

Laminin database: a tool to retrieve high-throughput and curated data for studies on laminins

Daiane C. F. Golbert^{1,2}, Leandra Linhares-Lacerda², Luiz G. Almeida¹, Eliane Correa-de-Santana², Alice R. de Oliveira¹, Alex S. Mundstein¹, Wilson Savino^{2,*} and Ana T. R. de Vasconcelos^{1,*}

¹Bioinformatics Laboratory, National Laboratory of Scientific Computation, Ave. Getúlio Vargas 333, 25651-075, Petrópolis and ²Laboratory on Thymus Research, Oswaldo Cruz Institute, Oswaldo Cruz Foundation, Ave. Brasil 4365, 21045-900, Rio de Janeiro, Brazil

Received August 12, 2010; Revised October 12, 2010; Accepted October 13, 2010

ABSTRACT

The Laminin(LM)-database, hosted at <http://www.lm.incc.br>, is the first database focusing a non-collagenous extracellular matrix protein family, the LMs. Part of the knowledge available in this website is automatically retrieved, whereas a significant amount of information is curated and annotated, thus placing LM-database beyond a simple repository of data. In its home page, an overview of the rationale for the database is seen and readers can access a tutorial to facilitate navigation in the website, which in turn is presented with tabs subdivided into LMs, receptors, extracellular binding and other related proteins. Each tab opens into a given LM or LM-related molecule, where the reader finds a series of further tabs for 'protein', 'gene structure', 'gene expression' and 'tissue distribution' and 'therapy'. Data are separated as a function of species, comprising *Homo sapiens*, *Mus musculus* and *Rattus norvegicus*. Furthermore, there is specific tab displaying the LM nomenclatures. In another tab, a direct link to PubMed, which can be then consulted in a specific way, in terms of the biological functions of each molecule, knockout animals and genetic diseases, immune response and lymphomas/leukemias. LM-database will hopefully be a relevant tool for retrieving information concerning LMs in health and disease, particularly regarding the hemopoietic system.

INTRODUCTION

Laminins (LM) correspond to a large number of heterotrimeric glycoproteins, playing a major role in several cell functions, including differentiation, proliferation, adhesion and migration. These glycoproteins are composed of various combinations of one alpha, one beta and one gamma chains. Once assembled, these large molecules are 400–900 kDa in molecular mass and exhibit a cross or T shape. To date, five α , four β and three γ chains have been identified, each one representing a given gene product, resulting in 16 known LMs (LMs 1–15) in mammals, each one bearing varying but large numbers of glycosylation sites (1–3). Considering the complexity of LMs, a nomenclature was first proposed in 1994 (4); being evolved for a second and simpler nomenclature presently used, in which each α , β and γ chain is identified by an Arabic number (3). Accordingly, the LM originally described as LM-1 is formed by the trimer $\alpha 1\beta 1\gamma 1$ and is presently named LM-111. Actually, this LM was the first to be identified, >30 years ago by Rupert Timpl in Germany (5). Since then a large number of isoforms have been described, with specificities in terms of tissue distribution as well as cellular functions (3,6). In addition to binding to other extracellular matrix proteins, LMs bind specific cell membrane receptors. There are at least 11 integrins that have been reported as LM receptors, such as $\alpha 1\beta 1$, $\alpha 2\beta 1$, $\alpha 2\beta 2$, $\alpha 3\beta 1$, $\alpha 6\beta 1$, $\alpha 6\beta 4$, $\alpha 7\beta 1$, $\alpha 9\beta 1$, $\alpha \nu\beta 3$, $\alpha \nu\beta 5$ and $\alpha \nu\beta 8$ (6–8). Taking together, one can easily realize the high degree of biological complexity in LM-mediated interactions, in both health and disease.

Worldwide research into many aspects of LMs is rapidly growing. Such a notion can be applied for the study of LMs in general, as well as more specifically

*To whom correspondence should be addressed. Tel: +55 24 2233 6065; Fax: +55 24 2233 6124; Email: atr@incc.br
Correspondence may also be addressed to Wilson Savino. Tel: +55 21 3865 8101; Fax: +55 21 3865 8250; Email: w_savino@hotmail.com

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

when we evaluate the expression and role of LMs in the hemopoietic system. For example LM-mediated interactions are relevant for the entrance of T cell precursors into the thymus (9,10), the migration of developing thymocytes, both in mice and humans (11–13), as well as in peripheral lymphoid organs (14,15). Also, activated T cells use LM receptors to migrate, so that effector immune function in rejection of heart grafts can be abrogated by blocking LM- α 6 β 1 interaction with antibodies specific for the ligand or the corresponding receptor (16,17). In fact, blockade of the α 6 β 1 receptor also prevented neutrophils from crossing the basement lamina (18). A role for LM isoforms (LMs-411 and 511) has also been demonstrated in leukocyte extravasation in the central nervous system (19,20).

