

MoonDB 2.0: an updated database of extreme multifunctional and moonlighting proteins

Diogo M. Ribeiro¹, Galadriel Briere¹, Benoit Bely², Lionel Spinelli¹ and Christine Brun^{1,3,*}

¹Aix-Marseille Univ, INSERM, TAGC, UMR_S1090, Marseille, France, ²The European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus CB10 1SD, UK and ³CNRS, Marseille, France

Received August 09, 2018; Revised September 17, 2018; Editorial Decision October 14, 2018; Accepted October 17, 2018

ABSTRACT

MoonDB 2.0 (<http://moondb.hb.univ-amu.fr/>) is a database of predicted and manually curated extreme multifunctional (EMF) and moonlighting proteins, i.e. proteins that perform multiple unrelated functions. We have previously shown that such proteins can be predicted through the analysis of their molecular interaction subnetworks, their functional annotations and their association to distinct groups of proteins that are involved in unrelated functions. In MoonDB 2.0, we updated the set of human EMF proteins (238 proteins), using the latest functional annotations and protein–protein interaction networks. Furthermore, for the first time, we applied our method to four additional model organisms - mouse, fly, worm and yeast - and identified 54 novel EMF proteins in these species. In addition to novel predictions, this update contains 63 human and yeast proteins that were manually curated from literature, including descriptions of moonlighting functions and associated references. Importantly, MoonDB's interface was fully redesigned and improved, and its entries are now cross-referenced in the UniProt Knowledgebase (UniProtKB). MoonDB will be updated once a year with the novel EMF candidates calculated from the latest available protein interactions and functional annotations.

INTRODUCTION

Moonlighting, multitask and extreme multifunctional proteins are proteins that perform multiple unrelated biological functions, regardless of their domain organisation and their evolutionary history (1–3). A canonical example of a moonlighting protein is the human aconitase, an enzyme of the tricarboxylic acid cycle (TCA cycle) that also functions as a translational regulator, upon a conformational change (4). Extreme multifunctional and moonlighting proteins are

present throughout the evolutionary tree, and their unrelated functions may be performed in different tissues or cellular locations, sometimes associated (either as a cause or a consequence) to a change in their interaction partners, conformation or oligomeric states (5). These proteins are often in the intersection—and may coordinate—several pathways or responses to different stimuli (6). Despite their importance, the moonlighting functions of proteins have usually been identified by serendipity, since clear procedures to identify secondary functions have not been proposed. As a consequence, the prevalence of moonlighting proteins in proteomes was unknown. This prompted us to provide in 2015, MoonGO, a computational pipeline to identify extreme multifunctional (EMF) proteins on a large scale (2). EMF proteins were identified by exploiting the topology of protein–protein interaction networks and protein GO term annotations, without any *a priori* knowledge of moonlighting. The first version of MoonDB (2) contained the EMF proteins predicted by MoonGO, complemented with a careful manual curation of literature of moonlighting or EMF proteins. Here, we present MoonDB 2.0, an update that, besides improving predictions and manual curation for human, also includes predicted and curated entries for four other model organisms - mouse, fly, worm and yeast. Our main focus is to provide users with an extensive set of predicted and curated EMF and moonlighting proteins, describing their functions comprehensively.

MATERIALS AND METHODS

Prediction of extreme multifunctional proteins

The method used to predict extreme multifunctional (EMF) proteins was first described in Chapple *et al.* (2). Briefly, we perform a large-scale search for EMF proteins by (i) identifying functionally-dissimilar pairs of Biological Process Gene Ontology (GO) terms with PrOnto (7) that uses two metrics of GO *functional dissimilarity* based on the frequency of co-occurrence of GO term pairs in protein annotations; (ii) clustering the protein interactome into overlapping clusters of proteins using the OCG algorithm (8); (iii) annotating each cluster with functions (Biological Pro-

