neoGRID: Enabling access to teragrid resources

Application for sialylmotif analysis in the protozoan Toxoplasma gondii

Arun K. Datta*
National University (NUCRI)
11255 N. Torrey Pines Road
La Jolla, CA 92037 (USA)
Ph: (001) 858-642-8535
adatta@nu.edu
*Corresponding author

Becca Eller
University of Kentucky
College of Medicine
800 Rose St.
Lexington, KY 40536 (USA)
(001) 859-323-1341
becca.eller@uky.edu

Anthony P. Sinai
University of Kentucky
College of Medicine
800 Rose St.
Lexington, KY 40536 (USA)
(001) 859-323-6680
sinai@email.uky.edu

ABSTRACT

The neoGRID is under development in Quarry, a virtual hosting environment, for working with Taverna-based workflow utilizing grid computing. Taverna[1] is a graphical workbench often used for biomedical informatics [1, and the references therein]. neoGRID is designed to offer a HPC-supported collaborative environment for the researchers from multidisciplinary scientific fields to gather data, integrate and analyze using teragrid resources. A significant number of resources including bioinformatics tools [2] has been deployed in TeraGrid[2], an NSF funded project. Besides, the myGrid team[3] produced a suite of tools that are made available for analysis of proteins[4]. In addition, their myExperiment site makes it easy to find, use and share scientific workflows and building scientific communities with common interests. This Cyberinfrastructure (CI)-supported neoGRID can be utilized for protein motifs analysis, particularly in the glycosyltransferase protein family and will be available to the Glyco-Community [3]. Workflows specifically designed for such analysis will be available in neoGRID. Moreover, a significant number of workflows are available in KEGG[5] that are useful for protein analysis and drug discovery research. These and modification of these workflows in a collaborative environment will also be available for drug discovery research using members of glycosyltransferase enzyme family as target protein(s) [4]. Here, we demonstrate the usage of neoGRID for analysis of sialylmotifs of sialyltransferases. The presence of sialylmotifs is the cardinal feature of mammalian sialyltransferases [5], a group of enzymes that transfers sialic acid from CMP-NeuAc to the terminal carbohydrates group of various glycoproteins and glycolipids [6]. Sialic acid has been recognized as the key determinant of a diverse oligosaccharide structures involved in a large variety of biological events as diverse as animal cell-cell interaction to oncogenic transformation [7]. These conserved protein domains, sialylmotifs, have been shown to be involved in binding either the donor or acceptor substrates or both [8, 9], and a disulfide linkage between these two motifs has been shown to be essential for catalytic activity [10, 11]. Protein sequence [12] and structural analysis [11, 13] showed that mammalian sialyltransferase has no similarities with the bacterial enzymes, although a His residue serves as a catalytic center for both [11]. This provides a unique opportunity for discovery research on potential drug development. A thorough analysis of these motifs is now under study for drug discovery research using sialyltransferase as a target protein. Such analysis demanding high-performance computing power is available in neoGRID.

The genome of protozoan parasite Toxoplasma gondii has been used here for bioinformatics analysis. T. gondii, which chronically infects roughly 30% of the world's population is typically asymptomatic but can cause life threatening disease in immune compromised individuals and birth defects of acquired during pregnancy [14]. Sequencing of this parasite genome of about 63 Mb in size with 14 chromosomes, has recently been completed [15, 16]. The availability of this sequence information from multiple isolates in a specifically designed database (www.toxodb.org) that also include transcriptomic and proteomic data sets [16], has made this parasite an ideal candidate for in-depth bioinformatic interrogation. At this time, little is known about the parasite glycome or the encoded capacity for glycosylation. This deficit is in spite of the fact that the parasite form associated with chronic infection is heavily glycosylated [17]. Our initial studies have identified a broad spectrum of lectin reactivities [18]. Lectins recognize their specific glycan targets with high specificity and affinity [19]. We were particularly surprised to find evidence for sialylation of the parasite tissue cyst forms given that the genome does not appear to encode recognizable sialyltransferases based on the 'sialylmotif' [5]. This finding suggests that Toxoplasma may hijack the sialyltrasferase activities of the infected host cell or alternatively possess activities with entirely novel functional signatures.

The integration of state of the art computational biology and bioinformatics with experimental validation provides a unique opportunity toward new discovery. By conducting a systematic survey of glycosyltransferase activities *in silico* we can establish both the presence and absence of specific activities in the Toxoplasma genome. Given the evolutionary antiquity of the

[1] http://www.taverna.org.uk/

[2] https://www.teragrid.org/

[3] http://www.mygrid.org.uk/

[4] http://utopia.cs.manchester.ac.uk/

[5] http://www.genome.jp/kegg/

TG'11, Jul 18-21 2011, Salt Lake City, UT, USA. ACM 978-1-4503-0888-5/11/07.

parasitic protozoa we expect the combination of *in silico* analysis and experimental validation to offer new insights into the biology of these parasites. The availability of sequenced genomes of several related parasites housed at www.eupathdb.org [20] provides an ideal resource for expanding these studies to other parasites including Plasmodium species, the agent of malaria. *In silico* identification of potentially unique enzymatic activities could open doors toward the discovery of novel drugs to treat these often deadly infections.

Keywords
Parasite, Toxoplasma gondii, Apicomplexa, glycan, sialylmotif, sialyltransferases, glycosyltransferases.

