Indian J. Applied & Pure Bio. Vol. 38(3), 1237-1242 (2023). **ISSN: 0970-2091**
A web of Science Journal

Analysis of Viral protein interactions with NAG (N-Acetylglucosamine)

[*]**A. Stecy and S. Arul Mugilan**

P.G. Research Department of Physics, Kamarajar Government Arts College, Surandai, Tamil Nadu-627859 (India), Affiliated to Manonmaniam Sundaranar University, Tirunelveli-627012 (India)

[*]Reg. No. 21111062132004

[*]Corresponding Author E-mail : **astecy2705@gmail.com**

Abstract

Bio-computing plays a key role in macromolecular research and has created a chief impact in evaluating the sequencing and structure of the protein. Protein-ligand interaction at atomic resolution enables the design of small-molecule drugs for disease treatment. Here we present the analysis of different SARS Cov proteins that shortly interacted with NAG (N-Acetylglucosamine). Amino acids interact preferentially with ligands that are close in space and are represented graphically. In the above study, most of the data from the PDB indicate that the amino acid Arginine easily binds with the ligand NAG at a distance of around 3 AÚ. The structure of various protein IDs were visualized and their bond lengths are calculated using Python programming. Using the visualization tool, we clearly get to know that the protein IDs of Sars CoV favors 90% of Beta turn structure. The above analysis is beneficial for computation in addition to experimental biologists in field of drug designing.

Key words : NAG, PDB, RasMol, Python.

The function of a protein is dependent on its interaction with other molecules[27]. One of the most important modes of protein function is the protein-ligand interaction. A wide range of applications for structure-based drug design has been made possible by the binding of proteins to small molecules (known as ligands)[20]. Interactions between proteins and ligands play a major role in many biological processes [30]. Protein function can be regulated by ligand binding. The most effective way to study interactions is to solve complex structures using X-ray or nuclear magnetic resonance techniques[12,23]. It is extremely difficult or impossible to determine the structure of some large proteins and membrane proteins using traditional techniques[4,8,15]. Experiments are often time-consuming and expensive, and many computational attempts have been made to facilitate the study of protein-ligand interactions[9,18,19,26,28]. A cheaper alternative to solving the structure is to locate the binding sites based on theory[9.11,14,29,31]. The number of protein sequences available today is

increasing, but the data is not being analysed[32]. Recently, computational methods have been used to investigate the atomic-level functions of proteins[1]. However, a structured approach has certain drawbacks. Sequence-based methods have been developed to overcome this problem, which has led to many novel predictions of protein-ligand interactions. Nag is a prominent member of the carbohydrate group, especially a monosaccharide and it plays a vital role in structural processes and it maintains human health[2]. In bacteria it forms the peptidoglycan that lines the cell wall and in plants it forms the chitin that lines the cell wall[22]. Cells of animals also contain glycosaminoglycans in their extracellular matrix[17,21]. Additionally, it regulates the expression of genes and cell signalling in bacteria and fungi. Nag is also used for cell signaling in plants and animals[22]. As a result of the attachment of OGlcNAc to the protein Nag acts as a sensor for nutrition. NAG has been proven to be efficacious in the treatment of autoimmune diseases. In humans, NAG signaling allows bacteria, fungi, and human cells to coexist successfully[16]. This paper presents a simple approach for predicting the NAG binding residues in viral proteins.

Data collection :

In the current study, we used crystallographic data from the Protein Data Bank of Brookhaven National Laboratory to analyse viral proteins[6,7,10]. Proteins selected for this study were nonhomologous, and their structures, as well as their interactions with NAG, were determined at a high level of resolution. In this work, viral proteins were used with similarities between 30% (1508) and 90% (1050), their differences (458) being taken into account. All data sets with repeat protein sequences were removed, and the remaining IDs with ligand-interacting proteins were retained. Once the protein has been selected, a Python program is built with a range less than 13 A^{o} to determine the bond length[5,13]. It will show the outcomes of amino acids that shortly interacted with the ligand NAG. After the amino acids have been separated from the results , the data is subjected to statistical analysis[25]. The frequency of occurrence of amino acid singlets and triplets is calculated using the formula for every 458 proteins.