Currently, despite the considerable amount of information on these proteins, the data remains scattered in the literature and in a variety of databases. This prompted us to build up a LM-database, which will hopefully be useful for the scientific community interested in the field, to easily retrieve the data available, so far dispersed in the net. We expect that the LM-database will be a relevant tool for retrieving information of LMs in health and disease, particularly in relation to the physiology of the hemopoietic system and related pathological dysfunctions.

LM-DATABASE GENERAL FORMAT AND IMPLEMENTATION

The LM-Database home page provides a tab with Pathway links for public databases showing the LM molecule in dynamic graphical models. It is structured as follows: in the menu bar, the rationale for the database is presented with tabs sub-divided into LMs, receptors, extracellular LM-binding proteins and other related proteins. Each tab opens into a given molecule (e.g. LM-111). When clicking onto the specific protein, the reader finds a series of further tabs for the protein, gene structure, gene expression and tissue distribution, as well as therapy. Data are separated as a function of species.

Part of the data inserted for each molecule, is carefully curated and annotated. In this respect, all links manually annotated will be periodically updated.

Furthermore, there is a direct link to PubMed, which can be then consulted in a specific way, in terms of the biological functions of the given molecule, knockout animals and genetic diseases, immune response (further filtered as B cells, T cells, autoimmune diseases and Inflammation) and lymphomas/leukemias. The last tab refers to results generated in the context of the research consortium. Lastly, in the database project overview, there is a tutorial (help) to guide the reader, so that to facilitate retrieval of information within the database.

The LM-database was created in the context of a multicentric project dealing with LM in the hemopoietic system. Accordingly, the research groups involved have generated and will generate significant amount of data that will be part of the database. To highlight this issue, in addition to the subjects listed above, there is one in which references derived from the research consortium

are available, as PMIDs, which in turn are directly linked to the abstract found in the PubMed database.

In a second vein, we found potentially useful for the readers that in the main menu of the website, they can have access to the corresponding Brazilian Laboratories and respective researchers. Also, there is a direct contact link to the official Email address to be used by any reader that intends to get any specific information and/or provide new ideas or criticism on the structure and functioning of the database.

Since there have been different names for various LMs, we added a specific tab in the main menu in which the reader can have direct access to the previous and the present LM nomenclature, as well as other names originally provided for some of the isoforms.

In the main menu, we can also access a number of links related to Biocarta and Kegg databases. These links will provide summarizing data on metabolic pathways; most of them being triggered by LM ligation on the cell surface.

As stated above, the LM-database appears in the context of a research network on LM and hemopoietic system, with new information being generated. For that, in a restricted area, the database will be also used for sharing information of ongoing research being done in each Laboratory. Nevertheless, once scientific articles are accepted for publication and appear in the PubMed, the corresponding PMIDs will be made immediately accessible to the readers.

The LM-database is based on the relational database MySQL and your web site is operated under the WWW server Apache with FastCGI. In its construction, three programming languages are used: The languages Perl and Shell Script was applied for the implementation of scripts that carry out the search, download, parse of the data files (text and xml) and import data. The languages Javascript and Perl were used for the implementation of the web site, through the JQuery (<http://www.jquery.com>) library and the Catalyst framework (<http://www.catalystframework.org>). The scripts that carry out the search and import of data are run periodically to the database and are kept updated. A restricted access interface on the web site is available for the insertion of the manually curated information. Lastly, the LM-database is hosted at <http://www.lm.lncc.br> and is freely accessible.

PROTEINS PRESENT IN THE DATABASE

A list of protein composed by LMs, integrin type and non-integrin type receptors, extracellular LM-binding proteins and other LM-related proteins was first compiled manually from the literature and the resultant LM-database comprises 42 proteins. This list will be updated periodically in accordance with the literature.

The interface of LM-database was structured so that the information on each protein is displayed in a straight format. Each entry is listed according to the official protein name (the name described in the literature), although other names (including previous names) are provided. The information is further sorted into domains, displayed as tabs sorted by subject: Summary,

Protein, Gene structure, Gene Expression and Tissue distribution, Therapy and PubMed.

For each protein, the information is allocated into taxonomic division, *Homo sapiens*, *Mus musculus* and *Rattus norvegicus* and when the given protein is a heterotrimer or a heterodimer, as LMs and integrins, the information is provided for each polypeptide chain composing the given protein.