*To whom correspondence should be addressed. Tel: +33 491828712; Email: christine-g.brun@inserm.fr
Present address: Galadriel Briere, Bordeaux Sciences Agro, 33175 Gradignan, France.

cess GO terms) based on the annotations of its constituent proteins; (iv) identifying proteins that belong to at least two clusters and are annotated to *dissimilar functions* (after having inherited the annotations of their clusters in addition to their own), hereby labeled as EMF proteins. We used protein–protein interaction data gathered on December 2017 from the PSICQUIC webservice (9), processed as described in Chapple *et al.* 2015 (2). We only include experimentally identified binary protein–protein interactions, by considering only interactions from certain experimental methods (Supplementary Table S2). GO term annotations and ontologies were collected from the Gene Ontology Consortium (10) on December 2017.

Criteria for manual curation

We provide a list of *bona fide* moonlighting and extreme multifunctional proteins manually curated from literature over the years. Each entry was confirmed independently by at least two members of our team. Specifically, we confirm that the several functions are indeed distinct to each other and not a by-product of the same function under different circumstances (e.g. regulation of two distinct pathways through the same mechanism, such as phosphorylation). In each case, publications describing the different functions of a protein are provided. When available, the conditions that may be related to the change in function are also described (e.g. cellular localisation, oligomerization).

Database architecture and web interface

MoonDB 2.0 has been developed using the SQLAlchemy (v1.2.0) Python (v2.7.6) library for data storage. The web interface is mainly written in PHP (v7.1.14) and JQuery (v3.2.1) and is powered by the Drupal (v8.4) Content Management System (CMS). The database was deployed with Docker (v17.09.0-ce) to ensure stability. We gathered information on protein domains, publications and diseases from UniProtKB (11) in January 2018.

DATABASE CONTENT AND WEB INTERFACE

A new dataset of extreme multifunctional proteins

In MoonDB 2.0 we have predicted 292 extreme multifunctional (EMF) proteins in human, and—for the first time—also in mouse, fly, worm and yeast model species (Supplementary Table S1). These have been identified *de novo* using the latest protein–protein interaction networks and GO term annotations. The power to predict EMF proteins is dependent on the underlying coverage and quality of the protein interactomes and GO term annotations used (Supplementary Table S1). The new human EMF protein dataset and the one in the previous MoonDB version overlap significantly (Supplementary Figure S1). Interestingly, the analysis of the EMF signature on the new set of EMF proteins (as performed in Chapple *et al.* (2)), shows that the new set of EMF proteins produces a similar signature in terms of network properties, tissue expression, as well as domain, structural disorder and Eukaryotic linear motif (ELM) presence, among other features (Supplementary Figure S2). Notably, as observed in other studies (2,12,13),

moonlighting and EMF proteins are often associated to disease, a feature also observed for the set of human EMF proteins when considering disease-associations from OMIM ($P = 2.2 \times 10^{-7}$, odds ratio = 2.10; one-tailed Fisher's Exact test).

Besides EMF predictions, in this update we also manually curated 15 yeast moonlighting proteins, describing their unrelated functions, specifying which conditions may influence moonlighting (e.g. cellular localisation) and referencing relevant publications. Similarly, we added functional descriptions for 47 human moonlighting proteins. All these proteins constitute the 'Reference Set'.

