PREDICTION OF BINDING SITE OF SIGLECS USING COMPUTATIONAL TOOLS AND COMPARISON WITH EXPERIMENTAL FINDINGS

Madhumita Dandopath Patra*

Department of Chemistry, Acharya Prafulla Chandra College, New Barrackpore,
Kolkata-700131, West Bengal, INDIA

*E-mail: madhumita_ucst@yahoo.co.in

ABSTRACT

Siglecs are the major homologous subfamily of I-type lectins with the ability to recognize sialylated glycans. Their functions include control of myeloid cell interactions, inhibition of B cell receptor signaling, regulation of neuronal cell growth and maintenance of myelination in the nervous system. To date, there is very little data available describing their binding sites. Prediction of the binding site of siglecs by computational tools can help us to understand the function of this class of proteins. In our present study, we have predicted the binding site of sialoadhesin using software tools such as insight II and compared the results with the experimental findings. We applied an ab initio method using energy minimization and molecular dynamics simulation to search potential ligand-binding sites on the entire surface of the protein. Our putative sites showed dissociation constant values ranging from 1.4 mM - 0.093 μ M which are smaller than the site found from x-ray study implying higher affinity. Therefore, it may be concluded that different binding sites may exist in a solution other than the one found in the crystal structure.

Keywords: Docking, Modeling, Siglec, Sialic Acid, Sialoadhesin, Structure.

© RASĀYAN. All rights reserved

INTRODUCTION

Siglecs are a sialic acid-binding cell surface adhesion receptors that constitute the major sub-group of the I-type lectin family.¹⁻³ Recognition of an acidic nine-carbon monosaccharide, sialic acid (Neu5Ac), which often caps the non-reducing end of glycans⁴ by siglecs is thought to influence cellular adhesion, recognition and signaling in the immune and nervous systems.⁵⁻⁹ In Siglecs a single N-terminal V-set immunoglobulin-like domain functions as a binding receptor for sialic acid and is extended from the cell surface by a string of C-set immunoglobulin-like domains.¹⁰ Siglecs are divided into two subgroups one is recently characterized CD33-related Siglecs, which appears to be rapidly evolving¹¹ and the other subgroup consists of evolutionary conserved Siglecs 1, 2 and 4 (more frequently known as sialoadhesin, CD22 and myelin-associated glycoprotein/MAG, respectively) and Siglec-15.¹² Siglec expression appears to be restricted to cells of the hemopoietic and immune systems and appears to mediate both adhesive and signaling functions^{4,6} but myelin-associated glycoprotein (MAG, Siglec-4), found in the nervous system^{13,14} and Siglec-6, expressed on placental trophoblast cells.¹⁵

Sialoadhesin, a conserved cell surface receptor protein across mammals, is expressed in highest levels under the non-inflammatory condition on subsets of macrophages within hemopoietic and secondary lymphoid tissues.^{16,17} But in chronic inflammatory conditions, such as atherosclerosis and rheumatoid arthritis, Siglec-1 is expressed in high levels on the inflammatory macrophages.¹⁸ Sialoadhesin is the largest member of the Siglec family, with 16 C2-set Ig domains and one N-terminal V-set Ig-like domain (SnD1). The long neck of Siglec-1 may allow it to interact with sialic acid residues on a variety of extracellular matrix and cell surface molecules.^{19,20} The structure of the functional N-terminal domain of sialoadhesin has been determined in complex with the oligosaccharide 3'-sialyllactose (PDB ID: 1QFO) at 1.85 Å resolution.¹⁶

To date, there is very little data available describing the binding sites of siglecs. Prediction of the binding site of siglecs by computational tools may help us to understand the function of this important class of proteins. In our present study, we have searched for the binding site of a sialyl oligosaccharide to a siglec molecule with an *ab initio* method using energy minimization and molecular dynamics simulation on the entire surface of the x-ray structure (1QFO) followed by the calculation of theoretical dissociation constant values for several probable complexes and compared with the results obtained using the experimental technique.