ACKNOWLEDGMENTS
The authors like to acknowledge Jeffrey Tilson of RENCI and Eric Jakobsson of UIUC for MotifNetwork analysis and helpful discussions. Contributions of Andi Sumargo and Yared Tadessi of NUCRI for setting up neoGRID are also acknowledged.

REFERENCES

[1] Datta A.K., Tilson, J.L., Rendon, G., and Jakobsson, E. 2009. High-throughput sialylmotif analysis in the glycosyltransferase protein family. *Proceedings of the TeraGrid '09* (Arlington, Virginia, June 22 – 25, 2009). DOI= http://archive.teragrid.org/tg09/files/tg09_submission_43.pdf

[2] Mulder, N.J. and Apweiler, R. 2001. Tools and resources for identifying protein families, domains and motifs. *Genome Biology,* 3: reviews 2001. DOI=10.1186/gb-2001-3-1-reviews2001. The electronic version at: http://genomebiology.com/2001/3/1/reviews/2001

[3] Datta, A.K. 2009. Cyberinfrastructure for Glycome Research. *Glycoconjugate J.*, 26(7), 849-850.

[4] Datta, A.K. 2010. High-throughput motif analysis for drug discovery research. *Proceedings of the BIT's 8th Annual Congress of International Drug Discovery Sciences and Technology* (Beijing, China, October 23-26, 2010). DOI= http://www.iddst.com/2010), p137.

[5] Datta, A.K. 2009. Comparative sequence analysis in the sialyltransferase protein family: Analysis of Motifs. *Current Drug Targets*, 10 (6), 483-98. Review.

[6] Tsuji, S., Datta, A.K., and Paulson, J.C. 1996. Systematic nomenclature for sialyltransferases. *Glycobiology*, 6 (7): v – vii.

[7] Varki, A. 1997. Sialic acids as ligands in recognition phenomena. *FASEB J.*, 11, 248-255.

[8] Datta, A.K. and Paulson, J.C. 1995. The sialyltransferase sialylmotif participates in binding the donor substrate CMP-NeuAc. *J. Biological Chemistry*, 270, 1497 – 1500.

[9] Datta, A.K., Sinha, A. and Paulson, J.C. 1998. Mutations of the sialyltransferase S-sialylmotif alters the kinetics of the donor and acceptor substrates. *J. Biol. Chem*, 273, 9608 – 9614.

[10] Datta, A.K., Chammas, R., and Paulson, J.C. 2001. Conserved Cysteines in the sialyltransferase sialylmotifs form an essential disulfide bond. *J. Biol. Chem*, 276,15200-15207. Angata, K., et al., 2001. Unique disulfide bond structures found in ST8Sia IV polysialyltransferase are required for its activity. *J Biol Chem*, 276(18), 15369-15377.

[11] Rao, F.V., Rich, J.R., Rakić, B., Buddai, S., Schwartz, M.F., Johnson, K., Bowe, C., Wakarchuk, W.W., Defrees, S., Withers, S.G., Strynadka, N.C. 2009. Structural insight into mammalian sialyltransferases. *Nat Struct Mol Biol*., 16(11), 1186-1188.

[12] Yamamoto, T. 2010. Marine bacterial sialyltransferases. *Mar Drugs,* 8(11), 2781-2794. Review.

[13] Chiu, C.P., Watts, A.G., Lairson, L.L., Gilbert, M., Lim, D., Wakarchuk, W.W., Withers, S.G., Strynadka, N.C. 2004. Structural analysis of the sialyltransferase CstII from Campylobacter jejuni in complex with a substrate analog. *Nat. Struct. Mol. Biol*., 11(2), 163-170.

[14] Tenter, A. M., Heckeroth, A. R., and Weiss, L. M. 2000. Toxoplasma gondii: from animals to humans. *Int. J. Parasitol.,* 30, 1217-1258.

[15] Bontell, I.L., Hall, N., Ashelford, K.E., Dubey, J.P., Boyle, J.P., Lindh, J., Smith, J.E. 2009. Whole genome sequencing of a natural recombinant Toxoplasma gondii strain reveals chromosome sorting and local allelic variants. *Genome Biol*., 10(5), R53. doi: 10.1186/gb-2009-10-5-r53.

[16] Gajria, B., Bahl, A., Brestelli, J., Dommer, J., Fischer, S., et. al., 2008. ToxoDB: an integrated Toxoplasma gondii database resource. *Nucleic Acids Res.,* 36, D553-556.

[17] Weiss, L.M. and Kim, K. 2000. The development and biology of bradyzoites of Toxoplasma gondii. *Front. BioSci.,* 5, D391-D405.

[18] Eller, B., Lynn, B.C., Datta, A.K., and Sinai, A.P. 2011. Characterization of T. gondii tissue cyst lectin reactivity and the identification of novel glycoproteins. *Proc. 11th International Congress on Toxoplasmosis* (Ottawa, Ontario, Canada, June, 25 – 29, 2011).

[19] Cummings, R.D. and Etzler, M.E. 2008. Antibodies and Lectins in Glycan Analysis, in *Essentials of Glycobiology*, 2nd edition. (Ed. Varki, A. et al.) Cold Spring harbor press, Cold Spring , NY.

[20] Aurrecoechea, C., Brestelli, J., Brunk, B.P., Fischer, S., 2010. EuPathDB: a portal to eukaryotic pathogen databases. *Nucleic Acids Res*., 38, D415-D419.

TG'11, Jul 18-21 2011, Salt Lake City, UT, USA. ACM 978-1-4503-0888-5/11/07.