Published online 9 November 2017 *Nucleic Acids Research, 2018, Vol. 46, Database issue* ***D645–D648***
doi: 10.1093/nar/gkx1066

MultitaskProtDB-II: an update of a database of multitasking/moonlighting proteins

Luís Franco-Serrano[1], Sergio Hernández[1], Alejandra Calvo[2], María A. Severi[2], Gabriela Ferragut[2], JosepAntoni Pérez-Pons[1], Jaume Piñol[1], Òscar Pich[1], Ángel Mozo-Villarias[1], Isaac Amela[1], Enrique Querol[1,*] and Juan Cedano[2,*]

[1]Institut de Biotecnologia i Biomedicina and Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, 08193 Cerdanyola del Vallès, Barcelona, Spain and [2]Laboratorio de Inmunología, Universidad de la República Regional Norte-Salto, Rivera 1350, CP 50000 Salto, Uruguay

Received September 15, 2017; Revised October 11, 2017; Editorial Decision October 12, 2017; Accepted October 20, 2017

ABSTRACT

Multitasking, or moonlighting, is the capability of some proteins to execute two or more biological functions. MultitaskProtDB-II is a database of multifunctional proteins that has been updated. In the previous version, the information contained was: NCBI and UniProt accession numbers, canonical and additional biological functions, organism, monomeric/oligomeric states, PDB codes and bibliographic references. In the present update, the number of entries has been increased from 288 to 694 moonlighting proteins. MultitaskProtDB-II is continually being curated and updated. The new database also contains the following information: GO descriptors for the canonical and moonlighting functions, three-dimensional structure (for those proteins lacking PDB structure, a model was made using Itasser and Phyre), the involvement of the proteins in human diseases (78% of human moonlighting proteins) and whether the protein is a target of a current drug (48% of human moonlighting proteins). These numbers highlight the importance of these proteins for the analysis and explanation of human diseases and target-directed drug design. Moreover, 25% of the proteins of the database are involved in virulence of pathogenic microorganisms, largely in the mechanism of adhesion to the host. This highlights their importance for the mechanism of microorganism infection and vaccine design. MultitaskProtDB-II is available at http://wallace.uab.es/multitaskII.

INTRODUCTION

Multitasking, or moonlighting, refers to those proteins presenting two or more functions performed by a single polypeptide chain. The term 'moonlighting' was coined by Constance Jeffery (1), whereas Joram Piatigorsky proposed gene sharing (2). Multitasking proteins present alternative functions, resulting from differences in cellular localization, cell type, oligomeric state, concentration of cellular ligands, substrates, cofactors, products or post-translational modifications. Although some findings suggest the involvement of a protein in extra functions, for example, finding them in different cellular localizations or in amounts exceeding those required for their canonical function, usually multitasking proteins are experimentally revealed by serendipity. The appearance of a new function can become an advantage for the cell or the organism because it reduces the number of proteins to be synthesized, making its genome more compact and coordinating cell activities better. In any case, these proteins complicate the interpretation of knock-out/knock-in, DNA array, metabolomic, systems biology, drug pharmacokinetic, pharmacodynamic and toxicity assays. In order to facilitate the work of researchers interested in this field, we decided to make our set of multitasking proteins freely available in the form of a web database, MultitaskProtDB (3). Additionally, two other databases are currently available: MoonProt (4) and MoonDB (5).

DATABASE IMPROVEMENTS

Information on multitasking proteins has been collected from the literature through the NCBI PubMed server using keywords like moonlighting/multitasking/multifunctional protein and gene sharing. When necessary, some important protein characteristics have been retrieved from the UniProt Consortium (6). In order to identify which proteins of our database are involved in human diseases, the information present in the Online Mendelian Inheritance

*To whom correspondence should be addressed. Tel: +34 93 586 8951; Fax: +34 93 581 2011; Email: enric.querol@uab.cat
Correspondence may also be addressed to Juan Cedano. Tel: +598 47 337 133; Email: jcedano@unorte.edu.uy

Table 1. Most frequent folds between moonlighting proteins in which the 3D structure is available

