Glycoconj J (2008) 25:191–198
DOI 10.1007/s10719-007-9085-5

CancerLectinDB: a database of lectins relevant to cancer

Deepa Damodaran · Justin Jeyakani · Alok Chauhan · Nirmal Kumar · Nagasuma R. Chandra · Avadhesha Surolia

Received: 28 August 2007 / Revised: 31 October 2007 / Accepted: 1 November 2007 / Published online: 24 November 2007

Abstract The role of lectins in mediating cancer metastasis, apoptosis as well as various other signaling events has been well established in the past few years. Data on various aspects of the role of lectins in cancer is being accumulated at a rapid pace. The data on lectins available in the literature is so diverse, that it becomes difficult and time-consuming, if not impossible to comprehend the advances in various areas and obtain the maximum benefit. Not only do the lectins vary significantly in their individual functional roles, but they are also diverse in their sequences, structures, binding site architectures, quaternary structures, carbohydrate affinities and specificities as well as their potential applications. An organization of these seemingly independent data into a common framework is essential in order to achieve effective use of all the data towards understanding the roles of different lectins in different aspects of cancer and any resulting applications. An integrated knowledge base (CancerLectinDB) together with appropriate analytical tools has therefore been developed for lectins relevant for any aspect of cancer, by collating and integrating diverse data. This database is unique in terms of providing sequence, structural, and functional annotations for lectins from all known sources in cancer and is expected to be a useful addition to the number of glycan related resources now available to the community. The database has been implemented using MySQL on a Linux platform and web-enabled using Perl-CGI and Java tools. Data for individual lectins pertain to taxonomic, biochemical, domain architecture, molecular sequence and structural details as well as carbohydrate specificities. Extensive links have also been provided for relevant bioinformatics resources and analytical tools. Availability of diverse data integrated into a common framework is expected to be of high value for various studies on lectin cancer biology. CancerLectinDB can be accessed through http://proline.physics.iisc.ernet.in/cancerdb.

Keywords Cancer · Lectins · Taxonomic · Biochemical · Domain architecture · Sequence · Carbohydrate specificities · Database

Availability: CancerLectinDB is available freely for academic use from http://proline.physics.iisc.ernet.in/cancerdb, Contact nchandra@serc.iisc.ernet.in for further information.

D. Damodaran · J. Jeyakani · N. Kumar · N. R. Chandra (✉)
Bioinformatics Centre and Supercomputer Education and Research Centre, Indian Institute of Science, Raman Building, Bangalore 560 012, India
e-mail: nchandra@serc.iisc.ernet.in

A. Chauhan · A. Surolia (✉)
Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India

A. Surolia
National Institute of Immunology, Aruna Asaf Ali Marg, New Delhi 110 067, India
e-mail: surolia@nii.res.in

Introduction

Lectins are carbohydrate-binding proteins, ubiquitously distributed in nature, that specifically recognise diverse sugar structures and mediate a variety of cellular processes [11, 27] providing biological scripts to decipher the complex codes embedded in the *glycome*. The involvement of lectins in processes such as cell–cell and host–pathogen interactions, serum-glycoprotein turnover and innate immune responses are of particular relevance to tumour growth and metastatic spread. Several lines of evidence

accumulated in the recent years implicate tumour cell lectins in cellular interactions such as adhesion, cell growth, tumour cell differentiation, and metastasis [15]. Lectins are also known to modify the cell cycle by inducing cell cycle arrest, apoptosis and activation of the caspase cascade. Their effect on the immune system by altering the production of various interleukins is also well documented. There is also data to suggest that some lectins down-regulate telomerase activity and hence inhibit angiogenesis [9, 18]. A natural outcome of these studies has been the application of several lectins as therapeutic agents, preferentially binding to cancer cell membranes or their receptors, causing cancer cell agglutination and thus cytotoxicity, apoptosis, and inhibition of tumour growth [18, 20]. Although lectins seem to have great potential as cancer markers and anticancer agents, several gaps in our knowledge still exist, warranting further research on many fronts. A molecular level understanding of the mechanisms by which these lectins perform these diverse roles, is very important. The recognition of the importance of lectins in cancer related processes has provided further impetus to studying various other lectins, as a result, a vast amount of data has been accumulated in various areas of lectin biology and chemistry in the recent years. The successful completion of the human and several other genome projects has made amino acid sequences of several lectins available. Such vast amount of data on cancer related lectins due to simultaneous progress in diverse areas, makes it difficult and time-consuming to simultaneously comprehend all the various advances, especially because the sources of these various pieces of data are highly diverse as also the areas of specializations they emerge from. In order to achieve an effective use of all the data towards understanding of the function and for any possible application, an organization of these seemingly independent data into a common framework is essential. This can be made possible by collating and integrating these various bits of data using an appropriate framework.