EXPERIMENTAL

The study was done initially by taking the x-ray crystallographically determined structure of sialoadhesin in complex with 3' sialyllactose (PDB ID: 1QFO). Two amino acids, Asn-34 and Asp-102 were missing in the structure due to crystallographic disorder. These amino acids were added using our in-house software package of MODELYN.²¹ The segments of the structure where the amino acids were inserted were regularized separately using the InsightII 2005 of Accelrys (San Diego, CA) equipped with DISCOVER as the energy minimization and molecular dynamics module.

To find out the binding sites, initially, the ligand of the crystal structure was detached and placed about 50 Å away from the protein roughly perpendicular to its surface. Another initial position was chosen on the opposite surface. Similarly, about 10 different initial positions were selected in different directions. After the initial placement, a theoretical complex was defined using the Assembly/Associate option of insightII. At first, 100 steps each of the steepest descent and conjugate gradient energy minimization were done and dynamic simulation was carried out keeping the protein molecule fixed. Within 20 picoseconds of the dynamics, the ligand moved towards the protein molecule and attached automatically at a vulnerable position on the surface of the protein molecule. This structure was taken as an initial structure of the complex.

The initial ligand-protein complexes, thus obtained, were optimized using the combination of energy minimization and long molecular dynamic simulations repeatedly to obtain the most stable structure of the complex²². The absolute binding energies were calculated using the relation $\Delta G_{\text{bind}} = \alpha \Delta \langle V_{\text{l-s}}^{\text{el}} \rangle + \beta \Delta \langle V_{\text{l-s}}^{\text{vdw}} \rangle$ of Åqvist *et al.*,²³ where ΔG_{bind} is the absolute binding energy, Δ stands for differences in the electrical ($V_{\text{l-s}}^{\text{el}}$) and van der Waals ($V_{\text{l-s}}^{\text{vdw}}$) components of the free energies of the ligand solvent (l-s) systems i.e. in pure water and protein-containing water environments. The α and β were taken respectively as 0.5 and 0.16 as proposed by Åqvist *et al.* and used by earlier workers.^{24,25} Using the thermodynamic relation $\Delta G_{\text{bind}} = -RT \ln K_a$, where R is the ideal gas constant and T, is the absolute temperature, association constant (K_a) was calculated; the inverse of K_a is the dissociation constant K_d .

RESULTS AND DISCUSSION

In this study, we searched the binding site of a sialic acid-containing ligand on the surface of sialoadhesin using dynamics simulation as described above. The placement of the ligand on the surface of the lectin during the initial long dynamics run was driven by the electrostatic attraction between the negatively charged sialyl derivative and the nearest surface of the siglec molecule. The structure of the optimally bound complex was dominated by the local interactions between the ligand and the protein. Based on the interaction energy between the protein and the ligand we selected six complexes for further analysis. Total free energy of all the complexes, taking the protein and the ligand, were negative, but when compared with the total interaction energy of the ligand alone in the aqueous environment, some of the complexes showed relatively higher energies (Table-1) indicating that such complexes were not stable (Complex-III to V). Therefore, the calculated free energy (ΔG_{bind}) values turned out to be positive. For the complexes, (Complex-I, II and the x-ray structure), the calculated free energy (ΔG_{bind}) values were negative for which the K_a and K_d values could be calculated using the thermodynamic relations given above. The theoretical dissociation constant values for these complexes in aqueous solution ranged from 1.4 mM to 0.093 μM as shown in Table-1.

Ligand Binding Site in the Complex in the Crystal Structure

The ligand, 3'-sialyllactose, occupies a position such that one face of the sialic acid pyranose ring lies toward the protein surface, while the other is exposed to solvent. The carboxylate group of sialic acid

makes an essential electrostatic interaction with conserved arginine residue (Arg-97) of Siglec-1. The indole ring of Trp-2 and Trp-106 is in van der Waals contact with N-acetyl group and C9 carbon of glycerol side chain.

Table-1: Empirical Free Energies, their Difference in Water and Water-protein Environments and Corresponding ΔG_{bind} and K_d Values for the Complex Formation between Sialoadhesin and 3'-Sialyllactose in the Aqueous Solution.