SCOPe ID	FOLDS	FREQUENCY
c1	TIM beta/alpha-barrel	9
c2	NAD(P)-binding Rossmann-fold domains	9
b1	Immunoglobulin-like beta-sandwich	6
c47	Thioredoxin fold	6
c37	P-loop containing nucleoside triphosphate hydrolases	5
c55	Ribonuclease H-like motif	4
c57	Molybdenum cofactor biosynthesis proteins	4
d144	Protein kinase-like (PK-like)	4
d54	Enolase N-terminal domain-like	4
i1	Ribosome and ribosomal fragments	4
a118	alpha-alpha superhelix	3
c8	The 'swivelling' beta/beta/alpha domain	3
d15	beta-Grasp (ubiquitin-like)	3
d162	LDH C-terminal domain-like	3
d58	Ferredoxin-like	3
a127	L-aspartase-like	2
a45	GST C-terminal domain-like	2
b26	SMAD/FHA domain	2
b29	Concanavalin A-like lectins/glucanases	2
b35	GroES-like	2
b42	beta-Trefoil	2
b69	7-bladed beta-propeller	2
b85	beta-clip	2
c23	Flavodoxin-like	2
c26	Adenine nucleotide alpha hydrolase-like	2
c42	Arginase/deacetylase	2
c58	Aminoacid dehydrogenase-like, N-terminal domain	2
c67	PLP-dependent transferase-like	2
c80	SIS domain	2
d2	Lysozyme-like	2
d41	alpha/beta-Hammerhead	2
d8	Urease, gamma-subunit	2
g37	beta-beta-alpha zinc fingers	2
b98	Zn aminopeptidase N-terminal domain	2

in Man (OMIM) (7) and Human Gene Mutation Database (HGMD) (8) databases, have been carefully inspected. Moreover, in order to check which proteins of our database are a drug target, the Therapeutic Target Database (TTD) (9) and the DrugBank database (10) have been scanned. The three-dimensional (3D) structure of those proteins without a previously-solved PDB structure has been modeled by applying the ITasser (11) and Phyre (12) servers. Both methods use template-based tertiary structure modeling.

Using the SCOP code associated with the PDB structure with which the sequence of the moonlighting protein aligned, a table of observed main fold frequencies was made (Table 1). In order to study any fold preference in our database of moonlighting proteins, all proteins were aligned with the astral95 database, and a protein subset was made considering only those with <95% of identity (moon95). With this strategy, we wished to prevent the abundance of the same protein of close species and avoid over-represented proteins because of the accumulation of the same protein with multiple moonlighting functions. To test whether the distribution of fold frequencies is similar to what we would see if the moonlighting proteins had the same distribution of folds as that observed in the astral95 database, the frequency distribution of subset moon95 was compared with the distribution present in the proteins of the astral95 database. This was done using a G-test calculated through a specific statistical R package (www.r-project.org). The *P*-value provided by R was $<2.2 \times 10–16$, which is below the acceptance threshold of the null hypothesis. We could then conclude that the distribution of frequencies in the structural classes of both subgroups of proteins is different.

USER INTERFACE

Upon opening the database web page (Figure 1), a large table that contains 694 entries of multitasking proteins is shown. Fifteen columns in the table give different information regarding the main characteristics of each protein. Alphabetically ordering is available by clicking on the title of each column, and this allows to order, for example, by organisms. From left to right, the following information is shown: Column 1 is a clickable button to see the complete record details. Column 2 is a clickable button to select the entry. Column 3 (UniProt) shows the UniProt accession number, which is linked to the corresponding database information. Column 4 (Protein Name) shows the name of the protein. Column 5 (Canonical Function) contains a detailed description of the canonical function of the protein. Columns 6 and 8 (GO and GO Moon) display GO numbers related to the canonical and moonlighting protein functions, respectively. Column 7 (Moonlighting Function) contains a detailed description of the moonlighting functions. Column 9 (Organism) indicates the organism in which the protein acts as a moonlighting protein according to the bibliography. Column 10 (Human Disease) indicates the associated diseases in the case of human moonlighting proteins