Lectins have been grouped into several broad classes based on their sources and functions. Such features can often be captured more systematically by the corresponding amino acid sequences. Sequences of lectins and their tertiary structures where available, provide a good framework upon which all other data can be integrated. They also provide a basis for a unique classification of this class of proteins [21] (Bettler, Loris and Imberty, http://www.cermav.cnrs.fr/lectines/). Further, chemical and biological data on lectins, as in the case of any family of proteins, when interpreted through their tertiary structures provide the greatest insight into their function and their role in biological systems [8, 24, 27]. The vast number of sequences, a significant amount of biochemical data as well as several crystal structures reported in literature, in fact necessitate a simultaneous analysis of all known members of the family to develop a broader perspective of the functionalities as well as potential uses of these lectins. Recently, we have made a similar effort to develop an integrated database of plant lectins (http://nscdb.bic.physics.iisc.ernet.in) [7], which are often model systems of choice to study the molecular basis of these recognition events [14, 22]. Refining such approaches for collating, integrating, classifying primary data and generating various derived data from several perspectives, we have developed an integrated knowledge base together with appropriate analytical tools for lectins relevant to cancer. While lectindb has information regarding plant lectins and any functional aspects associated with them, CancerLectinDB provides information specifically about lectins from any source with some functional link to cancer. There are also number of useful resources related to glycans that are now available over the internet. Some examples of which are glycan topology, glycan structure, glycan processing pathways, glycosylation sites and glycan binding proteins. This database on lectins specific to cancers will be an useful addition to the rapidly growing glycan and glycomics information.

Database schema and construction

A flowchart depicting the methodology used in constructing CancerLectinDB is illustrated in Fig. 1. A relationally organized database schema was designed to serve as a repository of lectins. The database was implemented using MySQL (http://www.mysql.com), a free-to-use RDBMS while the scripts were written in PERL (http://www.perl.org). CGI and Java applets for some modules were used for client-server programming and a Linux web server was used to deliver the interface. The schema has been designed to accommodate basic information about a lectin, its corresponding sequence and structural details, fold, family classification, and carbohydrate specificity and also enables easy addition of new information in the future. Derived data features such as domain boundaries, active site residues, structure prediction, fold classification, phylogenetic results are stored in various file formats and are processed and accessed through PERL scripts. An easy-to-navigate web interface has been built to allow the user to search through the database, which offers a variety of search options.

Data collection and curation

The initial dataset comprised all proteins from plants, animals, viruses, fungi, alveolata and bacteria most of the entries available in eukaryotic particularly *Homo sapiens* and *Mus musculus.* These were either annotated as one of