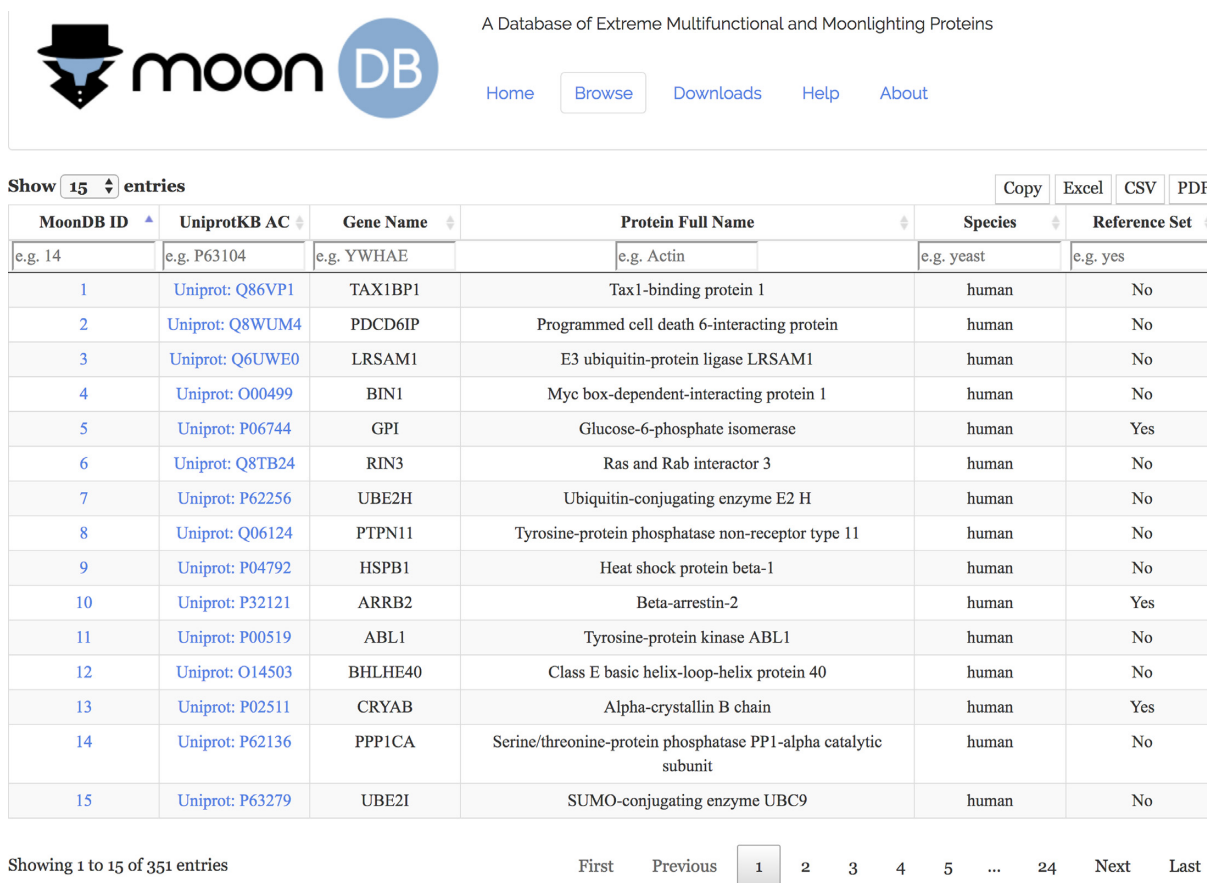
A new user-friendly interface and additional content

To provide our visitors with a clear, fast and easy-to-use database, we completely redesigned MoonDB's web interface and added new functionalities. It is now possible to search a MoonDB entry by gene name, UniProtKB identifier (ID) or UniProtKB accession (AC). Moreover, the 'Browse' page (Figure 1), which displays all protein entries in MoonDB, can now be filtered through any column, thus allowing searches by species, full name of the protein and its presence in the 'Reference Set'. These filters can also be used in combination with each other to make more advanced queries.

Importantly, MoonDB specifies which pairs of dissimilar (i.e. unrelated) functions led us to propose each predicted EMF and curated protein as moonlighting/extreme-multifunctional (Figure 2, under 'MoonDB Dissimilar Functions'). Furthermore, we provide the set of GO terms associated to the protein (Figure 2, under 'MoonDB Network Modules'), determined by its participation in network clusters with annotated functions (*guilt-by-association* principle), and the GO terms directly annotating the protein. This information is pertinent in the context of multifunctionality, since EMF proteins associate with several groups of proteins to perform alternative functions. In addition, since the ability to perform unrelated functions may be correlated with the presence of a protein in unrelated subcellular locations, in MoonDB 2.0 we identified pairs of unrelated cellular component GO terms associated to each protein with PrOnto (7) (Figure 2, under 'Protein GO Annotations'). Lastly, to fully describe moonlighting and extreme multifunctional proteins, we further cross-link functional data with other orthogonal information such as the protein association to disease, protein domains and publications associated to the protein.

