Published online 17 September 2010 *Nucleic Acids Research, 2011, Vol. 39, Database issue* ***D235–D240***
doi:10.1093/nar/gkq830

MatrixDB, the extracellular matrix interaction database

Emilie Chautard[1], Marie Fatoux-Ardore[1], Lionel Ballut[1], Nicolas Thierry-Mieg[2] and Sylvie Ricard-Blum[1,*]

[1]Institut de Biologie et Chimie des Protéines, UMR 5086 CNRS—Université Lyon 1, IFR 128 Biosciences Gerland-Lyon Sud, 7 passage du Vercors 69367, Lyon Cedex 07 and [2]TIMC—IMAG/TIMB, UMR 5525 CNRS—Université Grenoble 1—Faculté de Médecine, 38706 La Tronche Cedex, France

Received July 29, 2010; Revised August 30, 2010; Accepted September 3, 2010

ABSTRACT

MatrixDB (http://matrixdb.ibcp.fr) is a freely available database focused on interactions established by extracellular proteins and polysaccharides. Only few databases report protein–polysaccharide interactions and, to the best of our knowledge, there is no other database of extracellular interactions. MatrixDB takes into account the multimeric nature of several extracellular protein families for the curation of interactions, and reports interactions with individual polypeptide chains or with multimers, considered as permanent complexes, when appropriate. MatrixDB is a member of the International Molecular Exchange consortium (IMEx) and has adopted the PSI-MI standards for the curation and the exchange of interaction data. MatrixDB stores experimental data from our laboratory, data from literature curation, data imported from IMEx databases, and data from the Human Protein Reference Database. MatrixDB is focused on mammalian interactions, but aims to integrate interaction datasets of model organisms when available. MatrixDB provides direct links to databases recapitulating mutations in genes encoding extracellular proteins, to UniGene and to the Human Protein Atlas that shows expression and localization of proteins in a large variety of normal human tissues and cells. MatrixDB allows researchers to perform customized queries and to build tissue- and disease-specific interaction networks that can be visualized and analyzed with Cytoscape or Medusa.

INTRODUCTION

The extracellular matrix is comprised of proteins and complex polysaccharides that are organized in a tissue-specific manner. Major components of the extracellular matrix are collagens [~30% of proteins in humans; (1)], elastic fibers, proteoglycans and glycosaminoglycans. Several extracellular protein families (e.g. collagens, laminins and thrombospondins) form stable multimers in their native state, the multimers being comprised of either identical or different polypeptide chains. The extracellular matrix provides a structural scaffold contributing to the mechanical properties of tissues (2), and is a reservoir of bioactive fragments, called matricryptins, that are released upon limited proteolysis. These fragments exhibit biological and biomolecular recognition properties of their own and regulate a number of physiological and pathological processes including angiogenesis and tumor growth (3). The cohesion of the extracellular matrix is maintained by an intricate interaction network of protein–protein and protein–glycosaminoglycan interactions. These interactions are involved in the formation of supramolecular assemblies such as collagen fibrils and elastic fibers, in tissue architecture, and in cell-matrix interactions that regulate cell growth and behavior. The perturbation of the extracellular interaction network by mutations in genes coding for extracellular proteins lead to several diseases ranging from mild to severe phenotypes [e.g. osteogenesis imperfecta; (4)].

Interactions involving extracellular proteins are poorly represented in existing databases, and protein–glycosaminoglycan interactions are almost absent from databases although they contribute to the structural organization of the extracellular matrix, to the sequestration of growth factors and chemokines within the extracellular matrix, and to signalling at the cell surface (5). Furthermore, interactions involving multimers, which are frequent in the extracellular matrix (collagens, laminins, thrombospondins are trimers), are often reported as interactions established by individual polypeptide chains. This is a concern especially when molecules are heteromultimers. The above reasons prompted us to build an interaction database focused on interactions occurring between extracellular biomolecules

*To whom correspondence should be addressed. Tel: +33 4 37 65 29 26; Fax: +33 4 72 72 26 04; Email: s.ricard-blum@ibcp.fr