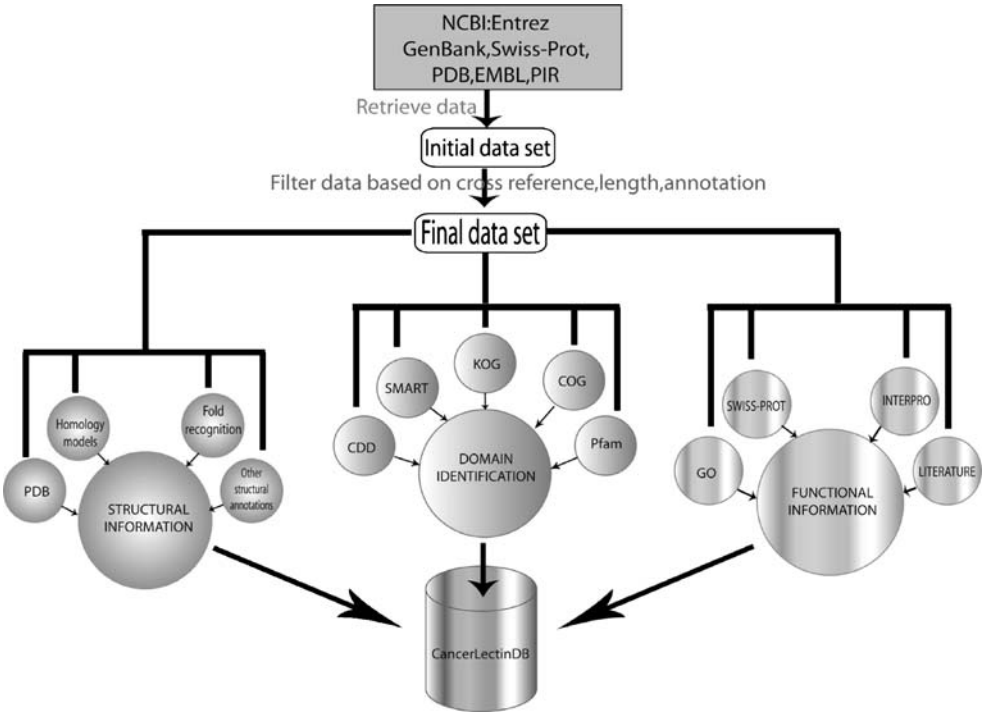

Fig. 1 A schematic representation of the pipeline used in the development of CancerLectinDB

the lectin classes (annexin, lectin, galectin, agglutinin, selectin, pentraxin, ficolin, ricin, collectin, tachylectin) or its receptors and in addition contained a keyword (such as cancer, carcinoma, tumor, malignant and other cancer related terms given in the list of cancers cited by the National Cancer Institute, http://www.cancer.gov/cancertopics/alphalist) referring to cancer from *NCBI protein* database, which resulted in an initial list of 2,548 sequence entries. Entries were also collected on the basis of lectins obtained from various carcinogenic cell or tissue cultures. All the entries that were obtained were grouped together based on the source, lectin class and their accession numbers cross-compared against different databases, so as to remove redundant entries, which resulted in a filtered list of 765 cancer related entries (Table 1). This list was verified for uniqueness and any further redundancies were systematically removed by following the protocol as described for the plant lectin database, which resulted in a final set of 506 sequence entries. This list includes isolectins or even the two chains of hetero-dimers as separate entries. For each entry, basic information pertaining to the lectin name, amino acid sequence length, molecular weight, source and carbohydrate specificity where available, associated disease (OMIM) and PDB identifiers were parsed from the respective entries in various databases. In addition to these a static list of lectins known to be relevant to cancer, but whose sequence could not as yet be mapped to any protein in the public databases is also made available in CancerLectinDB but for obvious reasons have not been analysed or discussed here further.

A comprehensive domain annotation has been carried out by scanning CDD, SMART, Pfam, KOG/COG databases using RPS-BLAST (Cut-offs employed were 30% identity and 60% coverage in length of the known domain sequence). Alignments of isolectins within a given plant were obtained using Clustalw package [26]. Information pertaining to sequence neighbours and active site residues identified by the SAS server based on related structures in PDB, was also obtained. In addition to the precompiled analyses results, dynamic links have been provided for obtaining sequence neighbours for each of the lectin entries against NR, PDB as well as within CancerLectinDB. The sequences have been scanned for possible signal-peptides using SignalP [4]. Functional information pertaining to carbohydrate and blood group specificities obtained through literature searches for each of the entries, have been integrated into the database. Layers of structural information has been obtained from PDB and through sequence analysis of structural templates as described previously for the plant lectin database. Quaternary structure information obtained from PDB, wherever available, have been mapped onto the respective lectin entries. Function annotations from Swissprot/Interpro/Gene ontology [1, 10], where such information was available has also been included. Further, precompiled lists of sequence neighbours for each of the lectin entries in the database, determined through BlastP searches have been made available on the database, besides which dynamic links to repeat such analysis with updated databases, have also been provided. A sequence-based search option that scans a user-input query sequence for