DISCUSSION AND CONCLUSION

The MoonDB 2.0 database is accessible at <http://moondb.hb.univ-amu.fr/> and now contains data for human, mouse, fly, worm and yeast. MoonDB 2.0 stands out compared to the two other current databases of moonlighting proteins MoonProt (14) and Multitask-II (12) because MoonDB 2.0 combines curated and predicted proteins. We consider our dataset to be more comprehensive as well as highly complementary to other available databases. Whereas other databases are dependent on the available literature, and thus limited to providing information which is already known,



A Database of Extreme Multifunctional and Moonlighting Proteins

Home Browse Downloads Help About

Show 15 entries Copy Excel CSV PDF

MoonDB ID	UniprotKB AC	Gene Name	Protein Full Name	Species	Reference Set
e.g. 14	e.g. P63104	e.g. YWHAE	e.g. Actin	e.g. yeast	e.g. yes
1	Uniprot: Q86VP1	TAX1BP1	Tax1-binding protein 1	human	No
2	Uniprot: Q8WUM4	PDCD6IP	Programmed cell death 6-interacting protein	human	No
3	Uniprot: Q6UWE0	LRSAM1	E3 ubiquitin-protein ligase LRSAM1	human	No
4	Uniprot: O00499	BIN1	Myc box-dependent-interacting protein 1	human	No
5	Uniprot: P06744	GPI	Glucose-6-phosphate isomerase	human	Yes
6	Uniprot: Q8TB24	RIN3	Ras and Rab interactor 3	human	No
7	Uniprot: P62256	UBE2H	Ubiquitin-conjugating enzyme E2 H	human	No
8	Uniprot: Q06124	PTPN11	Tyrosine-protein phosphatase non-receptor type 11	human	No
9	Uniprot: P04792	HSPB1	Heat shock protein beta-1	human	No
10	Uniprot: P32121	ARRB2	Beta-arrestin-2	human	Yes
11	Uniprot: P00519	ABL1	Tyrosine-protein kinase ABL1	human	No
12	Uniprot: O14503	BHLHE40	Class E basic helix-loop-helix protein 40	human	No
13	Uniprot: P02511	CRYAB	Alpha-crystallin B chain	human	Yes
14	Uniprot: P62136	PPP1CA	Serine/threonine-protein phosphatase PP1-alpha catalytic subunit	human	No
15	Uniprot: P63279	UBE2I	SUMO-conjugating enzyme UBC9	human	No

Showing 1 to 15 of 351 entries First Previous 1 2 3 4 5 ... 24 Next Last

Figure 1. MoonDB 2.0 browse page. The browse page displays the entries of all MoonDB 2.0 proteins and can be searched interactively. The ‘MoonDB ID’ can be clicked to access each individual MoonDB 2.0 protein entry.

our dataset of predictions goes beyond current propositions of moonlighting and provides novel candidates. EMF prediction is large-scale and detection does not require *a priori* knowledge besides protein interactions and GO term annotations.