[http://matrixdb.ibcp.fr; (6)]. The database has been updated to include additional interaction data, comprehensive extracellular interaction datasets (e.g. the elastic fiber interactome, extracellular interactions of leucine-rich repeat receptors), and new functionalities. MatrixDB is focused on mammalian molecules, but interaction data of a model organism (zebrafish) has been integrated in the updated database. MatrixDB provides direct links to Online Mendelian Inheritance in Man (OMIM), to databases recapitulating data on mutations occurring in genes encoding extracellular proteins, to UniGene and to the Human Protein Atlas that shows expression and localization of proteins in a large variety of normal human tissues, cancer cells and cell lines. MatrixDB allows researchers to perform customized queries and to build tissue- and disease-specific interaction networks.

BIOMOLECULE DATA

We have imported protein data from the UniProtKB/Swiss-Prot knowledgebase (7), and used UniProtKB accession numbers for proteins. We have created specific identifiers for multimers such as collagens, laminins, thrombospondins and integrins using the following format: MULT_x_species (e.g. MULT_3_human for human collagen I). These entries refer to the UniProtKB accession numbers of their constituent polypeptide chains. Complexes corresponding to stable multimers have been created by the IntAct database (European Bioinformatics Institute, UK) (e.g. EBI-2325312 for human collagen I), and MatrixDB identifiers are cross-referenced to these complexes. Protein isoforms are identified by a variant number (VARy), and the full MatrixDB identifier becomes MULT_x_VARy_species (e.g. MULT_4_VAR1_human). Matricryptins are identified as PFRAG_x_species and are cross-referenced to the feature identifier of UniProtKB. For example, the MatrixDB identifier of endostatin, a C-terminal fragment of collagen XVIII, is PFRAG_1_human and it is cross-referenced to the UniProtKB feature identifier PRO_0000005794. Glycosaminoglycans (GAG_x), lipids (LIP_x) and cations (CAT_x) are cross-referenced to ChEBI and KEGG compound databases (8, 9). Besides protein–protein and protein–glycosaminoglycan interactions, MatrixDB reports interactions involving cations (mostly calcium) and lipids because a number of extracellular molecules bind to cations and some of them to lipids. Detailed information on each molecule is displayed on the 'Biomolecule Report Page'.

INTERACTION DATA

MatrixDB is an active member of the International Molecular Exchange (IMEx) consortium (10) and is in charge of the curation of papers published in *Matrix Biology*, a journal focused on the extracellular matrix, since January 2009. MatrixDB has adopted the PSI-MI standards for annotating and exchanging interaction data. Interaction data stored in MatrixDB are (i) experimentally determined in the laboratory using surface plasmon resonance (SPR) binding assays, including protein and glycosaminoglycan arrays probed by SPR imaging (11), (ii) extracted from the literature by manual curation and (iii) imported from other interaction databases belonging to the IMEx consortium [IntAct (12), DIP (13), MINT (14), BioGRID (15)], as well as from the Human Protein Reference Database (16). Imported data are restricted to interactions involving at least one extracellular protein. The extracellular proteins are identified using UniProtKB/Swiss-Prot keywords and Gene Ontology (17), complemented with manual annotations when required. The text files containing known extracellular human proteins, membrane human proteins and secreted human proteins can be freely downloaded from the download page of MatrixDB. Our curation process has followed the MIMIx guidelines [Minimum Information about a Molecular Interaction experiment; (18)] and has been updated to adhere to the IMEx curation rules in 2010. Interaction data curated by MatrixDB are freely available for download in the PSI-MI XML and TAB 2.5 formats (19).