Table 1 Various statistics pertaining to CancerLectinDB construction and content

Dataset construction						
Primary set (lectin sequences from PIR/ PRF/ PDB/ DDBJ/ EMBL/Genbank/ SwissProt)						2548
Filtered set						
(No. of entries after filtering)	Xref redundancy/sequence redundancy					765
	Manual inspection of annotation					506
Final set (CancerLectinDB)						506
Number of lectin receptors in the final set						20
Number of organisms in the final set						37
No. species from various Kingdoms	Plant	Animal	Fungi	Bacteria	Virus	Alveolata
	29	472	2	3	3	1
Domain annotation						
Total Number of Domains in the database						3018
Total/Unique Domains identified from:	CDD 482/62	COG 91/20	KOG 1053/185	SMART 634/57		PFAM 758/125
Td: Sum of unique domains in all domain databases	449 (contains 33 lectin domains) 425 other domains					
Unique domains from Td	33 (contains 16 lectin domains)					
Receptor Domains	107/55					
Structural Annotation						
Structures available through PDB						33
Structure that can be built by homology modelling						226
Structure that can be predicted by fold recognition						233
Unknown structures						18
Receptor Structures	PDB	Homolog PDB	Threader PDB			
	6	8	8			
Distribution of lectin domains across various folds						
Annexin						49
Beta-prism I						3
Beta-trefoil						31
Concanavalin A						45
C-type lectin						84
Knottins (small inhibitors toxins lectins)						17
Other folds						76
Distribution of other domains co-occuring with lectins						31
Receptors with lectin domains						10
Functional Annotation						
No. of lectins with annotation from	InterPro					582
	GO					1644
	PRINTS					118
	PRODOM					58
	PROSITE					338
No. of lectins with known Carbohydrate Specificity						191
Number of OMIM entries						201

similarities within CancerLectinDB has also been incorporated using a Blast interface.

Database content

CancerLectinDb integrates information from Swiss-Prot [2], Genbank [5] and protein sequence databases from NCBI, Blink-sequence distribution across species (http://www.ncbi.nlm.nih.gov/sutils/static/blinkhelp.html), taxonomy database (http://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html/), PDB-protein structure database [6], SCOP-structural classification database [13], CDD [17], Pfam [3], SMART [23], KOG-domain databases [25], SAS-structure annotated sequence database [19] and quaternary structure information from PDB, thus providing a single point for accessing diverse information pertaining to lectins relevant to cancer.

In total, 506 entries were obtained as unique, non-redundant entries spread across 37 different sources (Table 1). Not surprisingly, *H. sapiens* and a close animal model *M. musculus* contained the largest number of cancer related lectin sequences summing up to 307 and 119 respectively. Some other sources such as *Rattus norvegicus* and *Sambucus nigra* also contained several lectins, most of them closely related to each other, thus qualifying as isolectins. Twenty-nine of the entries in the database belong to plants which are also cross mapped to the plant lectin database (LectinDB). These have been known to specifically bind to sugars on cancer cells or inhibit tumor formation or cytotoxicity. An example snapshot of the database querying is shown in Fig. 2a. Multiple sequence alignments of all lectins within each source help in identifying the similarities and the subtle but perhaps

Fig. 2 **a** A snapshot illustrating an example of a search of CancerLectinDB. The information is grouped into sequence-related, structure-related and function-related categories. The information retrieved for MLB1_VISAL Galactose specific lectin I B chain protein from *Viscum album* indicates that it may slow the growth of cancer cells and can have a role in the effective treatment for solid tumors and is also responsible for cell agglutination (lectin activity). This protein has available PDB structures and has two domains belonging to the lectin fold beta trefoil and the co-occurring ribosome inactivating protein. **b** A snapshot of a portion of the multiple alignment of lectins indicated in *Viscum album*. Identical and highly conserved stretches are displayed

important differences among them. A portion of such an alignment for lectins from *Macaca mulatta* is shown as an example (Fig. 2b). While many lectins were made of single domains, several others contained lectin domains as part of their larger polypeptide sequences that contained several other domains as well (for example, lectins from *H. sapiens* and *M. musculus* contained Trypsin-like serine protease and EGF domains in combination with the C-type lectin domains). In yet other cases, multiple copies of the same lectin or two or three different lectin domains were also found (e.g., Ricin repeats were present in *H. sapiens*, *M. musculus* and *S. nigra*). Majority of the sequences contained two or three domains, while there were about 50 sequences that contained more than 12 domains in a single polypeptide chain, of which one or more are lectin domains. Three thousand eighteen domains were identified in the dataset of 506 sequences, which contained 33 unique domain types, of which there were 16 unique lectin domains. Identification of domains obtained through a comprehensive analysis, integrated into the database, will help greatly in understanding domain architectures and their functional implications. Understanding lectin domains that co-exist either with other lectin domains or with any other protein provides a first step in gaining insights into their larger biological roles.