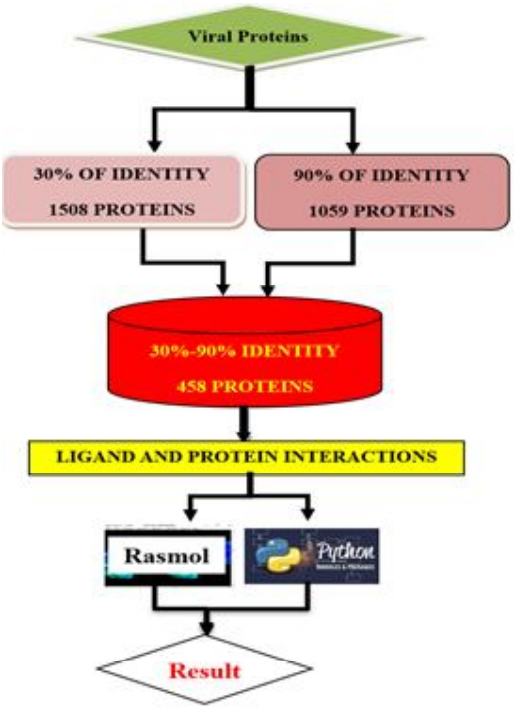
$$P(x)=\frac{No\ of\ \text{counts in particular AA}}{Total\ No\ of\ aminoacid\ in\ proteins}$$

In this approach, we used the values for the frequency of occurrence of singlet and triplet as expected count and observed count. The chi-square value for AA singlet and triplet in a given structural element is calculated using the formula[33],

$$\chi^2 = \frac{\{Observed - Expected\}^2}{Expected}$$

If the chi-square is positive, it indicates the preferred choice, and if it is negative, it indicates a non-preferred choice. Using this formula, we may be able to find the residue that preferentially binds to the NAG. Also, we may be able to find the AA that is in the non-preferential zone. Here, 15 proteins were chosen at random and the work was done similarly for all influenza proteins. With visualisation tool, the shortest distances between amino acids that interact with NAG are visualized. In addition, RasMol is used to analyze the structures of amino acids that interact with each other[3].

Flow Chart :



The results were retrieved using the program and are represented graphically (Fig. 1).

Graphical representation of the Chi-square test of various protein IDs with NAG as a ligand:

The above graphical analysis categorises Chi-square values into four categories: -1, which indicates that amino acids are not in range, and above 0.5, which indicates that amino acids are interacting with NAG. Finally, -0.5 denotes the non-preferred region, while 1.5 and higher denote the highly preferred region of amino acids. From graphical representation we might able to conclude Arginine is the most common residue found to bind with NAG, followed by Phenylalanine, Glutamic acid, Aspartic acid and Asparagine whereas residues like Alanine, Cysteine, Histidine, Methionine and Valine are not preferred.

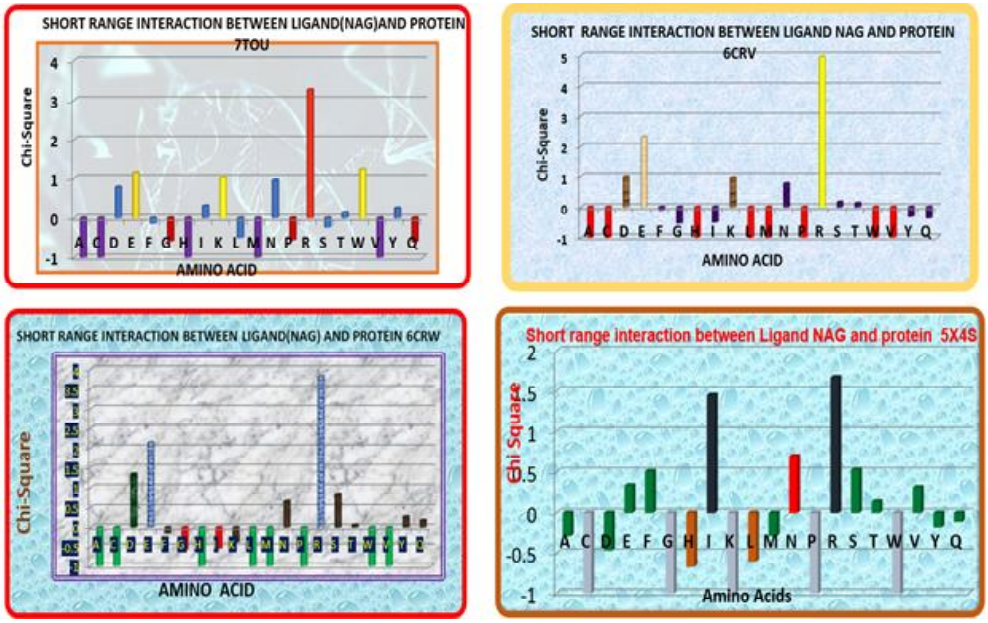


Fig. 1. Interactions of amino acids with the ligand NAG in the preferential and non-preferential regions.