Mammalian interaction data refer to human molecules in order to easily display the list of partners of a given molecule on the 'Biomolecule Report' page (cf. the schematic organization of MatrixDB, Figure 1). Clicking on an interaction gives access to the 'Interaction Report' page where the source of the data (name of the database) and the experiments supporting the interaction are listed along with links to the abstracts of the corresponding papers. The species experimentally used to demonstrate the interaction are indicated on the 'Experiment Report' page with a detailed report of the experiment according to MIMIx or IMEX standards (e.g. interaction detection method, partner detection method, biological and experimental roles of partners, binding sites, kinetics, and affinity when available). MatrixDB is focused on mammalian interactions, but a comprehensive extracellular interaction dataset (69 interactions) of zebrafish has been imported (20,21). We have also curated a recent dataset of the elastic fiber interactome (45 interactions) identified by affinity purification and mass spectrometry (22), and the interactions (~30) established by SPARC, an extracellular protein involved in a number of biological processes. The current release of MatrixDB contains 2174 extracellular matrix interactions including 1836 protein–protein and 119 protein–glycosaminoglycan interactions. We have curated 490 interactions, and 847 experiments from 192 articles, the other interaction data being imported from several databases (Figure 1). Statistics are available on the 'Statistics' page of MatrixDB.

INTEGRATION OF LOCALIZATION AND MUTATION DATA

The 'Biomolecule Report' page contains a direct link to data from the Human Protein Atlas that shows the expression and localization of proteins in a large variety of normal human tissues, cancer cells and cell lines but is not available for downloading (23). We have imported UniGene expressed sequence tag profiles that reflect

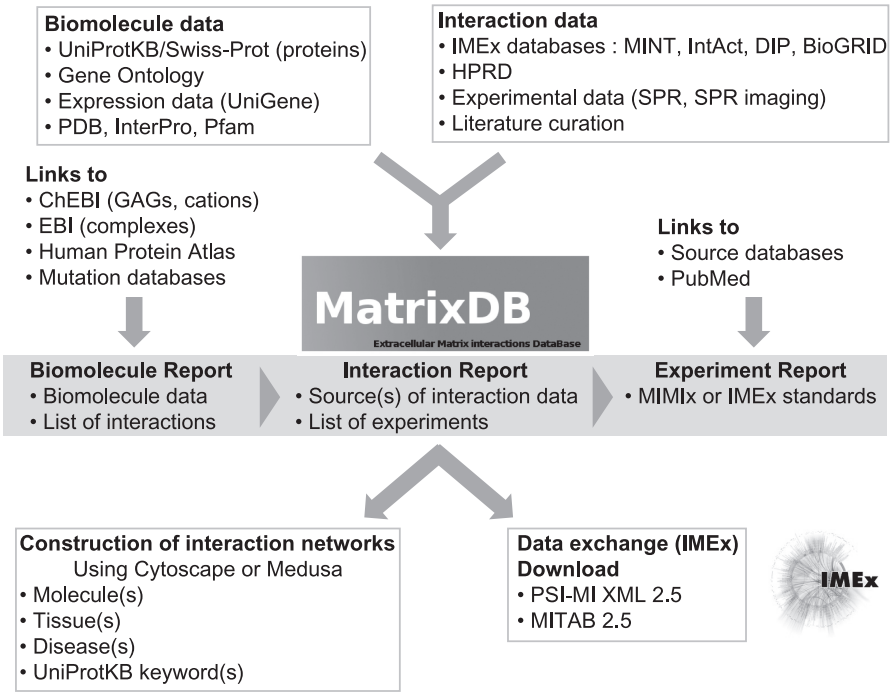

Figure 1. Organization of MatrixDB showing the sources of biomolecule and interaction data, the 'Biomolecule Report', 'Interaction Report' and 'Experiment Report' pages, the links to other web sites, the construction of interaction networks, data formats available for downloading and data exchange with the members of the IMEx consortium.

approximate expression patterns in tissues [http://www.ncbi.nlm.nih.gov/unigene; (24)] in order to create tissue-specific interaction networks.

We have also added on the Biomolecule Report page a link to databases recapitulating data on mutations occurring in the gene encoding the extracellular protein, including the osteogenesis imperfecta consortium [http://oiprogram.nichd.nih.gov/consortium.html; (25)], a database of osteogenesis imperfecta and Ehlers-Danlos syndrome variants [http://www.le.ac.uk/ge/collagen; (26,27)], and to COLdb, a database linking genetic data to molecular function in fibrillar collagens [http://collagen.stanford.edu/; (28)]. On the 'Biomolecule Report' page, and when appropriate, there is a link to the OMIM database of human genes and genetic disorders [http://www.ncbi.nlm.nih.gov/omim; (29)]. These data are used to build disease-specific interaction networks.