Of the 506 entries, 33 polypeptide chains, referring to lectins from eight sources, had structures determined experimentally, available through PDB. Structures of 226 more sequences in the database can be built by homology modelling using structural templates from PDB. Further overall structures of 233 sequences could be predicted by fold recognition methods. For 18 lectin sequences, no structural annotation has been possible with the existing databases, which has been indicated appropriately. A fold-based classification, utilizing the fold information of either the known lectin structures or high confidence structural templates of lectin sequences, from SCOP database, has been carried out. This classification scheme groups known lectin types into six major folds, which are the Annexin, Knottins, β-trefoil, β-Prism-I, Concanavalin A and C-type lectin (Fig. 3). Further, multiple alignment of lectins within each fold class have been carried out, followed by phylogenetic analyses, which are useful to understand the extent of divergence in detail, and hence subtle but definite functional differences/adaptations, within each fold. Beyond the overall fold information, the structural annotations also comprise information regarding quaternary structure of the different lectins in the database.

The next level of information in the database pertains to the known function(s) of the lectins. Here again, the information spans a wide hierarchical range, starting from individual monosaccharide specificities to larger roles in various cellular events. Analysis of the lectins in the databases using SAS (Sequence annotated by Structure) bioinformatics tool, helped in identifying the conservation of active residues in each class of lectins. Functional

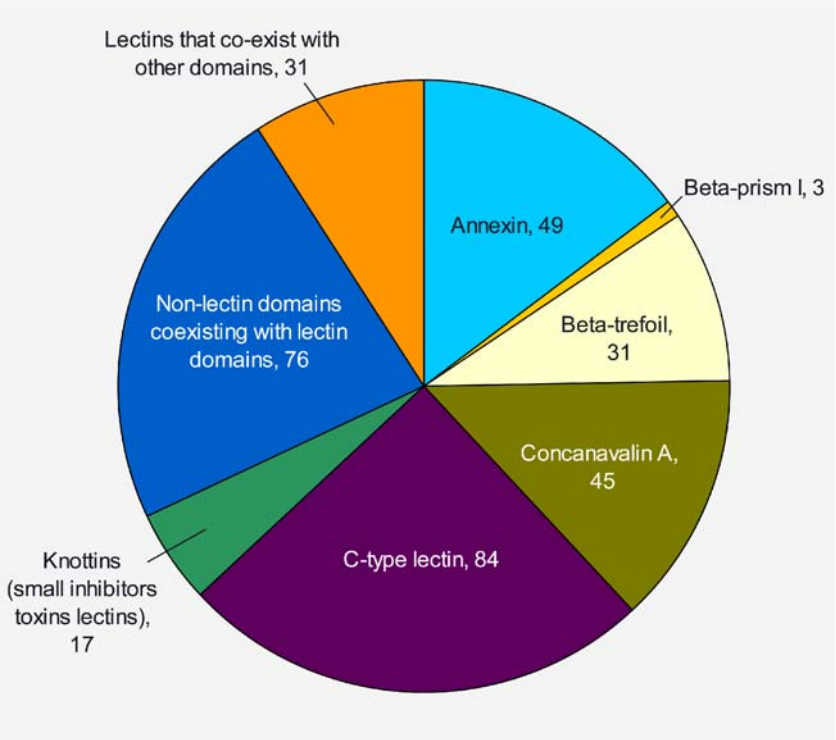

Fig. 3 The fold-profile of structures in the database across the different folds

annotations of the lectins have also been derived from Swissprot, Genbank function cards as well as from Interpro and GO. For example, a 111 amino-acid Sialic acid-binding lectin (P18839), has been annotated to agglutinate preferentially a large variety of tumor cells. It has pancreatic ribonuclease domain and has RNase A-like fold. The functional information obtained from literature, indicates this molecule in frog eggs may be involved in the fertilization and development of the frog embryo and lectin preferentially agglutinate a large variety of tumor cells, but it does not agglutinate non-transformed cells and erythrocytes. Cross references from Interpro, Pfam, Smart, Prosite, Prodom are included.

Utility and discussion

CancerLectinDB provides an easy-to-use web interface with flexibility to select an entry or a collective set of entries matching users' criteria such as name of the plant, sequence class, domain type, structural class, quaternary type, fold class or carbohydrate specificity, keyword searches. Additional searching schemes provided in the database are based on the (a) Conventional classification of lectins, (b) Lectin Receptor (c) Cancer type (d) Function/ application.

