

# BacterialLectinDb: An integrated bacterial lectin database

Dharmendra Kumar<sup>1\*</sup> & Yashoda Mittal<sup>2</sup>

<sup>1</sup>Department of Zoology, Government Post Graduate College, Rishikesh, Uttarakhand, India; <sup>2</sup>Principal, KLDVA, Post Graduate College, Roorkee, Uttarakhand, India; E-mail: research.bioinformatics@gmail.com; \*Corresponding author

Received June 10, 2011; Accepted March 03, 2012; Published March 31, 2012

## Abstract:

Studies of various diversified bacterial lectins/ lectin data may serve as a tool with enormous promise to help biotechnologists/geneticists in their innovative technology to explore a deeper understanding in proteomics/ genomics research for finding the molecular basis of infectious diseases and also to new approaches for their prevention and in development of new bacterial vaccines. Hence we developed a bacterial lectin database named 'BacterialLectinDb'. An organized database schema for BacterialLectinDb was designed to collate all the available information about all bacterial lectins as a central repository. The database was designed using HTML, XML.

**Availability:** <http://www.research-bioinformatics.in>

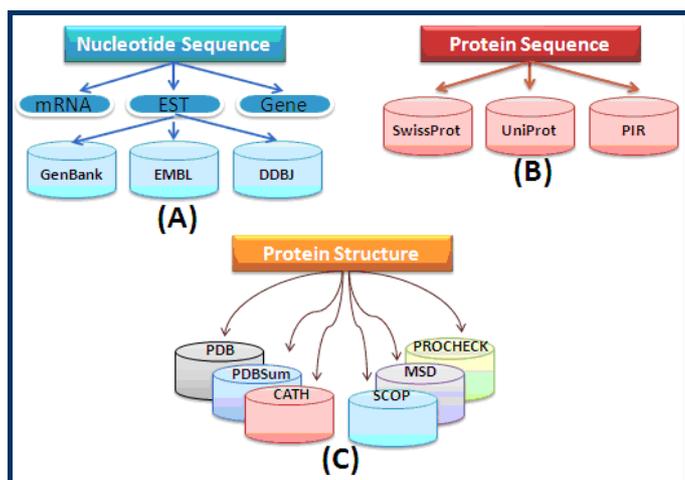
**Keywords:** Bacterial Lectins, Bacterial Vaccines, Glycan Specificity and Structure Annotation, Lectin- Genome/Proteome, Data Curation.

## Background:

Most lectins are carbohydrate binding proteins of non-immuneorigin that agglutinate cells/ bind glycans of glycoproteins, glycolipids and polysaccharides. Besides, havinghemagglutinating activity, Nathan Sharon and colleague's for the first time described bacterial surface lectins as adhesions because primarily lectins function as to facilitate attachment or adherence to host cells (receptors), a prerequisite for bacterial colonization [1]. Bacterial lectins occur commonly in the form of elongated, submicroscopic, multi subunit protein appendages known as fimbriae (hair) or pili (threads), which interact with glycoprotein and glycolipid receptors on host cells. Among the best characterized of bacterial surface lectins are mannose specific (Type 1) lectins [2]. Although bacterial colonization is not always pathogenic e.g. the normal flora of the lower gastrointestinal tract is determined by appropriate and desirable colonization by beneficial bacteria similarly colonization of *Rhizobium* in nitrogen-fixing nodules of leguminous root tips which involve lectins on the root tip binding to Nod factors generated by the bacterium, is beneficial yet most of the bacterial surface lectins appear to function primarily in the initiation of infection by mediating bacterial

adherence to the host epithelial cells The mannose specific lectins also act as recognition molecules in lectinophagocytosis (i.e. phagocytosis of the bacteria in the absence of opsonins) by mouse, rat and human peritoneal macrophages, and human polymorphonuclear leukocytes [3]. Detailed studies of then specificity of microbial lectins have led to the identification and synthesis of powerful inhibitors of adhesion that may form the basis for therapeutic agents for treating infection. Genome and proteome research on pathogenic and associative bacteria also provide important information on bacteria host interaction [4]. Progress has been made to understand the structure function relationship for several bacterial to lectins in order to know vaccine against several infections and other diseases [4]. Studies of various diversified bacterial lectins/ lectin data may serve as a tool with enormous promise to help biotechnologists/geneticists in their innovative technology to explore a deeper understanding in proteomics/ genomics research for finding the molecular basis of infectious diseases and also to new approaches for their prevention and in development of new bacterial vaccines. In the present work the efforts have been made to develop an integrated knowledge based bacterial

lectins database together with appropriate analytical tools which we named as "BacterialLectinDb".



**Figure 1** (A) Different Nucleotide sequences obtained from GenBank, EMBL and DDBJ; (B) Protein sequences obtained from SwissProt, UniProt and PIR databases; (C) Various forms of protein structures obtained from PDB, PDBSum, CATH, MSD, PROCHECK and SCOP.