MatrixDB: AN EXTRACELLULAR MATRIX WEB SITE

Links to individual extracellular interaction datasets are available on the homepage of MatrixDB. They include the map of candidate cell and matrix interaction domains on the human type I collagen fibril (30), the endostatin interaction network established in our laboratory (11), the elastic fiber interaction network (22) and the cell surface interaction network of neural leucine-rich repeat receptors identified in zebrafish (20,21). Comprehensive extracellular interaction datasets will be curated on a regular basis.

BROWSING MatrixDB

Two types of searches are offered by default. 'Biomolecule category' displays all the human molecules in a category (protein, glycosaminoglycan, fragment, lipid, cation and inorganic compound). Searching by 'Biomolecule name' can be performed with the biomolecule or gene name or with its UniprotKB and ChEBI accession number or MatrixDB identifier. Three other types of queries are available in the 'Advanced Search': free text search, search by PubMed identifier and dataset search. The dataset search displays all the interactions provided by a given database (IntAct, MINT, DIP, BioGRID and MatrixDB), or those reported in specific papers (11,20–22). Detailed data is displayed when a molecule is selected, and links are provided to access further information within MatrixDB or on external websites. For example, UniGene EST profiles or OMIM disease data associated with the gene coding for the protein(s) of interest are provided when available. A list of the protein partners is displayed with the number of experiments reporting each interaction. An interaction can be selected to examine these supporting experiments, and an experiment can be selected to access to kinetics, affinity, binding site and the experimental species.

BUILDING INTERACTION NETWORKS USING MatrixDB DATA

Several options are available for building customized networks. The user can create (i) the entire network of interactions involving at least one extracellular partner, combined or not with interactions established by