Several independent investigations have indicated that lectins present on the surface of cancer cells might be involved in tumor cell aggregation, embolization within blood vessels and adhesion to capillary epithelial cells, all of which are required for cancer metastasis [12]. Yet several other investigations have indicated that some other lectins, are capable of inhibiting cell adhesion, proliferation and colony formation [16] and hence of immense potential in understanding and managing cancers. These lectins are also known to stimulate the immune system, modulate growth and apoptosis of premalignant and malignant cells *in vitro* as well as *in vivo*. Many pre-cancer and cancerous cells are known to exhibit altered patterns of glycoconjugate expressions on their surfaces. Such changes are easily tractable by the use of appropriate lectins. In summary, the involvement of lectins in cancer is manifold, which can be exploited for our benefit. Some examples of potential applications of lectins in cancer are (a) Clinical diagnosis I: diagnosis of different types of cancer (b) Clinical diagnosis II: staging the progression in metastatic phenotypes through analysis of types and extents of lectin expression, (c) Laboratory use-I: as a probe to study various aspects of cell aggregation, attachment to substratum, anchorage independent growth and colonization, (d) Laboratory use-II: apart from their use as molecular biology and biotechnology (tools for protein purification, for e.g., in affinity chromatography), (e) Laboratory use-III: mapping and analysis of oligosaccharide structure, (f) Clinical use: mitogenic stimulants, inititating apoptosis and inhibiting metastasis, and (g) Drug delivery and targeting (lectin-mediated drug targeting/use of lectins in drug delivery to oral mucosa). In the pursuit of the above applications, it is our belief that this database will serve as a useful repository of manually curated information pertaining to sequence, structure and function, all integrated into a single framework.

Acknowledgements Financial support from Department of Biotechnology, Govt. of India, is gratefully acknowledged. Use of facilities at the Supercomputer Education & Research Centre, Bioinformatics Centre and Interactive Graphics facility supported by DBT is also acknowledged.

References

1. Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Birney, E., Biswas, M., Bucher, P., Cerutti, L., Corpet, F., Croning, M.D., Durbin, R., Falquet, L., Fleischmann, W., Gouzy, J., Hermjakob, H., Hulo, N., Jonassen, I., Kahn, D., Kanapin, A., Karavidopoulou, Y., Lopez, R., Marx, B., Mulder, N.J., Oinn, T.M., Pagni, M., Servant, F., Sigrist, C.J., Zdobnov, E.M.: InterPro Consortium, InterPro: an integrated documentation resource for protein families, domains and functional sites. Bioinformatics, **16**, 1145–1150 (2000)
2. Bairoch, A., Boeckmann, B.: The SWISS-PROT protein sequence data bank. Nucleic Acids Res. **20**, 2019–2022 (1992)
3. Bateman, A., Coin, L., Durbin, R., Finn, R.D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E.L., Studholme, D.J., Yeats, C., Eddy, S.R.: The Pfam protein families database. Nucleic Acids Res. **32**, D138–D141 (2004)
4. Bendtsen, J.D., Nielsen, H., von Heijne, G., Brunak, S.: Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. **340**, 783–795 (2004)
5. Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Wheeler, D.L.: GenBank. Nucleic Acids Res. **33**, D34–D38 (2005)
6. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E.: The protein data bank. Nucleic Acids Res. **28**, 235–242 (2000)
7. Chandra, N.R., Kumar, N., Jeyakani, J., Singh, D.D., Gowda, S. B., Prathima, M.N.: Lectindb: a plant lectin database. Glycobiology **16**, 938–946 (2006)
8. Chandra, N.R., Prabu, M.M., Suguna, K., Vijayan, M.: Structural similarity and functional diversity in proteins containing the legume lectin fold. Protein Eng. **14**, 857–866 (2001)
9. Choi, S.H., Lyu, S.Y., Park, W.B.: Mistletoe lectin induces apoptosis and telomerase inhibition in human A253 cancer cells through dephosphorylation of Akt. Arch Pharm Res. **27**, 68–76 (2004)
10. Harris, M.A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., Eilbeck, K., Lewis, S., Marshall, B., Mungall, C., Richter, J., Rubin, G.M., Blake, J.A., Bult, C., Dolan, M., Drabkin, H., Eppig, J.T., Hill, D.P., Ni, L., Ringwald, M., Balakrishnan, R., Cherry, J.M., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S., Fisk, D.G., Hirschman, J.E., Hong, E.L., Nash, R.S., Sethuraman, A., Theesfeld, C.L., Botstein, D., Dolinski, K., Feierbach, B., Berardini, T., Mundodi, S., Rhee, S. Y., Apweiler, R., Barrell, D., Camon, E., Dimmer, E., Lee, V., Chisholm, R., Gaudet, P., Kibbe, W., Kishore, R., Schwarz, E.M., Sternberg, P., Gwinn, M., Hannick, L., Wortman, J., Berriman, M., Wood, V., de la Cruz, N., Tonellato, P., Jaiswal, P., Seigfried,

