

Do protein–protein interaction databases identify moonlighting proteins?†

Antonio Gómez,^{‡§^{ab}} Sergio Hernández,^{§^a} Isaac Amela,^a Jaume Piñol,^a Juan Cedano^a and Enrique Querol^{*^a}

Received 12th May 2011, Accepted 31st May 2011

DOI: 10.1039/c1mb05180f

One of the most striking results of the human (and mammalian) genomes is the low number of protein-coding genes. To-date, the main molecular mechanism to increase the number of different protein isoforms and functions is alternative splicing. However, a less-known way to increase the number of protein functions is the existence of multifunctional, multitask, or “moonlighting”, proteins. By and large, moonlighting proteins are experimentally disclosed by serendipity. Proteomics is becoming one of the very active areas of biomedical research, which permits researchers to identify previously unseen connections among proteins and pathways. In principle, protein–protein interaction (PPI) databases should contain information on moonlighting proteins and could provide suggestions to further analysis in order to prove the multifunctionality. As far as we know, nobody has verified whether PPI databases actually disclose moonlighting proteins. In the present work we check whether well-established moonlighting proteins present in PPI databases connect with their known partners and, therefore, a careful inspection of these databases could help to suggest their different functions. The results of our research suggest that PPI databases could be a valuable tool to suggest multifunctionality.

Moonlighting proteins alternative functions are mostly related to cellular localization, cell type, oligomeric state and the cellular concentration of ligands, substrates, cofactors and products.^{1–5} In any case, moonlighting will complicate the analysis and interpretation of protein networks of interactions, functional genomics, metabolomics, knock-out and iRNA phenotypes, genetic analysis of diseases, drug-target identification, toxicology and so on. Although some findings suggest involvement of a protein in extra functions, *i.e.*, finding them in different cellular locations; in amounts exceeding those required for its catalytic known function, usually moonlighting proteins are experimentally disclosed by serendipity; therefore

any alternative method to identify these proteins would be very valuable. In a previous work, the possibility of identifying moonlighting proteins by bioinformatics was explored by our group.⁶ In the present work, we check whether the analysis of protein interacting partners of well-established moonlighting proteins can be reliable enough to disclose multifunctionality.

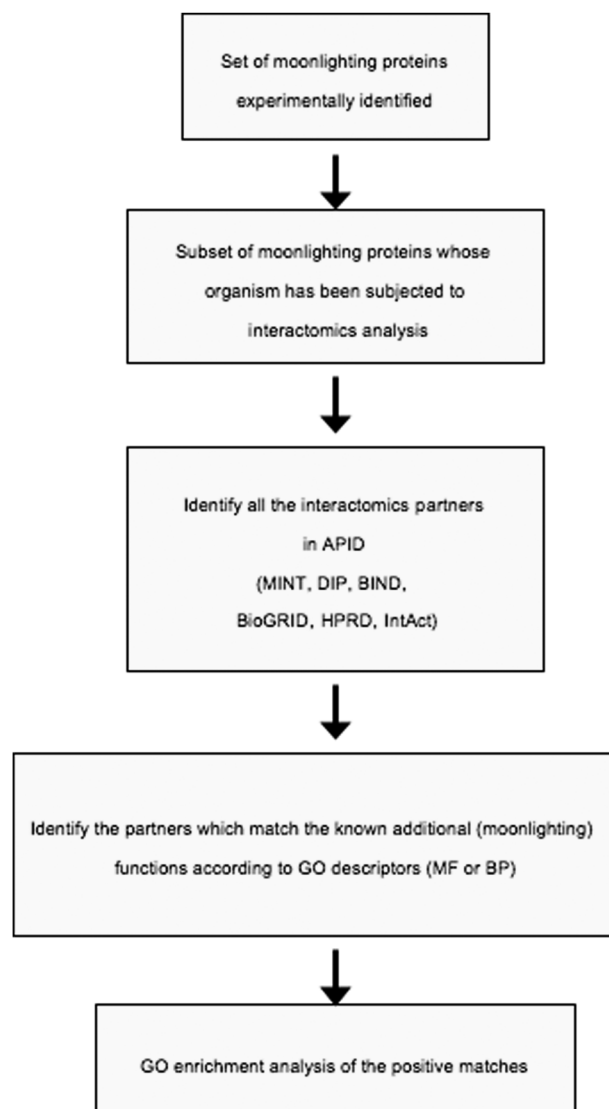


Fig. 1 Scheme of the method used in this work.