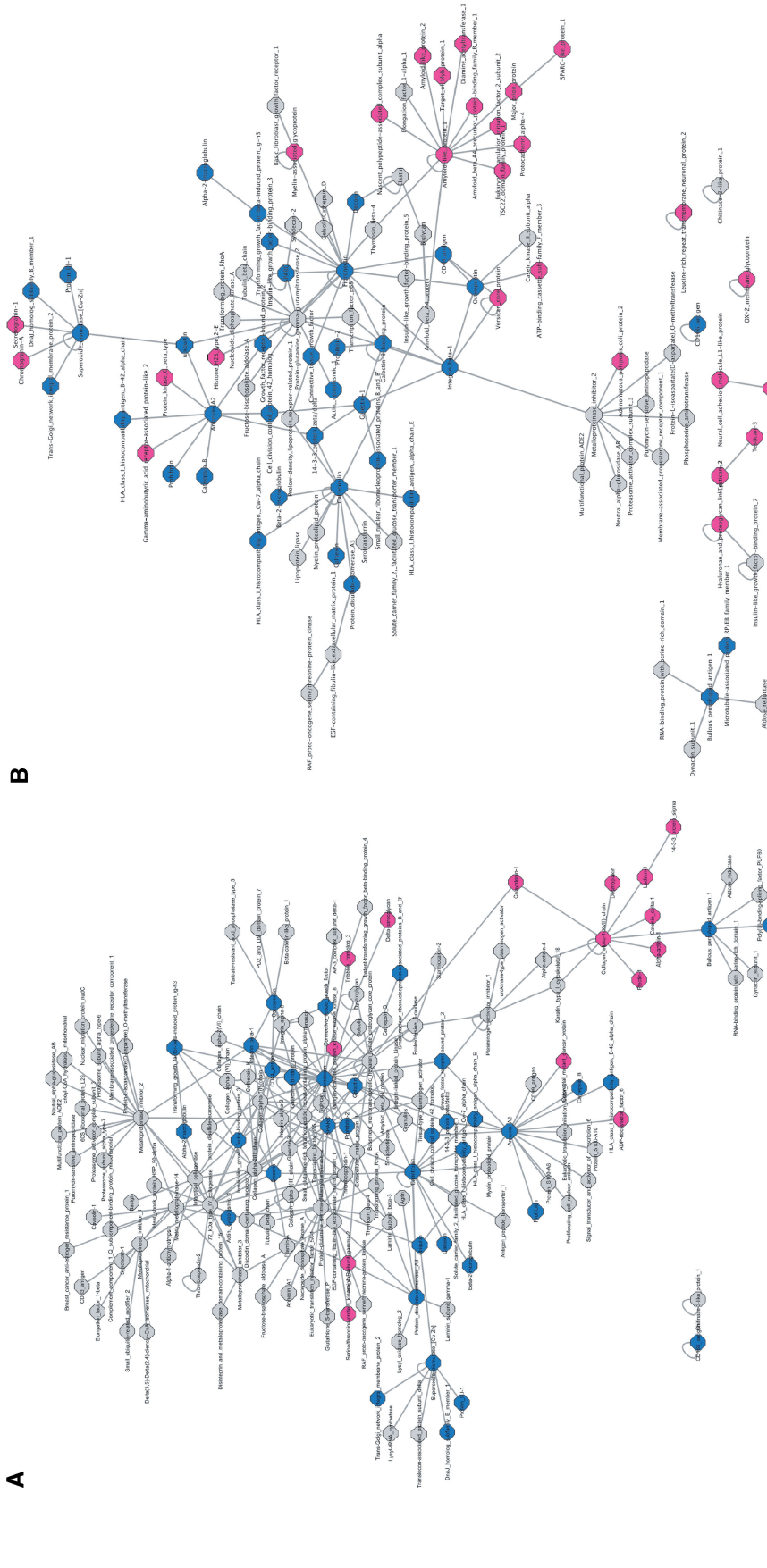

Figure 2. Protein–protein interaction networks of skin (**A**) and brain (**B**), built using MatrixDB and visualized with Cytoscape (threshold used for UniGene annotations: ≥100 transcripts per million). Edges: interactions. Pink nodes: proteins specific of skin (A) or brain (B); blue nodes: proteins present in five tissues (liver, lung, bone, skin and brain); grey nodes: other proteins present in two to four tissues out of five.

membrane and secreted molecules, (ii) the interaction network of proteins annotated with user-selected UniProtKB keywords, (iii) the interaction network of one or several molecule(s), including or not the interactions of its (their) partners, (iv) tissue-specific interaction networks (Figure 2) and (v) disease-specific interaction networks. The building of tissue-specific interaction networks is based on expression data imported from UniGene. One or several tissues can be selected and a threshold (minimum number of transcripts per million present in the tissue) can be defined to keep only interactions established by proteins expressed above this threshold in the selected tissues. An option restricts the interactions to those where the partners are specifically expressed in one or several selected tissues. This function allows the identification of tissue-specific partners. It is also possible to build disease-specific interaction networks, based on OMIM identifiers.

The interaction networks are visualized using Cytoscape (31) or Medusa (32), using customized styles that are described in the MatrixDB tutorial available on the website.

CONCLUSION

MatrixDB is a database providing interaction data involving extracellular proteins and glycosaminoglycans and interactions established by these two major constituents of the extracellular matrix with cations and lipids. Building the extracellular interactome is a prerequisite to delineate the molecular mechanisms underlying the assembly of the extracellular matrix and to understand how genetic diseases interfere with this process. Future releases will also include interaction data imported from the databases that will join the IMEx consortium. MatrixDB will increase its coverage by curation of interactions involving (i) matrix metalloproteinases and their inhibitors, which play a major role in tissue remodelling (links to the MEROPS database (33) will be provided), (ii) the adhesive matrix molecule family (microbial surface components recognizing adhesive matrix molecules, MSCRAMMs) responsible for the interaction of pathogens with the extracellular matrix (34) and (iii) other proteins and sugars of pathogens.

ACKNOWLEDGEMENTS

We would like to thank Christophe Blanchet (UMR 5086, Lyon, France) for helping us to install the MatrixDB server, Samuel Kerrien and Bruno Aranda (EBI, Hinxton, UK) for their help regarding data format exchange and Sandra Orchard (EBI, Hinxton, UK) for guiding us through the curation process.