The domains have been filled in using automatic recovery from public databases, providing computationally derived links, together with manual annotation and curation of each information.

THE PROTEIN TAB

The 'Protein' tab contains general information extracted from public protein databases and from literature, including name from UniProt (21) and RefSeq accession numbers (22). Important protein information are provided, as 2D gel link, recovered from UniProt (21), amino acid modifications shown using filters in agreement with each modification found on protein, data about folding linked for experimentally-determined structures of proteins available on PDB (23) and known protein domains linked to Pfam (24) and Interpro (25) databases. It also contains manually annotated sections on protein-protein interactions and protein function.

THE GENE STRUCTURE TAB

The 'Gene' *structure* tab contains information extracted from NCBI Entrez Gene database including mRNA RefSeq accession numbers (21) and chromosome location. In addition, this tab provides data concerning known homologs among the annotated genes of sequenced eukaryotic genome on NCBI, linked to HomoloGene (26), exon-intron structure and splicing variants available from Ensembl (27), as well as links to the genome browsers MapViewer (26), UCSC Genome Browser (28) and Single Nucleotide Polymorphisms (29), when available. It also contains manually annotated sections on gene identification and gene structure and chromosomal localization.

GENE EXPRESSION AND TISSUE DISTRIBUTION TAB

The 'Gene expression' and 'protein distribution' tab also provides general information on LMs. In terms of gene expression, data were sub-divided in Microarray and Expressed Sequence Tags (EST). The Microarray is linked to the Gene Expression Atlas database (30), which provides information concerning gene expression in different biological and experimental conditions. The EST data are linked to UniGene (26). Additionally, the expression and localization of proteins in a large variety of normal tissues, cancer cells and cell lines, can be visualized by immunohistochemistry (IHC). Herein, the data on humans are linked to the human protein atlas (30). The data regarding mice is linked to the Mouse Genome

Informatics (MGI) (31), whereas tissue distribution in the rat model is manually annotated from literature.

THE THERAPY TAB

This tab has annotated information concerning the use of LM and/or LM receptors as targets for therapeutic interventions. Additionally, we provide information related to commercial and non-commercial antibodies and inhibitors to each protein. The inhibitor list comprises inhibitors of biosynthesis, protein-receptor interaction and small RNA involved with mRNA silencing.

THE PUBMED TAB

Relevant literature has been automatically retrieved from PubMed (26), manually curated and included on the *PubMed* tab. This tab also provides a structure based on filters where the papers are assembled as a function of important subjects related with the hematopoietic system. The filter topics are physiological functions, knockout animals, innate and adaptive immune responses, leukemias/lymphomas and others. The subject immune response is further subdivided into: B cells, T cells, auto-immune diseases and inflammation. Furthermore, the reader within the PubMed tab, the readers find a specific link in which articles published by members of the consortium can be consulted. The links to PubMed references are provided through the corresponding PMID numbers.

DISCUSSION AND PERSPECTIVES

Despite the huge amount of information concerning extracellular matrix ligands and receptors, the organization of such literature as Databases is, surprisingly, very incipient. Actually, the three databases we could find on this field referred to collagen-related genetic diseases (32–34). In this context, the LM-database is the first one that deals specifically with a non-collagenous extracellular protein family and the first database related to extracellular matrix, providing information of both physiology and pathology.

In this context, the LM-database was a build up to allow a more objective and easier way to deal with the enormous amounts of information so far available on LMs, their receptors and their functions. In addition to the general molecular biology, genetics and biochemistry of this extracellular matrix protein family, literature on the role of LMs, particularly regarding the hemopoietic system, has been gathered together and sub-divided in more particular aspects.

In conclusion, we expect that LM-database will be a relevant tool for retrieving information of LMs in health and disease, particularly in relation to the hemopoietic system. In this regard, since it is clear that knowledge on various aspects of LMs and LM-related molecules is continuously expanding, we plan to update the LM-database by adding new, curate and annotated information as it appears in the literature.

FUNDING

Brazilian grant (Pronex) conjointly provided by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (Faperj); other Brazilian institutions, including the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), Laboratório Nacional de Computação Científica (LNCC), Oswaldo Cruz Foundation (Fiocruz) and Instituto de Metrologia (Inmetro), also contributed for funding the project, either directly or through scholarships. Funding for open access charge: LNCC.

Conflict of interest statement. None declared.

