Computational characterization of moonlighting proteins

Ishita K. Khan* **and Daisuke Kihara*†[1]**

*Department of Computer Science, Purdue University, West Lafayette, IN 47907, U.S.A.
†Department of Biological Science, Purdue University, West Lafayette, IN 47907, U.S.A.

Abstract

Moonlighting proteins perform multiple independent cellular functions within one polypeptide chain. Moonlighting proteins switch functions depending on various factors including the cell-type in which they are expressed, cellular location, oligomerization status and the binding of different ligands at different sites. Although an increasing number of moonlighting proteins have been experimentally identified in recent years, the quantity of known moonlighting proteins is insufficient to elucidate their overall landscape. Moreover, most moonlighting proteins have been identified as a serendipitous discovery. Hence, characterization of moonlighting proteins using bioinformatics approaches can have a significant impact on the overall understanding of protein function. In this work, we provide a short review of existing computational approaches for illuminating the functional diversity of moonlighting proteins.

Introduction

With the increase in the number of functionally well-characterized proteins, as well as the advancement of large-scale proteomics studies, more and more proteins have been observed to exhibit more than one cellular function. These proteins were named as 'moonlighting' proteins first by Jeffrey [1]. A moonlighting protein demonstrates multiple autonomous and usually unrelated functions. The diversity of dual functions of these proteins is, in principle, not a consequence of gene fusions, splice variants, multiple proteolytic fragments, homologous but non-identical proteins or varying post-transcriptional modifications. Moonlighting proteins are not limited to a certain type of organism or protein family, nor do they have common switching mechanisms through which they moonlight. The known mechanisms for switching functions include expression of cell type, cellular localization, oligomerization state and identity of binding ligand [1].

It was identified that crystallin, a structural protein in the eye lens of several species, also has enzymatic activity [2]; this was one of the first examples of multifunctional proteins. Many known moonlighting proteins were originally recognized as enzymes, but there are also others that are known as receptors, channel proteins, chaperon proteins, ribosomal proteins and scaffold proteins [1,3,4]. The secondary or moonlighting functions of these proteins include transcriptional regulation, receptor binding, involvement in apoptosis and other regulatory functions. So far, the identification of moonlighting proteins has been done by experiments and reviews of these proteins exist in the literature [1,3–6]. Studies suggest significant effects of moonlighting proteins in diseases and disorders [7–9]. Despite the potential abundance of moonlighting proteins in various genomes and their important roles in pathways and disease development, the number of currently confirmed moonlighting proteins is still too small to obtain a comprehensive picture of the cellular mechanisms underlying their functional diversity. This quantitative insufficiency is, in large part, due to the tendency for the additional function of these proteins to be found serendipitously in the course of unrelated experiments. Hence, a systematic bioinformatics approach could make substantial contributions in identifying novel moonlighting proteins and also in elucidating functional characteristics of moonlighting proteins.

In the present article, we review existing computational analyses on moonlighting proteins. First, we discuss two studies that investigated whether existing sequence-based function prediction methods can identify distinct dual functions of moonlighting proteins [10,11]. Secondly, we review another study by Gómez et al. [12] on analysis of protein–protein interactions (PPIs) of moonlighting proteins where they examined whether the interacting partners of moonlighting proteins disclose the moonlighting function or not. Thirdly, we analyse the study by Hernández et al. [13] where they explored structural aspects of known moonlighting proteins to identify whether the promiscuous functionality of these proteins are caused by the conformational fluctuations in their structures. Then we introduce recently developed databases of moonlighting proteins [14]. Lastly, we discuss the current situation of Gene Ontology (GO) [15] annotations of known moonlighting proteins in the UniProt database [16].

Key words: bioinformatics, computational analysis, moonlighting protein, multitasking protein, protein function.

Abbreviations: BLOSUM, Blocks Substitution Matrix; BP, biological process; ESG, Extended Similarity Group; GO, Gene Ontology; IDP, intrinsic disordered protein; MF, molecular function; PFP, Protein Function Prediction; PPI, protein–protein interaction; PSI-BLAST, Position-Specific Iterated-Basic Local Alignment Search Tool.

[1]To whom correspondence should be addressed (email dkihara@purdue.edu).

Moonlighting proteins pose a challenge in bioinformatics research

A review by Jeffery [4] discusses moonlighting proteins in the context of systems-level proteomics studies and present challenges for computational analyses. Most sequence-based function prediction methods are based on homology searches or motif/domain identifications. Moonlighting proteins can complicate this approach since there are cases in which orthologous proteins in different organisms do not share moonlighting functions. Moreover, the possibility of a moonlighting function would change how we treat the existence of motifs in the protein that have been identified with less confidence, since those hits may explain the moonlighting function of the protein. From a structural point of view, moonlighting proteins could be identified by discovery of multiple ligand-binding sites. Last but not least, moonlighting functions have implications in the discovery of drug-targets and biomarkers, since the knowledge of all functions of a target protein is necessary to design drugs that only affect the desired function of the target.