## Methodology:

### Data collection

Data curation was initiated with different protein databases available through National Center for Biotechnology Information (NCBI) and other protein information repositories as shown in (Figure 1). Initially a search was carried with the keyword "bacterial lectin". Continual searches were made with the same keyword but with an additional source keyword pertaining to each of the bacterial lectin, obtained in the first list through the NCBI search engines Entrez, using GenBank, Swiss-Prot [5], and Protein Data Bank (PDB) [6]. The results of these searches were downloaded onto local machines, along with other available associated details. The data was arranged with the help of Relational Database Management System technique by assigning primary key and foreign key to ensure there is no duplication of data. Any previously unidentified redundancies were removed. For each Entry, basic information pertaining to the lectin name, source, protein, amino acid sequence length, molecular weight, carbohydrate specificity, and PDB identifiers was parsed from the respective entries in various databases.

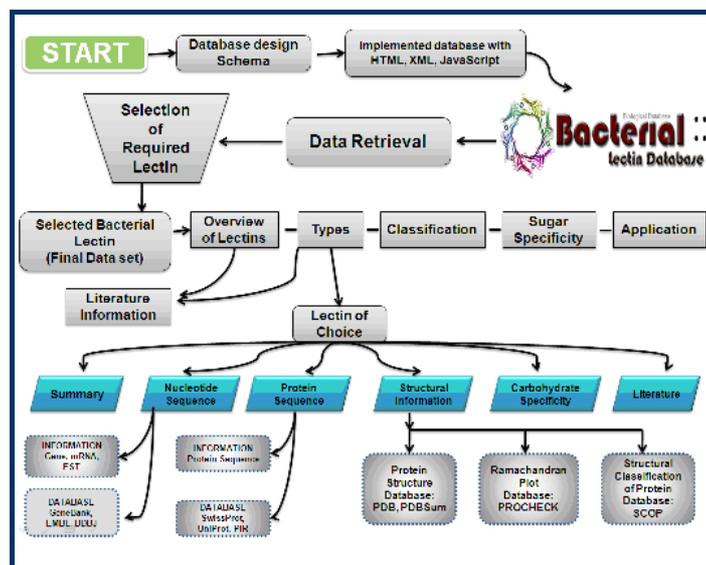
### Database schema

An organized database schema for BacterialLectinDb was designed to collate all the available information about all bacterial lectins as a central repository. The database was designed using HTML, XML. Some modules were implemented using JavaScript and Java applets. The database blueprint was designed to accommodate basic information about lectins viz. the structural details (Fold, family classification, primary structure of their proteins, the relevant nucleotide sequences) and also the details of the carbohydrate specificities. Derived data features such as domain boundaries, active site residues, structure prediction, fold classification, and phylogenetic results were stored in various file formats. Many of the links were connected to different databases to acquire more recent data. The database schema was designed in such a way that it can also enable an easy addition of new information about bacterial

lectins in the future. Moreover, it can also support addition of information on lectins from other sources such as plants, fungi, and viruses that are planned to be integrated in the future as a "Comprehensive Lectin Database".

### Database assembly

A flowchart depicting the methodology used in constructing BacterialLectinDb is illustrated in (Figure 2). The pipeline to construct the database has been automated in parts and also manually checked at specific stages, to ensure of minimizing errors in the database.



**Figure 2:** A schematic representation of the pipeline used in the development of BacterialLectinDb.

### Database search and contents

Using keywords "bacterial lectin" search through NCBI resulted in an initial list of 7 different types of bacterial lectins comprising of a total no. of 65 lectins viz. (i) ADO-ribosylating toxin: 27; (ii) Bacterial neurotoxin: 12; (iii) Cyanovirin: 5; (iv) Fucotoxin: 7; (v) Galactophilic lectin: 3; (vi) Other Bacterial toxin: 2; (vii) Pilli adhesion: 9. Further information was collected by repeated searches for each bacterial lectin individually using different databases. The database assembly was designed as discussed in methodology, it is then run and contents were analyzed.

### Database content

There are some specific databases which may provide information on bacterial lectins, publicly accessible computerized repositories of proteins, genes and other biological informations (e.g. Swiss-Prot, GenBank and EMBL) have served as integral research components for several decades. However, the utility of these repositories seems to be critical when the new forms of biological data are similarly archived and made available. Therefore, the public availability of bacterial lectins data would enable independent validation of published findings, which are currently problematic because the complete, untransformed and fully annotated data sets are not yet placed into the public domain. Indeed, a consensus opinion has emerged supporting the need to couple publication of bacterial lectins data with web-based availability of data to facilitate confirmation of findings. The database integrates lectins related data from sequence (GenBank, EMBL, and

DDBJ), taxonomy, and PDB, CATH, SCOP and MSD databases. Besides these, it provides functional information for all lectins fetched from literature, functional annotations derived from Swiss-Prot and GenBank function cards as well as from Protein Information Resource (PIR). The functional information pertaining to carbohydrate specificities, blood group specificities, and biological processes have been mapped. Furthermore, each lectin entry in the database has been tagged with structural annotation in a layered fashion, depending upon the extent of information available about them. The next level of information in the database pertains to the known function(s) of the lectins. The information spans a wide hierarchical range, starting from individual monosaccharide specificities to larger roles in various cellular events. Carbohydrate specificities obtained from the literature have often pertained to specific Bacterial lectin and a general functional annotation to lectins of a given Bacterial has also been provided. In our database some of the functions of lectins and the broad potential applications they lead have been provided as a different section of the database.

