residues using GlyVicinity. Within a distance of 4 Å in vicinity of D-Neup5Ac, polar amino acids are more frequently observed as compare to other amino acids. There is a high frequency of occurrence of tyrosine, serine, asparagine, arginine residues in the vicinity of D-Neup5Ac chain whereas cysteine is the rarest amino acid found in protein–carbohydrate interaction. The positional preference of certain potential amino acids in vicinity of Neup5Ac residues has been discussed in terms of absolute counts, percentage, and deviations from natural abundances. This analysis provides a basis for further studies on the molecular mechanism of carbohydrate–protein interaction.

KEYWORDS

Carbohydrate–protein complexes, GlyVicinity, PDB, N-acetyl-D-neuraminic acid, absolute counts

INTRODUCTION

For many of the biological functions of carbohydrates residues, predominantly in variety of cell adhesion phenomenon, host-pathogen recognition and signal transduction, a high selectivity is required for these cellular processes. Carbohydrates frequently found at the cell surface serve as a cellular address tag that is used to interact with appropriate proteinaceous receptors and also clearance from the circulatory system^{1–3}. Most of these selective recognition mechanisms are mediated by carbohydrate binding proteins, the lectins, most of which selectively recognise

certain carbohydrate. For carbohydrate binding proteins, a specific distinction between diverse carbohydrates residues is also indispensable^{3–5}. It is not fully discovered yet how these proteins are able to differentiate specifically between very similar carbohydrate structures^{6–7}. In order to uncover such mechanism the participation of a particular amino acid of a carbohydrate binding protein and specific residues of the carbohydrate in the interaction need to be assessed which would provide information on the role played by individual residues to the protein–carbohydrate interaction^{8–9}. Proteins with similar structural characteristics can be selectively identifying

different types of carbohydrates residues¹⁰. Therefore, the selective recognition of specific carbohydrates clearly depends on the local conformation of the carbohydrate binding site rather than the overall structure of the protein. 3D structures of carbohydrate–protein complexes from PDB provide useful information on the structural requirements for carbohydrate binding and selective recognition between different carbohydrates residues^{10–16}. The major publicly accessible deposition of such 3D structures is the Protein Data Bank (PDB)¹⁷ which contains more than 2500 sequences on carbohydrate–protein complexes. These complexes consist of around 5900 carbohydrate chains with more than 10,000 carbohydrate residues. N-acetyl-D-neuraminic acid (Neup5Ac) is one of most abundant terminal residue in glycoconjugates (glycoproteins and glycolipids) and this particular type of residue has great commercial application. It is an acidic sugar with a major function in several biological functions, including anti-viral and anti-bacterial defensive activities¹⁸. Therefore an attempt is made to identify the amino acids that are necessary for the selective binding of Neup5Ac using Insilico approach.

MATERIALS AND METHODS

In order to investigate carbohydrate–protein interactions, D-Neup5Ac residues in the PDB database were identified by using the PDB2linucs¹⁹ tool located at glycosciences.de

portal [www.glycosciences.de/tools/glyvicinity/] 106 PDB entries for D-Neup5Ac residues with 254 chains were selected for the present study. Total 2264 amino acids in the vicinity of 277 carbohydrate residues were statistically analysed using the GlyVicinity²⁰ tool, available at the glycosciences.de portal, displays numerical information on the kinds of amino acids in vicinity of carbohydrate chains. In order to identify the amino acids in the vicinity of non-covalently bound D-Neup5Ac residues, the “ligands only” option has been chosen in GlyVicinity. Information about the amino acids within a 4 Å radius around D-Neup5Ac residues can be displayed as absolute counts, percentage, and deviations from natural abundances as numerically in tables or graphically in diagrams.

RESULT AND DISCUSSION

The frequency of occurrence of each of the 20 amino acids within a distance of 4 Å in the Vicinity of D-Neup5Ac carbohydrate residues the PDB database, polar amino acids are more commonly found than other amino acids (Table 1). The aromatic amino acid especially tyrosine forms a considerable exception. There is a high frequency of occurrence of serine, asparagines, arginine residues in the vicinity of D-Neup5Ac residues whereas cysteine is the rarest amino acid found in amino acid interaction with D-Neup5Ac in the PDB (Graph 1).