Compound	Free Energy in kcal/mol			Difference in kcal/mol		ΔG_{bind} in kcal/mol	K_d
	Vdw	Elect	Total	Vdw	Elect		
Ligand*	-71.30	-245.09	-316.39				
X-ray complex	-64.52	-255.11	-319.63	+6.78	-10.02	-3.93	1.4 mM
Complex I	-57.77	-255.51	-313.28	+13.53	-10.42	-3.05	6.2 mM
Complex II	-78.84	-262.11	-340.95	-7.54	-17.02	-9.72	0.093 μ M
Complex III	-75.19	-235.75	-310.94	-3.89	+9.34	+4.05	
Complex IV	-76.29	-229.98	-306.27	-4.99	+15.11	+6.76	
Complex V	-70.81	-218.97	-289.78	+0.49	+26.12	+13.14	

* Values corresponding to the interaction energies in presence of water molecules only as needed for the calculation of K_d value using linear interaction energy approximation method.

The 8- and 9- hydroxyl groups of the glycerol side chain and amide nitrogen of the N-acetyl group form hydrogen bonds with the main-chain amide and carbonyl of Leu-107 and main-chain carbonyl of Arg-105, respectively. The main chain carbonyl of Ser-103 interacts with the 4-hydroxyl group of sialic acid. The phenolic hydroxyl of Tyr-44 forms only one hydrogen bond with 6-hydroxyl of galactose(Gal) residue (Fig.-1, Panel A). The calculated ΔG_{bind} value for sialoadhesin-3'-sialyllactose complex corresponds to dissociation constant (K_d) of 1.4 mM. The conformation of the ligand is stabilized by two intra-molecular hydrogen bonds as presented in Table-2.

Table-2: Hydrogen-bonding Network within the Binding Site of Complex between 3'-Sialyllactose and Sialoadhesin in the Crystal Structure. Distances are measured between Hydrogen and Acceptor or Donor Atom.

Ligand-protein Hydrogen Bonds		
Atoms of 3'-Sialyllactose	Atoms of Sialoadhesin	Distance(Å)
Neu5Ac		
O1A	Arg-97:NH2	1.83
O1B	Arg-97:NH1	1.78
O4	Ser-103: O	2.07
N5	Arg-105: O	2.01
O8	Leu-107:N	1.96
O9	Leu-107:O	1.84
Gal		
O6	Tyr-44:OH	1.87
Intramolecular Hydrogen Bonds		
Atoms of 3'-Sialyllactose	Atoms of 3'-Sialyllactose	Distance(Å)
Neu5Ac O10	Neu5Ac O7	2.01
Neu5Ac O8	Neu5Ac O1B	1.85

Ligand Binding Site in the Simulated Complex I

3'-Sialyllactose is found to bind on the surface of sialoadhesin very close that occupied by it in the x-ray structure. However, the orientation of the ligand is changed in the simulated structure of the complex I in solution compared to that observed in the crystal structure. The sialic acid moiety of the ligand is closer to the C-C' loop, which plays a major role in determining the binding specificities of siglecs.²⁶ Most of the residues of sialoadhesin participating in ligand binding are the same in the two structures as can be seen

from Fig.-1, Panel A & Panel B. The essential salt bridge formation interaction is maintained with the guanidino group of Arg-48 and sialic acid carboxylate group (Table-3).

Fig.-1: Orientation and Mode of Interaction of 3'-Sialyllactose in Complex with Sialoadhesin shown in Ribbons Representation; β Sheets are shown in Yellow with the Arrowheads pointing towards the C-termini; Random Coils and turns are in Maroon colored threads. The Residues of the Protein involved in Hydrogen bonding with the Ligand are shown in Stick representation and colored as Atoms (C = Green, O = Red, N = Indigo). The Ligand is shown as Red Sticks. Panel A: the Complex as in the Crystal Structure (1QFO); Panel B as in Simulated Complex I and Panel C as in Simulated Complex II.