T., White, R.: Gene Ontology Consortium. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res. **32**, D258–D261 (2004)
11. Lis, H., Sharon, N.: Lectins: carbohydrate-specific proteins that mediate cellular recognition. Chem. Rev. **98**, 637–674 (1998)
12. Liu, F.T., Rabinovich, G.A.: Galectins as modulators of tumour progression. Nat. Rev., Cancer **5**, 29–41 (2005)
13. Lo Conte, L., Ailey, B., Hubbard, T.J., Brenner, S.E., Murzin, A. G., Chothia, C.: SCOP: a structural classification of proteins database. Nucleic Acids Res. **28**, 257–9 (2000)
14. Loris, R., Hamelryck, T., Bouckaert, J., Wyns, L.: Legume lectin structure. Biochim. Biophys. Acta **383**, 9–36 (1998)
15. Lotan, R., Raz, A.: Lectins in cancer cells. Ann. N.Y. Acad. Sci. **551**, (1), 385–396 (1988)
16. Mannori, G., Santoro, D., Carter, L., Corless, C., Nelson, R.M., Bevilacqua, M.P.: Inhibition of colon carcinoma cell lung colony formation by a soluble form of E-selectin. Am. J. Pathol. **151**, 233–243 (1997)
17. Marchler-Bauer, A., Panchenko, A.R., Shoemaker, B.A., Thiessen, P.A., Geer, L.Y., Bryant, S.H.: CDD: a database of conserved domain alignments with links to domain three-dimensional structure. Nucleic Acids Res. **30**, 281–283 (2002)
18. De Mejia, E., Prisecaru, V.: Lectins as bioactive plant proteins: a potential in cancer treatment. Crit. Rev. Food Sci. Nutr. **45**, 425–445 (2005)
19. Milburn, D., Laskowski, R.A., Thornton, J.M.: Sequences annotated by structure: a tool to facilitate the use of structural information in sequence analysis. Protein Eng. **11**, 855–859 (1998)
20. Nakahara, S., Oka, N., Raz, A.: On the role of galectin-3 in cancer apoptosis. Apoptosis **10**, 267–275 (2005)
21. Peumans, W.J., Van Damme, E.J., Barre, A., Rouge, P.: Classification of plant lectins in families of structurally and evolutionary related proteins. Adv. Exp. Med. Biol. **491**, 27–54 (2001)
22. Rudiger, H.: Plant lectins—more than just tools for glycoscientists: occurrence, structure, and possible functions of plant lectins. Acta Anat. (Basel) **161**, (1–4), 130–152 (1998)
23. Schultz, J., Milpetz, F., Bork, P., Ponting, C.P.: SMART, a simple modular architecture research tool: identification of signaling domains. Proc. Natl. Acad. Sci. USA **95**, 5857–5864 (1998)
24. Sharma, V., Surolia, A.: Analyses of carbohydrate recognition by legume lectins: size of the combining site loops and their primary specificity. J. Mol. Biol. **267**, 433–445 (1997)
25. Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., Krylov, D.M., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., Rao, B.S., Smirnov, S., Sverdlov, A.V., Vasudevan, S., Wolf, Y.I., Yin, J.J., Natale, D. A.: The COG database: an updated version includes eukaryotes. BMC Bioinformatics **4**, 41 (2003)
26. Thompson, J.D., Higgins, D.G., Gibson, T.J.: CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighing, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**, 4673–4680 (1994)
27. Vijayan, M., Chandra, N.: Lectins. Curr. Opin. Struct. Biol. **9**, 707–714 (1999)