## Utility:

A number of repositories of Lectins data had been developed by government organizations viz. Plant Lectin Database of the Indian Institute of Sciences, Bangalore, India and The Centre de Recherche sur les Macromolécules Végétales, a research unit of the Centre National de la Recherche Scientifique (<http://www.cermav.cnrs.fr/cgi-bin/lectins/list.pl?menuitem=1>). Recently we have developed a phyto chemical database for diabetes named "Phyto-Mellitus" [7] and also an animal lectin Database "AnimalLectinDb" [8] providing important information about the chemical nature of various plant products and structural-functional annotations of various animal lectins, respectively. Our studies with analysis of databases for cystic fibrosis and mutational analysis of H5N1 and H1N1 were found significant for ascertaining test systems and provide an insight to understand the role of glycoprotein/lectins for vaccine development, respectively [9, 10].

Lectins bind to various sugars in a highly selective manner. This selectivity enables lectins to display many significant biological activities and a variety of functions. Mannose specific bacterial lectins (Type 1) possess an extended combining site corresponding to an oligosaccharide and preferentially bind to the carbohydrate moiety i.e. oligomannose (hybrid type). Type 1 bacterial surface lectin/ fimbriae possess a hydrophobic region close to the carbohydrate-binding site, since aromatic alpha-mannosides inhibit strongly (up to 1000-times more than methyl alphanmannoside) the agglutination of yeasts by the bacteria and the adherence of the latter to pig ilea epithelial cells [11]. The combining sites of Type 1 fimbriae of the *Salmonella* and of other enteric bacteria are different from those of *Escherichia coli* in that they are smaller and do not possess a hydrophobic region [12].

Molecular recognition is indeed a key event in many biological processes. The mannose specific lectins also act as recognition molecules in lectinophagocytosis (i.e. phagocytosis of the

bacteria in the absence of opsonins) by mouse, rat and human peritoneal macrophages, and human polymorphonuclear leukocytes [4]. Because lectins form an important class of carbohydrates containing molecules, therefore, studies with lectins have proved invaluable in the understanding of molecular mechanisms of various cellular processes and deciphering the code contained within the sugar molecules [13]. The biological role of lectin-carbohydrate interaction and proteomics research will lead to a deeper understanding of the molecular basis of infectious diseases, and perhaps also to new approaches for their prevention [14].

BacterialLectinDb provides an easy-to-use web interface with flexibility to select for an entry or a collective set of entries matching user's criteria such as name of the bacteria, sequence class etc. It also provides information such as Express Sequence Tags and mRNA detail along with complete nucleotide sequence and Protein sequence from different databases. Microscopic details of protein such as visualization of dihedral angles  $\psi$  against  $\phi$  of amino acid residues in protein structure, ligand information and domain etc are one of the most important data sets available in Bacterial lectin database, which is not available in any other lectin database. In the pursuit of all the applications, it is our belief that this database will serve as a useful repository of manually curreted information pertaining to sequence, structure, and function, all integrated into a single framework

## Conflict of interest:

The authors have no conflicts of interest to declare.

## References:

- [1] Sharon N *et al.* *Ciba Found Symp.* 1981 **80**: 119 [PMID: 6114817]
- [2] Pokorna M *et al.* *Biochemistry.* 2006 **45**: 7501 [PMID: 16768446]
- [3] Greco A *et al.* *Proc Nat Acad Sci USA.* 2006 **13**: 460 [PMID: 16622409]
- [4] Zhang LH *et al.* *Yi Chuan.* 2007 **29**: 675 [PMID: 17650483].
- [5] Bairoch A & Boeckmann B, *Nucleic Acids Res.* 1992 **20** : 2019 [PMID: 1598233]
- [6] Berman HM *et al.* *Nucleic Acids Res.* 2000 **28**: 235 [PMID: 10592235]
- [7] Middha SK *et al.* *Bioinformatics.* 2009 **4**: 78 [PMID: 20198174]
- [8] Kumar D & Mittal Yashoda, *Bioinformatics.* 2011 **6**: 134 [PMID: 21887013]
- [9] Kumar D & Mittal Yashoda, *Proc BIOCOMP'07.* 2007 **2**: 727
- [10] Mittal Yashoda *et al.* *J C Intelligence in Bioinformatics.* 2010 **3**: 91
- [11] Lacy DB *et al.* *Nat Struct Biol.* 1998 **5** : 898 [PMID: 9783750]
- [12] Firon N *et al.* *Carbohydrate Research.* 1983 **120**: 235 [PMID: 6138149]
- [13] Agrawal BB & Goldstein IJ, *Biochem J.* 1965 **96**: 23C [PMID: 5862398]
- [14] Sharon N, *Biochim Biophys Acta.* 2006 **1760**: 527 [PMID: 16564136]

Edited by P Kanguane

Citation: Kumar & Mittal, *Bioinformatics* 8(6): 281-283 (2012)

**License statement:** This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.