The residues of the protein involved in hydrogen bonding with 3'-sialyllactose are conserved but the interaction pattern is not conserved due to the altered position of the ligand at the binding site (Fig.-1, Panel B). The phenolic hydroxyl of Tyr-44 makes a hydrogen bond with carboxylate group of sialic acid. The 8- hydroxyl group of the glycerol side chain forms a hydrogen bond with Arg-48. The 4- and 6-hydroxyl groups of the galactose interact with Arg-97 via direct hydrogen bonds. Glucose (Glc) forms two direct and two water-mediated hydrogen bonds with the protein. The C2- hydroxyl group of glucose makes water-mediated hydrogen bonds with main-chain amide and carbonyl of Leu-107. The C1-hydroxyl group of glucose interacts via a water-mediated hydrogen bond with the main chain carbonyl of Arg-105. The C6-hydroxyl group of glucose forms hydrogen bonds with main-chain carbonyl groups of Ser-103 & Arg-105. The conformation of the ligand is stabilized by three intra-molecular hydrogen bonds (Table- 3). The value of ΔG_{bind} was calculated by the linear interaction energy approximation for the complex and presented in Table-1. It may be noted that the calculated ΔG_{bind} value for the complex I is negative indicating that the complex formation of 3'-sialyllactose in this conformation with the protein in the aqueous medium is thermodynamically favorable. Values of ΔG_{bind} for complex I correspond to the dissociation constant (K_d) of 6.2 mM. The value is comparable to that obtained for the crystal structure of the complex (Table-1).

Table-3: Hydrogen-bonding Network within the Binding Site of the Simulated Complex I. Distances are measured between Hydrogen and Acceptor or Donor Atom.

Ligand-protein Hydrogen Bonds		
Atoms of 3'-Sialyllactose	Atoms of Sialoadhesin	Distance(Å)
Neu5Ac		
O1B	Tyr-44:OH	1.79
O1B	Arg-48:NH1	2.07
O8	Arg-48:NH2	2.03
Gal		

O6	Arg-97:NH1/NH2	2.39/1.93
O4	Arg-97:NH2	2.12
Glc		
O6	Ser-103: O	2.47
O6	Arg-105: NE	1.97
Water mediated Ligand-protein Hydrogen Bonds		
Atoms of 3'-Sialyllactose	Atoms of Sialoadhesin	*Distance(Å)
Glc		
O1	Arg-105: O	4.76
O2	Leu-107: NH/O	5.05/5.47
Intra-molecular Hydrogen Bonds in 3'-Sialyllactose in the Complex		
Atoms of 3'-Sialyllactose	Atoms of 3'-Sialyllactose	Distance(Å)
Neu5Ac O8	Neu5Ac O6	2.26
Neu5Ac O8	Neu5Ac O1B	1.99
Neu5Ac O7	Neu5Ac O9	1.96

*Distances between atoms linked through hydrogen-bonding via water molecule

Ligand Binding Site in the Simulated Complex II

The binding site of 3'-sialyllactose is different from that observed in the crystal structure and complex I. Orientation of the ligand at the binding site is similar to that observed in complex I (Fig.-1, Panel C). 3'-Sialyllactose makes several hydrogen bonds with the protein with a different set of residues. These interactions are summarized in Table-4. The conformation of the ligand is stabilized by three intra-molecular hydrogen bonds. The sialic acid carboxylate group makes hydrogen bonds with Lys-8 and Lys-110. The essential salt-bridge formation interaction with Arg-97 is absent here but lysine residues lead to a strong interaction with the sialic acid moiety of the 3'-sialyllactose which is reflected in the calculated K_d (0.093 μ M) value (Table-1). The 3- and 6-hydroxyl group of glucose and galactose make hydrogen bonds with the main-chain carbonyl group of Trp-2. The 3- and 4-hydroxyl groups of galactose moiety form hydrogen bonds with Lys-110 and the main chain nitrogen of Val-4. C2-hydroxyl group of glucose makes a hydrogen bond with Thr-1 main chain nitrogen. The sialic acid makes one water-mediated hydrogen bond between Ser-5 and its carboxylate group. Values of ΔG_{bind} for complex II corresponds to dissociation constant (K_d) of 0.093 μ M which is much smaller compared to those of the crystal structure and the modeled complex I.