Importantly, as protein interactomes and GO term annotations of model organisms will continue to grow towards completion in the following years, MoonDB will be updated every year with EMF predictions made from the latest interactomes and GO term annotations. Both power and reliability will progressively increase with future releases. This will be particularly important in the cases of mouse, which possesses high-quality GO term annotations (average of ~19 GO terms per protein), but an incomplete protein interactome (<15% of the proteome covered), as well as in fly, whose interactome is better covered (>40% proteome), but GO term annotations are available for less than half of the interactome. Consequently, only few EMF proteins in mouse and fly were detected with our method. However, 5 out of 14 mouse EMF proteins are orthologs of human EMF proteins, suggesting that even when data is limited, the EMF proteins predicted are reliable. Indeed, the ability for genes to be multifunctional is conserved across orthologs of different organisms (15) and some orthologous proteins are known to have moonlighting functions in dif-

ferent organisms (16). Notably, orthologs were also found between human and worm (UBE2I/ubc-9 and SUMO1; 2 out of 6 MoonDB 2.0 worm entries) and even between the distant human and yeast species (SKP1 gene), although our method does not use ortholog relationships for EMF prediction. Together, these findings further underline the quality of our predictions and designates MoonDB 2.0 as a valuable data repository for one interested in studying the extreme multifunctionality and moonlighting of proteins, possibly across species.

We believe that MoonDB is of interest not only to bioinformaticians working on multifunctionality, but also to any biologist who may profit from knowing whether their protein of study is likely to perform unexpected functions aside from the ones generally known. Due to the high frequency of EMF proteins involved in multiple diseases, often in comorbidity (13), the extensive functional information provided in MoonDB 2.0 is of interest to help designing therapies that are aware of the several functions of the protein. Importantly, MoonDB 2.0 is now cross-referenced in the UniProt Knowledgebase (UniProtKB) (11). We consider that this greatly magnifies the exposure of our database to the general scientific community, as UniProtKB is the reference database for protein-related data and widely used by biologists, biochemists, bioinformaticians and others.

UniprotKB AC [▲]	UniprotKB ID	Gene name	Full name	Species	Curated set
P04156 (Uniprot)	PRI0_HUMAN	PRNP	Major prion protein	human	No

Protein Function [❏]

MoonDB Dissimilar Functions [❏]

Showing 1 to 2 of 2 entries Search:

GO ID 1 [▲]	Function 1	Module ID 1 [❏]	GO ID 2 [❏]	Function 2	Module ID 2 [❏]	Association Probability [❏] (PrOnto)	Interaction Probability [❏] (PrOnto)
GO:0019538	protein metabolic process	665	GO:0016070	RNA metabolic process	75	5.15e-08	2.82e-13
GO:0036211	protein modification process	106	GO:0016070	RNA metabolic process	75	5.81e-04	8.44e-05

Copy Excel CSV PDF

Previous Next

MoonDB Network Modules [❏]

Biological Processes

Showing 1 to 10 of 31 entries Search:

Module ID [❏] (MoonGO) [▲]	GO ID (BP) [❏]	GO Name
75	GO:0016071	mRNA metabolic process
75	GO:0051252	regulation of RNA metabolic process
75	GO:0010468	regulation of gene expression

Protein GO Annotations [❏]

Biological Processes

Cellular Components

Dissimilar Cellular Components (PrOnto) [❏]

Showing 1 to 3 of 3 entries Search:

GO ID 1 [▲]	Component 1 [❏]	GO ID 2 [❏]	Component 2	Association Probability [❏] (PrOnto)	Interaction Probability [❏] (PrOnto)
GO:0005829	cytosol	GO:0009986	cell surface	3.26e-22	6.81e-06
GO:0005886	plasma membrane	GO:0043231	intracellular membrane-bounded organelle	0.00e+00	3.79e-47
GO:0009986	cell surface	GO:0043231	intracellular membrane-bounded organelle	1.41e-26	4.81e-14

Copy Excel CSV PDF

Previous Next

All Protein Interactions [❏]

Publications Associated [❏]

OMIM Diseases Associated [❏]

Figure 2. Example of a MoonDB's protein entry. Protein entries provide extensive functional information such as the dissimilar function annotations and GO term annotations from network modules, as well as publications, diseases and domains associated with the protein.