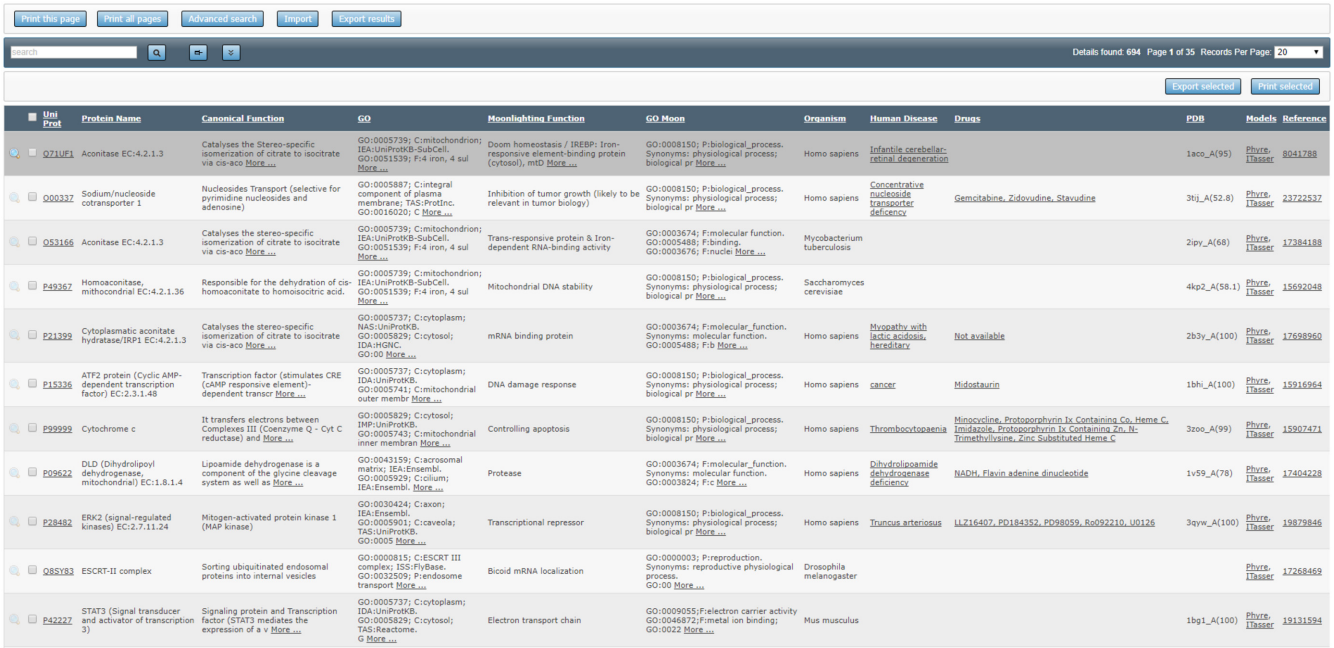

Figure 1. A screenshot of MultitaskProtDB-II web page. Currently, the database contains information of 694 multitasking proteins that can be easily viewed with the search button and other display facilities.

that are linked to the OMIM database. Column 11 (Drugs) indicates whether the protein is a known target of current drugs and is linked to the DrugBank database information. Column 12 (PDB) points to the PDB structure with which the structure was modeled (sequence identity is shown as a percentage), or to the experimentally solved PDB structure of the protein entry (homology is 100). Column 13 (Models) gives the 3D structure model with highest score according to ITasser or Phyre servers. These models were obtained by using the same sequence of the moonlighting protein and they are highly reliable, as they were modeled starting from high-homology templates. Even so, the reliability and coverage might be low in some cases, particularly if they are based on very remote homologs. Column 14 (Reference) provides a link to the PubMed bibliographic reference. Some facilities like display, print or search buttons are provided by the web page. Moreover, an export process can be easily performed to the whole database or to some selected entries. The type of data file obtained through the export option can be selected depending on the type of data file required by the user (Excel, Word, CSV or XML). The database is accessible at http://wallace.uab.es/multitaskII.

CONCLUSION

An important issue included in the present version of the database is the relation between multitasking proteins and human diseases. We have seen that 78% of the human moonlighting proteins are involved in diseases. Furthermore, 48% of the human multifunctional proteins are targets of current drugs. According to UniProt, the number of human proteins with a reviewed status is 26 199, and 13.74% of them are related to human diseases (see cross-reference between UniProt and OMIM at www.uniprot.org/help/involvement_in_disease). Thus, a percentage as high as the previously mentioned 78% indicates that moonlighting proteins are prone to be involved in human diseases, probably because of the two or more exhibited molecular functions. This is the case of fumarate hydratase, in which each molecular function is related to a different disease (fumarate deficiency and leiomiomatosis). A more unexpected result is the 48% of moonlighting proteins that are targets of current drugs (9,10), since only 9.8% of the human proteins present in UniProt are specified as drug targets. Moreover, targets of current drugs represent only successful cases, because not all the theoretically druggable human genome (5000–10 000 proteins, according to Drews, (13)) can be targetable. Still, a drug acting on a moonlighting protein can trigger more complicated toxic side-effects. Another interesting issue is that 25% of the database entries correspond to proteins involved in pathogenic microorganisms' virulence, mostly in the mechanism of host adhesion and colonization through interaction with plasminogen or extracellular matrix components. From the *reverse vaccinology* point of view, it is a very important fact. Several authors (14,15) have previously reviewed the involvement of moonlighting proteins in pathogen virulence.