FUNDING

This work was supported by a CPER grant from the Région Rhône-Alpes; by Institut des Systèmes Complexes (IXXI 2010); and by the EU FP7 'PSIMEx' grant (contract number FP7-HEALTH-2007-223411). Funding for open access charge: EU FP7 'PSIMEx' grant (contract number FP7-HEALTH-2007-223411).

Conflict of interest statement. None declared.

REFERENCES

1. Ricard-Blum,S. and Ruggiero,F. (2005) The collagen superfamily: from the extracellular matrix to the cell membrane. *Pathol. Biol. (Paris)*, **53**, 430–442.
2. Hynes,R.O. (2009) The extracellular matrix: not just pretty fibrils. *Science*, **326**, 1216–1219.
3. Ricard-Blum,S. and Ballut,L. Matricryptins derived from collagens and proteoglycans. *Front Biosci*, in press.
4. Bateman,J.F., Boot-Handford,R.P. and Lamandé,S.R. (2009) Genetic diseases of connective tissues: cellular and extracellular effects of ECM mutations. *Nat. Rev. Genet.*, **10**, 173–183.
5. Heinegård,D. (2009) Proteoglycans and more–from molecules to biology. *Int. J. Exp. Pathol.*, **90**, 575–586.
6. Chautard,E., Ballut,L., Thierry-Mieg,N. and Ricard-Blum,S. (2009) MatrixDB, a database focused on extracellular protein-protein and protein-carbohydrate interactions. *Bioinformatics*, **2**, 690–691.
7. UniProt Consortium. (2010) The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Res.*, **38**, D142–D148.
8. de Matos,P., Alcantara,R., Dekker,A., Ennis,M., Hastings,J., Haug,K., Spiteri,I., Turner,S. and Steinbeck,C. (2010) Chemical entities of biological interest: an update. *Nucleic Acids Res.*, **38**, D249–D254.
9. Kanehisa,M., Goto,S., Furumichi,M., Tanabe,M. and Hirakawa,M. (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.*, **38**, D355–D360.
10. Orchard,S., Kerrien,S., Jones,P., Ceol,A., Chatr-Aryamontri,A., Salwinski,L., Nerothin,J. and Hermjakob,H. (2007) Submit your interaction data the IMEx way: a step by step guide to trouble-free deposition. *Proteomics*, **7**, 28–34.
11. Faye,C., Chautard,E., Olsen,B.R. and Ricard-Blum,S. (2009) The first draft of the endostatin interaction network. *J. Biol. Chem.*, **284**, 22041–22047.
12. Aranda,B., Achuthan,P., Alam-Faruque,Y., Armean,I., Bridge,A., Derow,C., Feuermann,M., Ghanbarian,A.T., Kerrien,S., Khadake,J. *et al.* (2010) The IntAct molecular interaction database in 2010. *Nucleic Acids Res.*, **38**, D525–D531.
13. Salwinski,L., Miller,C.S., Smith,A.J., Pettit,F.K., Bowie,J.U. and Eisenberg,D. (2004) The database of interacting proteins: 2004 update. *Nucleic Acids Res.*, **32**, D449–D451.
14. Ceol,A., Chatr,A.A., Licata,L., Peluso,D., Briganti,L., Perfetto,L., Castagnoli,L. and Cesareni,G. (2010) MINT, the molecular interaction database: 2009 update. *Nucleic Acids Res.*, **38**, D532–D539.
15. Breitkreutz,B.J., Stark,C., Reguly,T., Boucher,L., Breitkreutz,A., Livstone,M., Oughtred,R., Lackner,D.H., Bahler,J., Wood,V. *et al.* (2008) The BioGRID interaction database: 2008 update. *Nucleic Acids Res.*, **36**, D637–D640.
16. Keshava Prasad,T.S., Goel,R., Kandasamy,K., Keerthikumar,S., Kumar,S., Mathivanan,S., Telikicherla,D., Raju,R., Shafreen,B., Venugopal,A. *et al.* (2009) Human protein reference database: 2009 update. *Nucleic Acids Res.*, **37**, D767–D772.
17. The Gene Ontology Consortium. (2010) The Gene Ontology in 2010: extensions and refinements. *Nucleic Acids Res.*, **38**, D331–D335.
18. Orchard,S., Salwinski,L., Kerrien,S., Montecchi-Palazzi,L., Oesterheld,M., Stumpflen,V., Ceol,A., Chatr-Aryamontri,A., Armstrong,J., Woollard,P. *et al.* (2007) The minimum information required for reporting a molecular interaction experiment (MIMIx). *Nat. Biotechnol.*, **25**, 894–898.
19. Kerrien,S., Orchard,S., Montecchi-Palazzi,L., Aranda,B., Quinn,A.F., Vinod,N., Bader,G.D., Xenarios,I., Wojcik,J., Sherman,D. *et al.* (2007) Broadening the horizon–level 2.5 of the HUPO-PSI format for molecular interactions. *BMC Biol.*, **5**, 44.