Interactions and structure of aminoacid with ligand using Ras Mol :
(Figs 2 & 3)

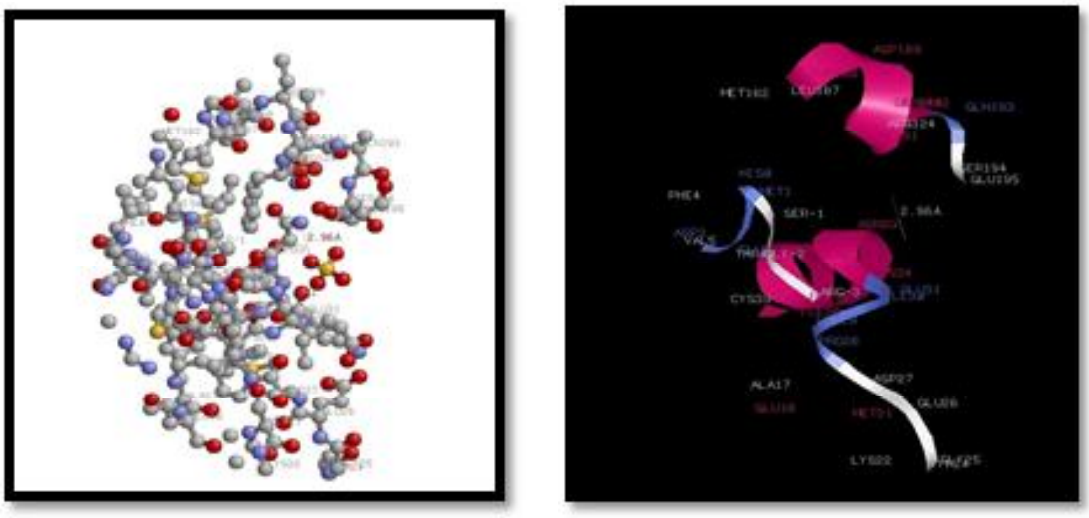


Figure 2 & 3 Depicts the interactions and structure of 7TOU

The above figure represents the binding portion of amino acid residue with NAG and the structure of protein.

To understand the process of NAG, we need to know which amino acid residues interact with Nag. We have developed a Python program to find the binding residue. Python was chosen due to its widespread use and wide range of applications in biosciences and this method is simple and time consuming, it only requires a protein sequence. We can perform and analyse many sequences simultaneously without prior knowledge of structural information. The main conclusion is that the residues of the NAG binding site are highly conserved and NAG is bound to the beta turn structure.

I would like to thank my professor Arul Mugilan for his guidance.

References :

1. Agarwal P, H Singh, HK Srivastava, S Singh, G Kishore, and GPS Raghava, (2019). Benchmarking of different molecular docking methods for protein-peptide docking, *BMC Bioinformatics., 19*(S13): 426.
2. Alvarez-Anorve, L.I., M.L. Calcagno, and J. Plumbridge, (005) Why Does *Escherichia coli* Grow More Slowly on Glucosamine Than on *N*-Acetylglucosamine? Effects of Enzyme Levels and Allosteric Activation of Glcn6p Deaminase (NagB) on Growth Rates, *J. Bacteriol., 187:* 2974–2982.
3. Amber D. Fraley, E. Katherine, Victor Odegaard, H. Vilchiz, Kang Chulhee