^a Institut de Biotecnologia i Biomedicina and Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain.
E-mail: enric.querol@uab.es; Fax: +34-93-5812011;
Tel: +34-93-5868951

^b Department of Systems Biology, Universitat de Vic, Vic, Spain

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c1mb05180f

‡ Present address: Cancer Epigenetics and Biology Program (PEBC), Bellvitge Biomedical Research Institute (IDIBELL), 08908 L'Hospitalet de Llobregat, Barcelona, Catalonia, Spain.

§ The first two authors contributed equally to this work.

Table 1 Demonstrated moonlighting proteins and their reported interacting partners found in APID. The first column shows the protein analyzed; the second column, the known additional function (or functions in some cases). The third column shows those interacting partners contained in APID database, which agree with the reported extra-function according to their GO descriptors (fourth column)

Protein	Known moonlighting functions	Database interacting partners	GO related functions	GO enrichment P-value		
Aconitase	mtDNA maintenance	ATP-dependent DNA helicase MER3	GO:0017111: nucleoside-triphosphatase activity	0.00461		
			GO:0030554: adenylyl nucleotide binding	0.00648		
			GO:0001883: purine nucleoside binding	0.00664		
			GO:0001882: nucleoside binding	0.00685		
			GO:0008135: translation factor activity, nucleic acid binding	0.00017		
Aldolase	Vacuolar H ⁺ -ATPase assembly	V-type proton ATPase subunit E 1	GO:0008553: hydrogen-exporting ATPase activity	0.00361		
			GO:0042623: ATPase activity, coupled	0.00615		
			GO:0051117: ATPase binding	0.00677		
			GO:0046961: proton-transporting ATPase activity, rotational	0.00857		
			GO:0016887: ATPase activity	0.00857		
Enolase	Bind to cytoskeletal structures	Actin	GO:0034621: cellular macromolecular complex organization	7.54×10^{-5}		
			GO:0032506: cytokinetic process	0.0053		
			GO:0007109: cytokinesis, completion of separation	0.0021		
			GO:0007017: microtubule-based process	0.00286		
			GO:0051488: activation of anaphase-promoting complex	0.00314		
		Microtubule-associated protein 4	GO:0000920: cytokinetic cell separation	0.00418		
			GO:0051015: actin filament binding	0.0071		
		Glyceraldehyde-3-phosphate dehydrogenase	Microtubule bundling	Tubulin polymerization-promoting protein	GO:0001948: beta-catenin binding	0.00594
					GO:0008017: microtubule binding	0.00251
			Phosphate group transfer	Phosphoglycerate kinase 1	GO:0017111: nucleoside-triphosphatase activity	0.00222
GO:0016462: pyrophosphatase activity	0.00316					
GO:0016772: transferase activity, transferring phosphorus-containing groups	0.00104					
GO:0003727: single-stranded RNA binding	0.00788					
Binds to RNA, RNA polymerase	Heterogeneous nuclear ribonucleoprotein Q		GO:0008266: poly(U) RNA binding	0.00094		
			GO:0003723: RNA binding	5.11×10^{-13}		
Decrease blood insulin levels	Growth factor receptor-bound protein 2		GO:0043567: regulation of insulin-like growth factor receptor signaling pathway	0.00593		
Nuclear tRNA export	Ataxin-1		GO:0050658: RNA transport	0.001658		
		GO:0050657: nucleic acid transport	0.001658			
		GO:0051236: establishment of RNA localization	0.001658			
Significant role in apoptosis	TNF receptor-associated factor 1	GO:0042981: regulation of apoptosis	4.03×10^{-6}			
		GO:0006915: apoptosis	1.93×10^{-5}			
		GO:0043065: positive regulation of apoptosis	0.00053			
		GO:0006917: induction of apoptosis	0.00199			
		GO:0016192: vesicle-mediated transport	0.00168			
Glycogen synthase kinase 3 Beta	Vesicular transport	Protein kinase C iota type Synphilin-1	GO:0048667: cell morphogenesis involved in neuron differentiation	1.25×10^{-6}		
			GO:0048812: neuron projection morphogenesis	1.59×10^{-6}		
			GO:0048699: generation of neurons	5.73×10^{-6}		
			GO:0030182: neuron differentiation	3.53×10^{-5}		
			GO:0010975: regulation of neuron projection development	0.00013		
			GO:0045664: regulation of neuron differentiation	0.00032		
			GO:0051402: neuron apoptosis	0.0007		
			GO:0005497: androgen binding	0.00619		
			GO:0008266: poly(U) RNA binding	0.0093		
			GO:0045744: negative regulation of G-protein coupled receptor protein signaling pathway	0.00053		
Pyruvate kinase	Regulates transcriptional responses of the thyroid hormone-binding protein	Nuclear hormone receptor family member nhr-111	GO:0043433: negative regulation of transcription factor activity	0.00464		