Table-4: Hydrogen-bonding Network within the Binding Site of the Simulated Complex II. Distances are measured between Hydrogen and Acceptor or Donor Atom

Ligand-protein Hydrogen Bonds		
Atoms of 3'-Sialyllactose	Atoms of Sialoadhesin	Distance(Å)
Neu5Ac		
O1A	Lys-8:NZ	1.64
O1B	Lys-110:NZ	1.70
Gal		
O3	Lys110:NZ	2.25
O4	Val-4:N	2.08
O6	Trp-2:O	2.03
Glc		
O2	Thr-1: N	1.69
O3	Trp-2: O	2.09
Water mediated Ligand-protein Hydrogen Bonds		
Atoms of 3'-Sialyllactose	Atoms of Sialoadhesin	*Distance(Å)
Neu5Ac O1A	Ser-5: OG	3.36
Intramolecular Hydrogen Bonds in the Ligand		
Atoms of 3'-Sialyllactose	Atoms of 3'-Sialyllactose	Distance(Å)
Neu5Ac O8	Neu5Ac O1B	2.22
Neu5Ac O7	Neu5Ac O9	1.94

*Distances between atoms linked through hydrogen-bonding via water molecule

CONCLUSION

In this study, we have attempted to predict the binding site on sialoadhesin of sialic acid-containing ligands using an *ab initio* molecular dynamics simulation technique. Most of the programs for docking of small molecules into the binding site of a protein molecule use static structures of the ligand and the protein. Our method, presented here, takes the static x-ray structure of the siglec, sialoadhesin, at the beginning when the small molecule approaches from a distance focusing on a selected surface on the protein determined by the solid angle subtended by the center of the molecule. In subsequent steps of a complex structure optimization, the structures of both the small molecule and the residues within the reach of 10 Å are free to move around within the constraints imposed by the chemical bonding. At the same time, we provide the aqueous environment by putting the water molecules in the vicinity of the complex.

It is evident from our results that there may be diverse potential ligand binding sites in addition to that obtained from the crystal structure. The crystallization conditions some time differ from the solution conditions, as a result, the ligand may not fit into the more specific site of native *in vivo* biochemical environment leading to an unusual site in the crystal. One of our predicted ligand-binding site is very close to the site found in the x-ray crystal structure but with a little different orientation of the ligand. This result indicates that during crystallization, the position of the ligand may undergo some rearrangement in forming the complex. The binding constants of the complexes in this location are very close to each other supporting this possibility.

The binding site in the other complex (II), obtained in this search, is away from the x-ray crystal structure. Most intriguingly, it has a very high binding affinity as reflected by its submicromolar affinity. If such a binding site exists, this complex will exist in a solution that is very much different from the complex structure found in the crystal. From the point of structural bioinformatics, this structure is highly probable in solution but experimental solution complex structure determination using NMR spectroscopic studies may, in the future, can verify the possibility of such an alternative structure.

ACKNOWLEDGMENT

This work was financially supported by the CSIR MMP Project No CMM0017.

REFERENCES

1. A.Varki and T. Angata, *Glycobiology*, **16**, 1R (2006), DOI:10.1093/glycob/cwj008.
2. P. Crocker, J. Paulson, A. Varki, *Nature Reviews Immunology*, **7**, 255(2007), DOI:10.1038/nri2056.
3. P. R. Crocker, and A. Varki, *Trends in Immunology*, **22**, 337(2001a), DOI:10.1016/s1471-4906(01)01930-5.
4. R. Kannagi, *Current Opinion in Structural Biology*, **12**, 599 (2002), DOI:10.1016/s0959-440x(02)00365-2.
5. P. R. Crocker, *Current Opinion in Structural Biology*, **12**, 609 (2002), DOI:10.1016/s0959-440x(02)00375-5.
6. S. Kelm, and R. Schauer, *International Review of Cytology*, **175**, 137 (1997), DOI:10.1016/s0074-7696(08)62127-0.
7. S. Kelm, J. Gerlach, R. Brossmer, C. P. Danzer, and L. Nitschke, *Journal of Experimental Medicine*, **195**, 1207 (2002), DOI:10.1084/jem.20011783.
8. C. Li, M. B. Tropak, R. Gerlai, S. Clapoff, W. Abramow-Newerly, B. Trapp, A. Peterson and J. Roder, *Nature*, **369**, 747(1994), DOI: 10.1038/369744a0.
9. J. Munday, H. Floyd, P.R. Crocker, *Journal of Leukocyte Biology*, **66**, 705(1999), DOI:10.1002/jlb.66.5.705.
10. K.F. Bornhöfft, T. Goldammer, A. Rebl, S.P. Galuska, *Developmental & Comparative Immunology*, **86**, 219 (2018), DOI:10.1016/j.dci.2018.05.008.
11. T. Angata, R. Hingorani, N. M. Varki, and A. Varki, *Journal of Biological Chemistry*, **276**, 45128 (2001), DOI:10.1074/jbc.M108573200.
12. P. R. Crocker, *Current Opinion in Pharmacology*, **5**, 431(2005), DOI:10.1016/j.coph.2005.03.003.
13. T. Angata, *Journal of Biomedical Science*, **27**, 10(2020), DOI:10.1186/s12929-019-0610-1.