and Cathy Lee, (2022) *The American Biology Teacher., 84*(6): 369–373.
4. Arkin, M.R., Y. Tang, and J.A. Wells, (2014). *Chem. Biol., 21:* 1102–1114.
5. Bassi S, (007) *PLoS Comput Biol., 3*(11): E199.
6. Berman, H.M., J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, and P.E Bourne, (2000) *Nucleic Acids Research., 28:* 235–242.
7. Bernstein FC, TF Koetzle, GJ Williams, EF Meyer, MD Brice, JR Rodgers, Kennard Olga, Shimanouchi, Takehiko and Tasumi Mitsuo, (1977) *Journal of Molecular Biology., 112*(3): 535–542.
8. Blundell, T.L. and S Patel, (2004) *Curr. Opin. Pharmacol., 4:* 490–496.
9. Brylinski, M., and J. A Skolnick, (2008) *Proc. Natl. Acad. Sci. U.S.A., 105:* 129134.
10. Burley, S.K., H.M. Berman, G.J. Kleywegt, J.L. Markley, H. Nakamura, S Velankar, (2017) 627–641.
11. Capra, J.A., R.A. Laskowski, J.M. Thornton, M. Singh, and T.A. Funkhouser, (2009) *PLoS Comput. Biol., 5:* E1000585.
12. Cavalli, A., X. Salvatella, C. M. Dobson, and M. Vendruscolo, (2007) *Proceedings of the National Academy of Sciences., 104*(23): 9615–9620.
13. Ekmekci B, CE Mcanay and C Mura, (2016) *PLoS Comput Biol., 12*(6): 2e1004867.
14. Heo, L., W.H. Shin,, M.S Lee, and C Seok, (2014) *Nucleic Acids Res., 42:* W210–W214.
15. Higueruelo, A.P., H. Jubb and T.L Blundell (2013) *Curr. Opin. Pharmacol., 13:* 791–796.
16. Jermy K. Nicholson, Elane Holmes, James Kinross, Remy Burcelin, Glenn G, Ibson, Eei Jia, and Sven Petterson, (2012) *Science., 336:* 1262–1267.
17. Konopka JB, (2012) *Scientifica.*, 1–15.
18. Li, J., A. Fu, and L. Zhang (2019) *Interdisciplinary Sciences : Computational Life Sciences., 11(2):* 320–328.
19. Lin, X., X Li, and X. A. Lin, (2020) *Molecules., 25(6):* 1375.
20. Malathi, K., S. Ramaiah, K. Malathi, and S Ramaiah, (2018) *Biotechnology and Genetic Engineering Reviews.,* 1–
21. Moussian B., (2008) *Comp Biochem Physiol B Biochem Mol Biol., 149:* 215–226.
22. Naseem S, SM Parrino, DM Buenten and JB. Konopka, (2012) *Commun Integr Biol., 5:* 156–159.
23. Opella, S.J, (2013) *Annu. Rev. Anal. Chem., 6:* 305–328.
24. Overington, J.P., B. Al-Lazikani, and A.L Hopkins, (2006) *Nature Reviews Drug Discovery., 5*(12)*:* 993–996.
25. Patrick Aloy, Cedano Juan, and Oliva Baldomero, (1997) 'Transmem': a neural network implemented in Excel spreadsheets for predicting transmembrane domains of proteins, *Oxford University Press.,* Vol. *13:* 231-234.
26. Pellecchia, M., I. Bertini, D. Cowburn, C. Dalvit, E. Giralt, W. Jahnke, and G Siegal, (2008) *Nature Reviews Drug Discovery., 7*(9)*:* 738–745.
27. Perozzo, R., G. Folkers, and L Scapozza, (2004) *J. Recept. Signal. Transduct. Res., 24:* 1–52.
28. Roche, D.B., M.T Buenavista, and L.J

Mcguffin, (2013) *Nucleic Acids Res., 41:* W303-W307.

29. Roy, A., J. Yang, Y. Zhang, A. Roy, J. Yang, and Y Zhang, (2012) *Nucleic Acids Res., 40:* W471–W477.
30. Shasank S. Swain and Hussain Tahziba, Role of Bioinformatics in Early Drug Discovery: An overview and Perspective, Computation in Bioinformatics: Multidisciplinary Applications., 2021.
31. Yang, J., A. Roy, and Y Zhang, (2013) *Bioinformatics., 29:* 2588–2595.
32. Yu D-J, J Hu, H Yan, X-B Yang, J-Y Yang, and H-B. Shen, (2014) *BMC Bioinformatics*., *15*(1): 297.
33. Zhu, P., P. Bowden, M. Tucholska and J.G Marshall, (2011) *Analytical Biochemistry., 409*(2): 189–194.