Table 1 (continued)

Protein	Known moonlighting functions	Database interacting partners	GO related functions	GO enrichment <i>P</i> -value
Alpha-crystallin A	Heat-shock protein	Heat shock protein beta-1	GO:0043066 = negative regulation of apoptosis	3.91×10^{-6}
			GO:0010941 = regulation of cell death	1.05×10^{-5}
			GO:0006915 = apoptosis	3.85×10^{-5}
			GO:0012501 = programmed cell death	4.01×10^{-5}
			GO:0051082: unfolded protein binding	0.0023
Alpha-crystallin B	Heat-shock protein	Heat shock protein beta-1 Heat shock protein beta-8	GO:0043027: caspase inhibitor activity	0.0077
			GO:0051082: unfolded protein binding	0.0011
			GO:0060561 = apoptosis involved in morphogenesis	3.07×10^{-6}
			GO:0006916 = anti-apoptosis	0.00015
			GO:0043066 = negative regulation of apoptosis	0.000717

Not only interacting partners of a protein could suggest its function (“guilty-by-association”) but also the existence of putative additional functions. It is known that function is closely linked to interaction. In any case, it has to be kept in mind that the current set of demonstrated moonlighting proteins is small and the space of interacting proteins is quite incomplete.

Some authors have pointed out that there is a relationship between protein conformational fluctuations and promiscuous functions in proteins. This promiscuity is possible thanks to the structural properties of the structurally disordered regions⁷ that can facilitate the evolution of these proteins to achieve extra functions.⁸ In solution, proteins exist in a range of conformations,⁹ and structurally disordered regions can alter their secondary-structure propensities as well as conformational flexibility in response to different environments or to the interacting partners.¹⁰ Depending on these factors, this produces a selection of certain conformations or even a population shift that would be responsible for this functional promiscuity. Recently, it has been stated that protein disorder could be related to moonlighting in order to act as a hub.¹¹ In fact, it has been suggested that disordered regions are involved in the moonlighting process in several neurological diseases.¹¹

Fig. 1 shows a scheme of the procedure followed in our analysis. Specifically, a number of moonlighting protein examples, which are present in the protein–protein interacting (PPI) databases, have been taken from the literature.^{2,4,12,13} All the proteins of these references which have one or more partners in the APID database have been analyzed. The protein partners for these moonlighting proteins have been checked in the APID PPI database¹⁴ at <http://bioinfow.dep.usal.es/apid/index.htm>, which comprises most of the proteomics data reported elsewhere (MINT, DIP, BioGRID, IntAct, HPRD and BIND). To be functional, proteins need to bind their partners; expressing function in the cell entails a network of binding events. Function is based on events taking place following binding, however, the binding may or may not specify function.¹⁵ Some partners function descriptors are closely linked to a specific function, while others function descriptors are insufficiently linked to define a specific function. To deal with this, a Gene Ontology enrichment analysis was conducted. We have considered that the proteomics data disclose a second function for the moonlighting protein if the PPI database identifies a *Molecular Function* or, in some