REFERENCES

- Miner, J.H. and Yurchenco, P.D. (2004) Laminin functions in tissue morphogenesis. *Annu. Rev. Cell Dev. Biol.*, **20**, 255–84.
- Miner, J.H. (2008) Laminins and their roles in mammals. *Microsc. Res. Tech.*, **71**, 349–56.
- Aumailley, M., Bruckner-Tuderman, L., Carter, W.G., Deutzmann, R., Edgar, D., Ekblom, P., Engel, J., Engvall, E., Hohenester, E., Jones, J.C. *et al.* (2005) A simplified laminin nomenclature. *Matrix Biol.*, **24**, 326–332.
- Robert, E., Burgeson, R.E., Chiquet, M., Deutzmann, R., Ekblom, P., Engel, J., Kleinman, H., Martin, G.R., Meneguzzi, G., Paulsson, M. *et al.* (1994) A new nomenclature for the laminins. *Matrix Biol.*, **14**, 209–211.
- Timpl, R., Rohde, H., Robey, P.G., Rennard, S.I., Foidart, J.M. and Martin, G.R. (1979) Laminin—a glycoprotein from basement membranes. *J. Biol. Chem.*, **254**, 9933–9937.
- Durbeej, M. (2010) Laminins. *Cell Tissue Res.*, **339**, 259–68.
- Belkin, A.M. and Stepp, M.A. (2000) Integrins as receptors for laminins. *Microsc. Res. Tech.*, **51**, 280–301.
- Barczyk, M., Carracedo, S. and Gullberg, D. (2010) Integrins. *Cell Tissue Res.*, **339**, 269–80.
- Savino, W., Mendes-Da-Cruz, D.A., Smaniotto, S., Silva-Monteiro, E. and Villa-Verde, D.M. (2004) Molecular mechanisms governing thymocyte migration: combined role of chemokines and extracellular matrix. *J. Leukoc. Biol.*, **75**, 951–961.
- Stimamiglio, M.A., Jiménez, E., Silva-Barbosa, S.D., Alfaro, D., García-Ceca, J.J., Muñoz, J.J., Cejalvo, T., Savino, W. and Zapata, A. (2010) EphB2-mediated interactions are essential for proper migration of T cell progenitors during fetal thymus colonization. *J. Leukoc. Biol.*, **88**, 1–12.
- Vivinus-nebot, M., Rousselle, P., Cenciarini, C., Berrih-aknin, S., Spong, S., Nokelainen, P., Cottrez, F., Marinkovich, P. and Bernard, A. (2010) Mature human thymocytes migrate on laminin-5 with Activation of metalloproteinase-14 and cleavage of CD44. *J. Immunol.*, **172**, 1397–1406.
- Drumea-Mirancea, M., Wessels, J.T., Müller, C.A., Essl, M., Eble, J.A., Tolosa, E., Koch, M., Reinhardt, D.P., Sixt, M., Sorokin, L. *et al.* (2006) Characterization of a conduit system containing laminin-5 in the human thymus: a potential transport system for small molecules. *J. Cell Sci.*, **119**, 1396–1405.
- Ocampo, J.S., de Brito, J.M., Correa-de-Santana, E., Borojevic, R., Villa-Verde, D.M. and Savino, W. (2008) Laminin-211 controls thymocyte-thymic epithelial cell interactions. *Cell Immunol.*, **254**, 1–9.
- Gorfu, G., Virtanen, I., Hukkanen, M., Lehto, V.P., Rousselle, P., Kenne, E., Lindbom, L., Kramer, R., Tryggvason, K. and Patarroyo, M.D. (2008) Laminin isoforms of lymph nodes and predominant role of alpha 5-laminin(s) in adhesion and migration of blood lymphocytes. *J. Leukoc. Biol.*, **84**, 701–712.
- Smaniotto, S., Mendes-da-Cruz, D.A., Carvalho-Pinto, C.E., Araujo, L.M., Dardenne, M. and Savino, W. (2010) Combined role of extracellular matrix and chemokines on peripheral lymphocyte migration in growth hormone transgenic mice. *Brain Behav. Immun.*, **24**, 451–461.
- Silva-Barbosa, S.D., Cotta-de-Almeida, V., Riederer, I., De Méis, J., Dardenne, M., Bonomo, A. and Savino, W. (1997) Involvement of laminin and its receptor in abrogation of heart graft rejection by autoreactive T cells from *Trypanosoma cruzi*-infected mice. *J. Immunol.*, **159**, 997–1003.
- Riederer, I., Silva-Barbosa, S.D., Rodrigues, M.L. and Savino, W. (2002) Local antilaminin antibody treatment alters the rejection pattern of murine cardiac allografts: correlation between cellular infiltration and extracellular matrix. *Transplantation*, **74**, 1515–1522.