DATA AVAILABILITY

MoonDB 2.0 is freely available at <http://moondb.hb.univ-amu.fr/>. Files containing EMF protein lists for each species, as well as the protein–protein interaction networks used in this study are freely available for download in MoonDB 2.0, and can be used in accordance with the GNU Public License and the license of primary data sources.

SUPPLEMENTARY DATA

[Supplementary Data](#) are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to thank Zacharie Menetrier for creating the MoonDB 2.0 logo, Benoit Ballester for tips regarding biological databases and Andreas Zanzoni for critically reading the manuscript and testing the database.

FUNDING

Excellence Initiative of Aix-Marseille University—A*MIDEX, a French ‘Investissements d’Avenir’ programme (to C.B.). Funding for open access charge: Excellence Initiative of Aix-Marseille University—A*MIDEX, a French ‘Investissements d’Avenir’ programme.

Conflict of interest statement. None declared.

REFERENCES

1. Jeffery, C.J. (1999) Moonlighting proteins. *Trends Biochem. Sci.*, **24**, 8–11.
2. Chapple, C.E., Robisson, B., Spinelli, L., Guien, C., Becker, E. and Brun, C. (2015) Extreme multifunctional proteins identified from a human protein interaction network. *Nat. Commun.*, **6**, 7412.
3. Chapple, C.E. and Brun, C. (2015) Redefining protein moonlighting. *Oncotarget*, **6**, 16812–16813.
4. Volz, K. (2008) The functional duality of iron regulatory protein 1. *Curr. Opin. Struct. Biol.*, **18**, 106–111.
5. Jeffery, C.J. (2014) An introduction to protein moonlighting. *Biochem. Soc. Trans.*, **42**, 1679–1683.
6. Jeffery, C.J. (2015) Why study moonlighting proteins? *Front. Genet.*, **6**, 211.
7. Chapple, C.E., Herrmann, C. and Brun, C. (2015) PrOnto database: GO term functional dissimilarity inferred from biological data. *Front. Genet.*, **6**, 200.
8. Becker, E., Robisson, B., Chapple, C.E., Guenoche, A. and Brun, C. (2012) Multifunctional proteins revealed by overlapping clustering in protein interaction network. *Bioinformatics*, **28**, 84–90.
9. del-Toro, N., Dumousseau, M., Orchard, S., Jimenez, R.C., Galeota, E., Launay, G., Goll, J., Breuer, K., Ono, K., Salwinski, L. *et al.* (2013) A new reference implementation of the PSICQUIC web service. *Nucleic Acids Res.*, **41**, W601–W606.
10. The Gene Ontology Consortium (2017) Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res.*, **45**, D331–D338.
11. UniProt Consortium, T. (2018) UniProt: the universal protein knowledgebase. *Nucleic Acids Res.*, **46**, 2699.
12. Franco-Serrano, L., Hernández, S., Calvo, A., Severi, M.A., Ferragut, G., Pérez-Pons, J., Piñol, J., Pich, Ó., Mozo-Villarias, Á., Amela, I. *et al.* (2018) MultitaskProtDB-II: an update of a database of multitasking/moonlighting proteins. *Nucleic Acids Res.*, **46**, D645–D648.
13. Zanzoni, A., Chapple, C.E. and Brun, C. (2015) Relationships between predicted moonlighting proteins, human diseases, and comorbidities from a network perspective. *Front. Physiol.*, **6**, 171.
14. Chen, C., Zabad, S., Liu, H., Wang, W. and Jeffery, C. (2018) MoonProt 2.0: an expansion and update of the moonlighting proteins database. *Nucleic Acids Res.*, **46**, D640–D644.
15. Pritykin, Y., Ghersi, D. and Singh, M. (2015) Genome-Wide detection and analysis of multifunctional genes. *PLoS Comput. Biol.*, **11**, e1004467.
16. Copley, S.D. (2012) Moonlighting is mainstream: paradigm adjustment required. *BioEssays*, **34**, 578–588.