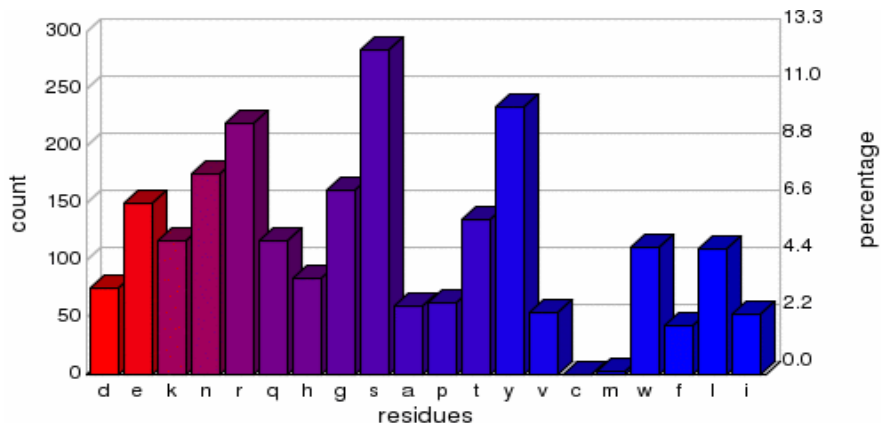
Table 1

Absolute count, percentage, and deviation of 20 aa in vicinity of 277 D-Neup5Ac residues

AA	Asp	Glu	Lys	Asn	Arg	Gln	His	Gly	Ser	Ala	Pro	Thr	Tyr	Val	Cy s	Met	Trp	Phe	Leu	Ile	Total
A Count	76	151	118	176	220	118	85	162	285	60	63	137	235	55	1	4	112	43	110	53	2264
Perc. (%)	3.4	6.7	5.2	7.8	9.7	5.2	3.8	7.2	12.6	2.7	2.8	6.1	10.4	2.4	0.0	0.2	4.9	1.9	4.9	2.3	100
Dev. (%)	-34	7.9	10	69	87	29	66	3.9	71	-64	-44	2.2	220	-62	-97	-92	269	-53	-47	-58	---

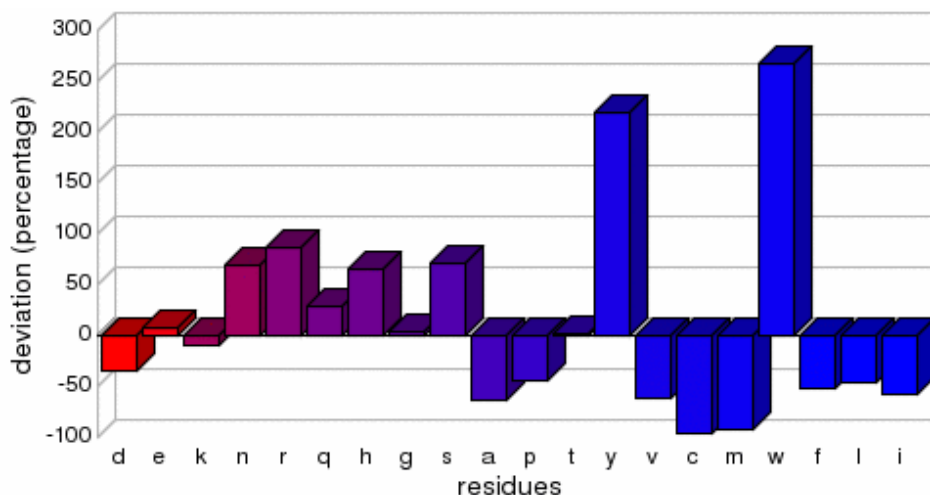
Graph 1

Amino acid count and percentage in vicinity of D-Neup2Ac residues



The GlyVicinity methodology can be utilised for identification of the specific binding of certain carbohydrate residues and amino acids. This is demonstrated for D-Neup5Ac residues. Deviation from average occurrence showed that tryptophan and tyrosine are over-expressed whereas threonine and glycine are under-expressed in statistical analysis using GlyVicinity (Graph 2). As far as preference of amino acids in the spatial vicinity of D-Neup2Ac residue is concern tryptophan shows the maximum deviation from its natural abundance, whereas tyrosine is poorly represented in addition methionine, cysteine and valine. Collectively these results suggest that effective binding of carbohydrate to D-Neup2Ac require aromatic amino acids for hydrogen bond formation and these amino acids also strengthen hydrophobic interaction of aromatic ring with D-Neup2Ac surface²¹

Graph 2
Amino acid Deviation from average occurrence in vicinity of D-Neup2Ac residues



The data presented in the present work clearly indicate that there is a prominent inclination for the amino acids in spatial vicinity of D-Neup5Ac residues this indicates that certain amino acids occur preferentially to provide

CONCLUSION

recognition specificity. Statistical analyses of such protein–carbohydrate interaction can be useful for identification of molecular basis of selective binding of glycan in terms of primary sequence hence, the specificity for certain carbohydrates residues observably lies in the local arrangement of the glycan binding site rather than the overall protein structure.