cases, a *Biological Process* according to the Gene Ontology annotation,¹⁶ which is in agreement with the expected additional function. Gene Ontology screening (www.geneontology.org) can be performed directly from the APID database. Then a Gene Ontology enrichment analysis was performed. For each moonlighting protein detected in APID, we collected the GO terms from its interacting partners and computed GO term enrichment using GOSTat R package.¹⁷ This function will compute hypergeometric *p*-values for overrepresentation of each GO term in the specified category among the GO annotations for the interesting genes. We selected as a true moonlighting function indicator, those GO terms with a *p*-value lower than 0.05. These values are depicted in the last column of Table 1.

Table 1 shows the interacting partners for a number of moonlighting proteins. In ESI† can be found an extension of Table 1 for other well known moonlighting proteins, mainly ribosomal. As can be seen in these tables, it becomes evident that PPI databases disclose most of the moonlighting activities. Therefore, PPI databases could be a valuable tool to suggest multifunctionality. It deserves to be mentioned that, with the exception of G3PDH, none of the set of moonlighting proteins can be considered putative sticky proteins (considering sticky as being those with > 50 partners).¹⁸

APID database has been built from several interactomics databases. In order to compare the achievement of identifying a moonlighting protein—according to the procedure described above—of these different databases, we have selected as true positives those proteins which achieve a high Connectivity and Clustering Coefficient with the query protein. All of the true positives were counted for all of the cases analysed for each PPI database and were divided by the total number of each analysed case (*i.e.* collect all the proteins that interact with the analysed proteins in each PPI). Table 2 shows the accuracy results.

Table 2 Percentage of identification of known moonlighting proteins for the different PPI databases contained in APID

PPI database	Accuracy
MINT	0.6
DIP	0.833
BioGRID	0.75
IntAct	0.756
HPRD	0.714
BIND	0.65

It is generally considered that experimental data from proteomics contain many false positives, estimated to be up to about 20%.^{19–21} This may easily induce proteomics researchers to consider most of the unexpected partners as false positives. This may represent a drawback for identifying true multifunctional proteins. For example, ribosomal proteins are generally considered false positives in the yeast two-hybrid method. However, this protein class is prone to moonlighting,¹³ and a number of them could be true positives. Another example is Aconitase: using the PPI Curate Yeast High Confidence Network database (HC),²² a well-demonstrated moonlighting protein with several interactors such as Aconitase (YLR304C) has been deleted in the curing process. Therefore, searching for putative moonlighting proteins should be performed on less curated databases such as DIP, MINT, APID.

Obviously, the number of partners of the moonlighting proteins found in all of the PPI database can be high. The proteins in Table 1 present from 6 to 20 partners (human Glyceraldehyde 3 phosphate dehydrogenase is an exception with 83 partners). Therefore, although PPI databases can disclose multifunctional proteins (which is the objective of our communication) to pick out the true partners that lead to identify extra-functionality, this will not be, for the moment, an easy task. For the moment, moonlighting true positives would be found experimentally, and likely by serendipity.

Increasing the number of protein functions without augmenting the number of genes can be achieved by several main mechanisms: alternative splicing,²³ moonlighting.^{2–4} Contrary to splicing, moonlighting can be used by microorganisms. For example, the classic example of a minimal cell, *Mollicutes (Mycoplasmas)*, seems to make extensive use of moonlighting.^{24,25} As stated by Jeffery,²⁶ current moonlighting proteins “appear to be only the tip of the iceberg”.

Acknowledgements

This research was supported by Grants BIO2007-67904-C02-01 and BFU2010-22209-C02-01 from the MCYT (Ministerio de Ciencia y Tecnología, Spain) and from the Centre de Referència de R+D de Biotecnologia de la Generalitat de Catalunya. EQ is Chair Holder of a “Càtedra de Transferència de Tecnologia Parc de Recerca UAB-Grupo Santander”. AG acknowledges a postdoctoral fellowship from “Parc de Recerca UAB-Grupo Santander”. The English of this manuscript has been corrected by Mr. Charles J. Simmons, a native English-speaking Instructor of English of UAB University.