20. Bushell,K.M., Söllner,C., Schuster-Boeckler,B., Bateman,A. and Wright,G.J. (2008) Large-scale screening for novel low-affinity extracellular protein interactions. *Genome Res.*, **18**, 622–630.
21. Söllner,C. and Wright,G.J. (2009) A cell surface interaction network of neural leucine-rich repeat receptors. *Genome Biol.*, **10**, R99.
22. Cain,S.A., McGovern,A., Small,E., Ward,L.J., Baldock,C., Shuttleworth,A. and Kielty,C.M. (2009) Defining elastic fiber interactions by molecular fishing: an affinity purification and mass spectrometry approach. *Mol. Cell Proteomics*, **8**, 2715–2732.
23. Pontén,F., Jirström,K. and Uhlen,M. (2008) The human protein atlas–a tool for pathology. *J. Pathol.*, **216**, 387–393.
24. 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.
25. Marini,J.C., Forlino,A., Cabral,W.A., Barnes,A.M., San Antonio,J.D., Milgrom,S., Hyland,J.C., Körkkö,J., Prockop,D.J., De Paepe,A. *et al.* (2007) Consortium for osteogenesis imperfecta mutations in the helical domain of type I collagen: regions rich in lethal mutations align with collagen binding sites for integrins and proteoglycans. *Hum. Mutat.*, **28**, 209–221.
26. Dalgleish,R. (1997) The human type I collagen mutation database. *Nucleic Acids Res.*, **25**, 181–187.
27. Dalgleish,R. (1998) The human collagen mutation database 1998. *Nucleic Acids Res.*, **26**, 253–255.
28. 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.
29. Amberger,J., Bocchini,C.A., Scott,A.F. and Hamosh,A. (2009) McKusick's Online Mendelian Inheritance in Man (OMIM). *Nucleic Acids Res.*, **37**, D793–D796.
30. Sweeney,S.M., Orgel,J.P., Fertala,A., McAuliffe,J.D., Turner,K.R., Di Lullo,G.A., Chen,S., Antipova,O., Perumal,S., Ala-Kokko,L. *et al.* (2008) Candidate cell and matrix interaction domains on the collagen fibril, the predominant protein of vertebrates. *J. Biol. Chem.*, **283**, 21187–21197.
31. Shannon,P., Markiel,A., Ozier,O., Baliga,N.S., Wang,J.T., Ramage,D., Amin,N., Schwikowski,B. and Ideker,T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–2504.
32. Hooper,S.D. and Bork,P. (2005) Medusa: a simple tool for interaction graph analysis. *Bioinformatics*, **21**, 4432–4433.
33. Rawlings,N.D., Barrett,A.J. and Bateman,A. (2010) MEROPS: the peptidase database. *Nucleic Acids Res.*, **38**, D227–D233.
34. Speziale,P., Pietrocola,G., Rindi,S., Provenzano,M., Provenza,G., Di Poto,A., Visai,L. and Arciola,C.R. (2009) Structural and functional role of *Staphylococcus aureus* surface components recognizing adhesive matrix molecules of the host. *Future Microbiol.*, **4**, 1337–1352.