Do protein–protein interaction databases identify moonlighting proteins?†

Antonio Gómez,⁎a Sergio Hernández,⁎a Isaac Amelia, 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.

† 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.

Fig. 1 Scheme of the method used in 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).

<table>
<thead>
<tr>
<th>Protein</th>
<th>Known moonlighting functions</th>
<th>Database interacting partners</th>
<th>GO related functions</th>
<th>GO enrichment P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aconitase</td>
<td>mtDNA maintenance</td>
<td>ATP-dependent DNA helicase MERT3</td>
<td>GO:0017111: nucleoside-triphosphatase activity &lt;br&gt; GO:0030554: adenyl nucleotide binding &lt;br&gt; GO:0001883: purine nucleoside binding &lt;br&gt; GO:0001882: nucleoside binding &lt;br&gt; GO:0008135: translation factor activity, nucleic acid binding</td>
<td>0.00461</td>
</tr>
<tr>
<td>Aldolase</td>
<td>Vacuolar H + -ATPase assembly</td>
<td>V-type proton ATPase subunit E 1</td>
<td>GO:0008553: hydrogen-expressing ATPase activity &lt;br&gt; GO:0042623: ATPase activity, coupled &lt;br&gt; GO:0051117: ATPase binding &lt;br&gt; GO:0046961: proton-transporting ATPase activity, rotational</td>
<td>0.00361</td>
</tr>
<tr>
<td>Enolase</td>
<td>Bind to cytoskeletal structures</td>
<td>Actin</td>
<td>GO:0034621: cellular macromolecular complex organization &lt;br&gt; GO:0032506: cytokinetic process &lt;br&gt; GO:0007109: cytokinesis, completion of separation</td>
<td>7.54 × 10⁻⁵</td>
</tr>
<tr>
<td>Glyceraldehyde-3-phosphate dehydrogenase</td>
<td>Microtubule bundling</td>
<td>Tubulin polymerization-promoting protein</td>
<td>GO:0051015: actin filament binding &lt;br&gt; GO:0000920: microtubule-based process &lt;br&gt; GO:0051488: activation of anaphase-promoting complex</td>
<td>0.00254</td>
</tr>
<tr>
<td>Phosphate group transfer</td>
<td></td>
<td>Phosphoglycerate kinase 1</td>
<td>GO:0017111: nucleoside-triphosphatase activity &lt;br&gt; GO:0016687: ATPase activity &lt;br&gt; GO:0016772: transferase activity, transferring phosphorus-containing groups</td>
<td>0.000857</td>
</tr>
<tr>
<td>Binds to RNA, RNA polymerase</td>
<td>Heterogeneous nuclear ribonucleoprotein Q</td>
<td></td>
<td>GO:0003727: single-stranded RNA binding &lt;br&gt; GO:008266: poly(U) RNA binding &lt;br&gt; GO:0003723: RNA binding</td>
<td>0.000788</td>
</tr>
<tr>
<td>Decrease blood insulin levels</td>
<td>Growth factor receptor-bound protein 2</td>
<td></td>
<td>GO:0043567: regulation of insulin-like growth factor receptor signaling pathway</td>
<td>0.00593</td>
</tr>
<tr>
<td>Nuclear tRNA export</td>
<td>Ataxin-1</td>
<td></td>
<td>GO:00050658: RNA transport &lt;br&gt; GO:00050657: nucleic acid transport &lt;br&gt; GO:0051236: establishment of RNA localization</td>
<td>0.001658</td>
</tr>
<tr>
<td>Significant role in apoptosis</td>
<td>TNF receptor-associated factor 1</td>
<td></td>
<td>GO:0042981: regulation of apoptosis &lt;br&gt; GO:0006915: apoptosis &lt;br&gt; GO:0043065: positive regulation of apoptosis</td>
<td>4.03 × 10⁻⁶ &lt;br&gt; 1.93 × 10⁻⁵ &lt;br&gt; 0.00053</td>
</tr>
<tr>
<td>Glycogen synthase kinase 3 Beta</td>
<td>Establishment and maintenance of neuronal polarity</td>
<td>Protein kinase C iota type Synphilin-1</td>
<td>GO:0006917: induction of apoptosis &lt;br&gt; GO:0016192: vesicle-mediated transport &lt;br&gt; GO:0048667: cell morphogenesis involved in neuron differentiation &lt;br&gt; GO:0048812: neuron projection morphogenesis</td>
<td>0.00199 &lt;br&gt; 0.00168 &lt;br&gt; 1.25 × 10⁻⁶ &lt;br&gt; 1.59 × 10⁻⁶</td>
</tr>
<tr>
<td>Pyruvate kinase</td>
<td>Regulates transcriptional responses of the thyroid hormone-binding protein</td>
<td>Nuclear hormone receptor family member nhr-111</td>
<td>GO:005497: androgen binding &lt;br&gt; GO:0005497: cell morphogenesis involved in neuron differentiation &lt;br&gt; GO:0045744: negative regulation of G-protein coupled receptor protein signaling pathway</td>
<td>0.0007 &lt;br&gt; 0.00619 &lt;br&gt; 0.00032 &lt;br&gt; 0.00053</td>
</tr>
</tbody>
</table>
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 regions\(^1\) that can facilitate the evolution of these proteins to achieve extra functions.\(^9\) In solution, proteins exist in a range of conformations,\(^7\) 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.\(^11\)

Table 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 database\(^14\) at http://bioinfo.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 GOSepStat 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\(^1\) 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 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.
It is generally considered that experimental data from proteomics contain many false positives, estimated to be up to about 20%. 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, and a number of them could be true positives. Another example is Aconitase: using the PPI Curate Yeast High Confidence Network database (HC), 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, moonlighting. Contrary to splicing, moonlighting can be used by microorganisms. For example, the classic example of a minimal cell, Mollicutes (Mycoplasma), seems to make extensive use of moonlighting. As stated by Jeffery, 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 Referencia de R+D de Bioteconomía de la Generalitat de Catalunya. EQ is Chair Holder of a “Catedra de Transferencia 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