The percentages described above highlight the interest of moonlighting proteins for gaining insight into the molecular basis of genetic-based diseases, the rational drug-design upon target identification, and the infection mechanism of pathogens and vaccine design. Databases such as MultitaskProtDB (3), MoonProt (4) and MoonDB (5) can be used as a source of data to create models or validate hypotheses about these proteins. Our database contains close to 700 experimentally demonstrated moonlighting proteins, with much information related to each one, and is a valuable resource for this growing class of proteins.

FUNDING

Ministerio de Economía y Competitividad of Spain [BIO2013–48704-R, BFU2013–50176-EXP]; Centre de Referència de R+D de Biotecnologia de la Generalitat de Catalunya; Comisión Coordinadora del Interior de Uruguay. Funding form open access charge: Ministerio de Economía y Competitividad of Spain [BIO2013–48704-R, BFU2013–50176-EXP].

Conflict of interest statement. None declared.

REFERENCES

1. Jeffery,C.J. (1999) Moonlighting proteins. *Trends Biochem.*, **24**, 8–11.
2. Piatigorsky,J. (1999) Gene sharing and evolution. Gene sharing in lens and cornea: facts and implications. *Prog. Retin. Eye Res.*, **18**, 552.
3. Hernández,S., Ferragut,G., Amela,I., Perez-Pons,J., Piñol,J., Mozo-Villarias,A., Cedano,J. and Querol,E. (2014) MultitaskProtDB: a database of multitasking proteins. *Nucleic Acids Res.*, **42**, D517–D520.
4. Mani,M., Chen,C., Amblee,V., Liu,H., Mathur,T., Zwicke,G., Zabad,S., Patel,B., Thakkar,J. and Jeffery,C.J. (2015) MoonProt: a database for proteins that are known to moonlight. *Nucleic Acids Res.*, **43**, D277–D282.
5. 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.
6. The UniProt Consortium (2017) UniProt: the universal protein knowledgebase. *Nucleic Acids Res.*, **45**, D158–D169.
7. Hamosh,A., Scott,A., Amberger,S., Bocchini,C. and McKusick,V. (2005) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.*, **33**, 514–517.
8. Stenson,P.D., Ball,E.V., Mort,M., Phillips,A.D., Shiel,J.A., Thomas,N.S., Abeysinghe,S., Krawczak,M. and Cooper,D.N. (2003) The human gene mutation database (HGMD): 2003 Update. *Hum. Mutat.*, **21**, 577–581.
9. Qin,C., Zhang,C., Zhu,F., Xu,F., Chen,S.Y., Zhang,P., Li,Y.H., Yang,S.Y., Wei,Y.Q., Tao,L. *et al.* (2014) Therapeutic target database update 2014: a resource for targeted therapeutics. *Nucleic Acids Res.*, **42**, D1118–D1123.
10. Law,V., Knox,C., Djoumbou,Y., Jewison,T. and Guo,A.C. (2014) DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res.*, **42**, D1091–D1097.
11. Wang,Y., Virtanen,J., Xue,Z. and Zhang,Y. (2017) I-TASSER-MR: automated molecular replacement for distant-homology proteins using iterative fragment assembly and progressive sequence truncation. *Nucleic Acids Res.*, doi: 10.1093/nar/gkx349.
12. Kelley,L.A., Mezulis,S., Yates,C.M., Wass,M.N. and Sternberg,M.J. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.*, **10**, 845–858.
13. Drews,J. (2000) Drug discovery: a historical perspective. *Science*, **287**, 1960–1964.
14. Henderson,B. and Martin,A. (2011) Bacterial virulence in the moonlight: Multitasking bacterial moonlighting proteins are virulence determinants in infectious disease. *Infect. Immun.*, **79**, 3476–3491.
15. Amblee,V. and Jeffery,C.J. (2015) Physical features of intracellular proteins that moonlight on the cell surface. *PLOS One*, **10**, e0130575.