14. M. Fruttiger, D. Montag, M. Schachner, and R. Martini, *European Journal of Neuroscience*, **7**, 511(1995), DOI:10.1111/j.1460-9568.1995.tb00347.x.
15. N. Patel, E. C. Brinkman-Van der Linden, S. W. Altmann, K. Gish, S. Balasubramanian, J. C. Timans, D. Peterson, M. P. Bell, J. F. Bazan, A. Varki, and R. A. Kastelein, *Journal of Biological Chemistry*, **274**, 22729 (1999), DOI: 10.1074/jbc.274.32.22729.
16. A.P. May, R.C. Robinson, M. Vinson, P.R. Crocker, E.Y. Jones, *Molecular Cell*, **1**, 719(1998), DOI:10.1016/s1097-2765(00)80071-4.
17. P.R. Crocker, S. Kelm, C. Dubois, B. Martin, A.S. McWilliam, D.M. Shotton, J.C. Paulson, S. Gordon, *EMBO Journal*, **10**, 1661(1991), DOI:10.1002/j.1460-2075.1991.tb07689.x.
18. A.S. G. O'Neill, T.K. van den Berg, G.E. D. Mullen, *Immunology*, **138(3)**, 198(2013), DOI: 10.1111/imm.12042.
19. P.R. Crocker, A. Hartnell, J. Munday, D. Nath, *Glycoconjugate Journal*, **14**, 601(1997), DOI:10.1023/A:1018588526788.
20. J. Munday, H. Floyd, P.R. Crocker, *Journal of Leukocyte Biology*, **66**, 705(1999), DOI:10.1002/jlb.66.5.705.
21. C. Mandal, MODELYN-A Molecular Modelling Program Version PC-1.0 Indian Copyright No 9/98(1998).
22. V. Gayam, S. Ravi, G.M.V.N.A.R. Ravikumar and A. Thangamani, *Rasayan Journal of Chemistry*, **12 (2)**, 880 (2019), DOI:10.31788/RJC.2019.1225137.
23. J. Aqvist, C. Medina, J.E. Samuelsson, *Protein Engineering*, **7**, 385(1994), DOI:10.1093/protein/7.3.385.
24. J. Aqvist, S. L. Mowbray, *Journal of Biological Chemistry*, **270(17)**, 9978(1995), DOI:10.1074/jbc.270.17.9978.
25. J. Hultén, N.M. Bonham, U. Nillroth, T. Hansson, G. Zuccarello, A. Bouzide, J. Aqvist, B. Classon, U.H. Danielson, A. Karlén, I. Kvarnström, B. Samuelsson, A. Hallberg, *Journal of Medicinal Chemistry*, **40**, 885(1997), DOI:10.1021/jm960728j.
26. T. Yamaji, T. Teranishi, M. S. Alpey, P. R. Crocker and Y. Hashimoto, *Journal of Biological Chemistry*, **277**, 6324(2002), DOI: 10.1074/jbc.M110146200.

[RJC-5574/2019]