This analysis is only a predictive and needed to be verified experimentally, testing in vitro protein–carbohydrate interaction which could

constitute an important initial way towards the validation of for the discovery of carbohydrate recognition specificity. Ser and Tyr residues are greatly favored, suggesting a structural role for these amino acids in interaction. In future it will be of much significant to study further the potential structural and conformational implications of some of these identified positional preferences of the various amino acids involved in protein–carbohydrate interaction.

REFERENCES

1. Lis H, Sharon and N., Lectins: carbohydrate-specific proteins that mediate cellular recognition. *Chem Rev* 98: 637–674, (1998)
2. Helenius A, and Aebi M., Intracellular functions of N-linked glycans. *Science*, 291: 2364–2369, (2001)
3. Dennis JW, Granovsky M, and Warren E., Protein glycosylation in development and disease. *Bioessays*, 21: 412–421, (1999)
4. Smith AE, and Helenius A: How viruses enter animal cells. *Science*, 304: 237–242, (2004)
5. Dube DH, Bertozzi CR: Glycans in cancer and inflammation – potential for therapeutics and diagnostics. *Nat Rev Drug Discov*, 4: 477–488, (2005)
6. Loris R: Principles of structures of animal and plant lectins. *Biochim Biophys Acta*, 1572:198–208, (2002)
7. Uñligil UM, Rini JM: Glycosyltransferase structure and mechanism. *Curr Opin Struct Biol*, 10: 510–517, (2000)
8. Elgavish S, Shaanan B Lectin–carbohydrate interactions: different folds, common recognition principles. *Trends Biochem Sci*, 22: 462–467, (2004)
9. Sujatha MS, Balaji PV Identification of common structural features of binding sites in galactose specific proteins. *Proteins* 55: 44–65, ((2004))
10. Drickamer K Making a fitting choice: common aspects of sugar-binding sites in plant and animal lectins. *Structure* 5: 465–468, ((1997))
11. Rini JM Lectin structure. *Annu Rev Biophys Biomol Struct* 24: 551–577, (1995)
12. Weis WI, Drickamer K Structural basis of lectin–carbohydrate recognition. *Annu Rev Biochem* 65: 441–473, (1996)
13. Flint J, Nurizzo D, Harding SE, *et al.* Ligand-mediated dimerization of a carbohydrate-binding molecule reveals a novel mechanism for protein–carbohydrate recognition. *J Mol Biol* 337: 417–426, (2004)
14. Kulkarni KA, Katiyar S, Surolia A, *et al.* Structural basis for the carbohydrate-specificity of basic winged-bean lectin and its differential affinity for Gal and GalNAc. *Acta Crystallogr Sect D: Biol Crystallogr* 62: 1319–1324, (2006)
15. Blanchard H, Yu X, Coulson BS, von Itzstein M Insight into host cell carbohydrate-recognition by human and porcine rotavirus from crystal structures of the virion spike associated carbohydrate binding domain (VP8). *J Mol Biol*. 367: 1215–1226, (2007)
16. Banerji S, Wright AJ, Noble M, *et al.* Structures of the Cd44–hyaluronan complex provide insight into a fundamental carbohydrate–protein interaction. *Nat Struct Mol Biol* 14: 234–239, (2007)
17. Berman HM, Westbrook J, Feng Z, *et al* The Protein Data Bank. *Nucleic Acids Res* 28:235–24, (2000)
18. Von Itzstein M, and Wu W-Y, Rational design of potent sialidase-based inhibitors

- of influenza virus replication. *Nature*, 363:418-423, (1993)
19. Bohne-Lang A, Lang E and Forster T, LINUCS: linear notation for unique description of carbohydrate sequences, *Carbohydr Res*, 336:1-11, (2001)
 20. Lütteke T and Frank M, Carbohydrate Structure Suite (CSS): analysis of carbohydrate 3D structures derived from the PDB *Nucleic Acids Res*, 33: D242-6,(2005)
 21. Bradley W. McLean and Mark Bray, Analysis of binding of the family 2a carbohydrate-binding module from *Cellulomonas fimi* xylanase 10A to cellulose: specificity and identification of functionally important amino acid residues , *Protein Engineering*,13:801-809,(2000)