- Dangerfield, J., Larbi, K.Y., Huang, M.T., Dewar, A. and Nourshargh, S. (2002) PECAM-1 (CD31) hemophilic interaction up-regulates $\alpha 6 \beta 1$ on transmigrated neutrophils *in vivo* and plays a functional role in the ability of $\alpha 6$ integrins to mediate leukocyte migration through the perivascular basement membrane. *J. Exp. Med.*, **196**, 1201–1211.
- Wu, C., Ivars, F., Anderson, P., Hallmann, R., Vestweber, D., Nilsson, P., Robenek, H., Tryggvason, K., Song, J., Korpos, E. *et al.* (2009) Endothelial basement membrane laminin alpha5 selectively inhibits T lymphocyte extravasation into the brain. *Nature Med.*, **15**, 519–527.
- Sorokin, L. (2010) The impact of the extracellular matrix on inflammation. *Nature Rev. Immunol.*, **10**, 712–723.
- UniProt Consortium. (2010) The universal protein resource (UniProt) 2010. *Nucleic Acids Res.*, **38**, D142–D148.
- Pruitt, K.D., Tatusova, T., Klimke, W. and Maglott, D.R. (2009) NCBI reference sequences: current status, policy and new initiatives. *Nucleic Acids Res.*, **37**, D32–D36.
- Velankar, S., Best, C., Beuth, B., Boutselakis, C.H., Cogley, N., Sousa Da Silva, A.W., Dimitropoulos, D., Golovin, A., Hirshberg, M., John, M. *et al.* (2010) The Protein Data Bank in Europe. *Nucleic Acids Res.*, **38**, 308–317.
- Finn, R.D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J.E., Gavin, O.L., Gunasekaran, P., Ceric, G., Forslund, K. *et al.* (2010) The Pfam protein families database. *Nucleic Acids Res.*, **38**, D211–222.
- Hunter, S., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Das, U., Daugherty, L., Duquenne, L. *et al.* (2009) InterPro: the integrative protein signature database. *Nucleic Acids Res.*, **37**, D224–D228.
- Sayers, E.W., Barrett, T., Benson, D.A., Bolton, E., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., Dicuccio, M., Federhen, S. *et al.* (2010) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **38**, D5–D16.
- Flicek, P., Aken, B.L., Ballester, B., Beal, K., Bragin, E., Brent, S., Chen, Y., Clapham, P., Coates, G., Fairley, S. *et al.* (2010) Ensembl's 10th year. *Nucleic Acids Res.*, **38**, D557–D562.
- Rhead, B., Karolchik, D., Kuhn, R.M., Hinrichs, A.S., Zweig, A.S., Fujita, P.A., Diekhans, M., Smith, K.E., Rosenbloom, K.R., Raney, B.J. *et al.* (2010) The UCSC Genome Browser Database: update 2010. *Nucleic Acids Res.*, **38**, D613–D619.
- Kapushesky, M., Emam, I., Holloway, E., Kurnosov, P., Zorin, A., Malone, J., Rustici, G., Williams, E., Parkinson, H. and Brazma, A. (2010) Gene Expression Atlas at the European Bioinformatics Institute. *Nucleic Acids Res.*, **38**, D690–D698.
- Berglund, L., Björling, E., Oksvold, P., Fagerberg, L., Asplund, A., Al-Khalili, S.C., Persson, A., Ottosson, J., Wernérus, H., Nilsson, P. *et al.* (2008) A gene-centric human protein atlas for expression profiles based on antibodies. *Mol. Cell Proteomics*, **10**, 2019–2027.
- Bult, C.J., Kadin, J.A., Richardson, J.E., Blake, J.A. and Eppig, J.T. (2010) Mouse Genome Database Group. (2010) The Mouse Genome Database: enhancements and updates. *Nucleic Acids Res.*, **38**, D586–D592.
- Dalgleish, R. (1998) The Human Collagen Mutation Database. *Nucleic Acids Res.*, **26**, 253–255.
- Bodian, D.L. and Klein, T.E. (2009) COLdb, a database linking genetic data to molecular function in fibrillar collagens. *Hum. Mutat.*, **30**, 946–951.
- Crockett, D.K., Pont-Kingdon, G., Gedge, F., Sumner, K., Seamons, R. and Lyon, E. (2010) The Alport syndrome COL4A5 variant database. *Hum. Mutat.*, **31**, E1652–E1657.