References

- 1 C. Gancedo and C. L. Flores, *Microbiol Mol. Biol. Rev.*, 2008, **72**, 197–210, table of contents.
- 2 C. J. Jeffery, *Trends Biochem. Sci.*, 1999, **24**, 8–11.
- 3 C. J. Jeffery, *Trends Genet.*, 2003, **19**, 415–417.
- 4 C. J. Jeffery, *Mol. Biosyst.*, 2009, **5**, 345–350.
- 5 I. Nobeli, A. D. Favia and J. M. Thornton, *Nat. Biotechnol.*, 2009, **27**, 157–167.
- 6 A. Gomez, N. Domedel, J. Cedano, J. Pinol and E. Querol, *Bioinformatics (Oxford, England)*, 2003, **19**, 895–896.
- 7 C. J. Tsai, B. Ma, Y. Y. Sham, S. Kumar and R. Nussinov, *Proteins*, 2001, **44**, 418–427.
- 8 G. Amitai, R. D. Gupta and D. S. Tawfik, *HFSP J.*, 2007, **1**, 67–78.
- 9 B. Ma, S. Kumar, C. J. Tsai and R. Nussinov, *Protein Eng.*, 1999, **12**, 713–720.
- 10 C. J. Tsai, B. Ma and R. Nussinov, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 9970–9972.
- 11 S. Raychaudhuri, S. Dey, N. P. Bhattacharyya and D. Mukhopadhyay, *PLoS One*, 2009, **4**, e5566.
- 12 G. Sriram, J. A. Martinez, E. R. McCabe, J. C. Liao and K. M. Dipple, *Am. J. Hum. Genet.*, 2005, **76**, 911–924.
- 13 I. G. Wool, *Trends Biochem. Sci.*, 1996, **21**, 164–165.
- 14 C. Prieto and J. De Las Rivas, *Nucleic Acids Res.*, 2006, **34**, W298–W302.
- 15 B. Ma, C. J. Tsai, Y. Pan and R. Nussinov, *ACS Chem. Biol.*, 2010, **19**, 265–272.
- 16 M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin and G. Sherlock, *Nat. Genet.*, 2000, **25**, 25–29.
- 17 T. Beissbarth and T. P. Speed, *Bioinformatics (Oxford, England)*, 2004, **20**, 1464–1465.
- 18 R. Sharan, S. Suthram, R. M. Kelley, T. Kuhn, S. McCuine, P. Uetz, T. Sittler, R. M. Karp and T. Ideker, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 1974–1979.
- 19 H. Huang and J. S. Bader, *Bioinformatics (Oxford, England)*, 2009, **25**, 372–378.
- 20 S. Lievens, S. Eyckerman, I. Lemmens and J. Tavernier, *Expert Rev. Proteomics*, 2010, **7**, 679–690.
- 21 M. Zhang and L. J. Lu, *BMC Bioinf.*, 2010, **11**, 466.
- 22 N. N. Batada, T. Regul, A. Breitkreutz, L. Boucher, B. J. Breitkreutz, L. D. Hurst and M. Tyers, *PLoS Biol.*, 2006, **4**, e317.
- 23 T. W. Nilsen and B. R. Graveley, *Nature*, **463**, 457–463.
- 24 S. Kuhner, V. van Noort, M. J. Betts, A. Leo-Macias, C. Batisse, M. Rode, T. Yamada, T. Maier, S. Bader, P. Beltran-Alvarez, D. Castano-Diez, W. H. Chen, D. Devos, M. Guell, T. Norambuena, I. Racke, V. Rybin, A. Schmidt, E. Yus, R. Aebersold, R. Herrmann, B. Bottcher, A. S. Frangakis, R. B. Russell, L. Serrano, P. Bork and A. C. Gavin, *Science (New York, N.Y.)*, 2009, **326**, 1235–1240.
- 25 J. D. Pollack, M. V. Williams and R. N. McElhane, *Crit. Rev. Microbiol.*, 1997, **23**, 269–354.
- 26 C. J. Jeffery, *Curr. Opin. Struct. Biol.*, 2004, **14**, 663–668.