Sequence-based function prediction for moonlighting proteins

Conventional sequence-based functional annotation methods are based on the concept of homology [17,18] or conserved motifs/domains [19–21]. Two studies have investigated how well current sequence-based methods identify the distinct dual functions of moonlighting proteins. In one of the works we have benchmarked performance of three sequence-based function prediction methods, the Protein Function Prediction (PFP) algorithm [22,23], the Extended Similarity Group (ESG) algorithm [24] and Position-specific Iterated-Basic Local Alignment Search Tool (PSI-BLAST) [25], on a set of experimentally known moonlighting proteins [5]. PFP extends a traditional PSI-BLAST search by extracting and scoring GO annotations from distantly similar sequences and then applying contextual associations of GO terms observed in the annotation database. ESG performs an iterative sequence database search and assigns probabilities to GO terms. PFP and ESG have different characteristics: PFP is designed to have larger coverage by retrieving annotations from weakly similar sequences whereas ESG provides better specificity by taking consistently predicted GO terms from an iterative search.

In the performance evaluation for predicting the diverse functions of moonlighting proteins [5], we compared the predicted GO terms by PFP, ESG and PSI-BLAST with those from both primary and moonlighting functions. In the average precision-recall for the 19 moonlighting proteins, ESG showed the highest precision for a recall range of 0.4–0.7, whereas PFP out-performed the other methods in recall. ESG had lowest recall among the three methods except for five cases. For PSI-BLAST, we used BLOSUM45 (Blocks Substitution Matrix 45) and BLOSUM30 (Blocks Substitution Matrix 30) in addition to the default BLOSUM60 (Blocks Substitution Matrix 60) in order to consider more distantly related sequences. Recall by PSI-BLAST improved using BLOSUM45. In the head-to-head comparison against PFP, PSI-BLAST with BLOSUM45 showed a higher recall than PFP for eight proteins whereas PFP had a higher recall in 10 cases (one protein had a tie). PSI-BLAST with BLOSUM30 failed to predict any GO terms above an *E*-value of 0.01 for 12 proteins. Overall, PFP and PSI-BLAST with BLOSUM45 showed higher recall than the rest of the methods.

These results highlighted the advantage of PFP in predicting the diverse functions of moonlighting proteins with high recall. Incorporating the BLOSUM45 matrix improved recall of PSI-BLAST greatly, which provides another indication that considering weakly similar sequences enhances the prediction of moonlighting functions of proteins.

The second work, by Gómez et al. [11], compared the performance of homology-based and motif/domain-based methods in retrieving sequences with primary and/or moonlighting functions using a dataset of 46 moonlighting proteins. They compared PSI-BLAST and ten motif/domain-based methods. For a dataset of 46 moonlighting proteins, the authors ran the 11 methods and retrieved all the sequences that matched a query protein above a certain standard score cut-off. If any of the retrieved sequences had the primary or secondary function of the query moonlighting protein, it was considered as a 'positive match' for that function. For example, for the moonlighting protein FtsH (primary function: protease, moonlighting function: chaperone), the PSI-BLAST output contained two matched sequences (both with *E*-value of 0.0): gi5231279 and gi12724524, which are a proteinase and a heat-shock protein respectively. In this case, both sequences were considered a 'positive match', the first for the primary function and the second for the moonlighting function. Among the methods tested, PSI-BLAST out-performed others in finding positive matches for both the functionalities of the moonlighting proteins. Among the ten motif/domain-based methods, ProDom performed best. Among the 46 proteins in the dataset, the authors performed structural analysis on four proteins (BirA biotin synthetase, thymidine synthase, aconitase and fructose-1,6-biphosphatase) and found two different functional sites for three of them.

Exploring moonlighting proteins in protein–protein interaction networks

Protein–protein interaction networks provide a useful clue of protein function because proteins of the same biological function or pathways tend to interact [26–31]. Gómez et al. [12] analysed PPI networks of known moonlighting proteins to determine whether interacting proteins of moonlighting proteins possess the secondary functions of the moonlighting proteins.

A set of experimentally identified moonlighting proteins that have known interacting partners in the Agile Protein Interaction DataAnalyzer (APID) database were selected for this analysis [32]. Among these interacting partners, 605 proteins were selected that have GO annotations [in the biological process (BP) or molecular function (MF) categories] that match the function description of the moonlighting function of the query protein. For each of these selected interacting partners of a moonlighting protein, GO terms related to the moonlighting function were collected and a GO term enrichment score (*P*-value from the hypergeometric distribution) was computed using the GOStat package in R. Using a *P*-value cut-off of 0.05, the authors analysed whether secondary functions of query proteins could be predicted. Among the six PPI databases they analysed (MINT, DIP, BioGRID, IntAct, HPRD and BIND), DIP had the highest percentage of identifying the moonlighting function from its interacting partners (0.833) and MINT had the lowest percentage (0.6). The authors concluded that PPI networks contain information that discloses moonlighting functions of proteins.

Moonlighting proteins and disordered regions

Intrinsically disordered regions have been found to have important roles in protein function [33]. The functional diversity of moonlighting proteins could be caused by structurally disordered regions as different conformations of disordered regions may facilitate different functions of a protein or allow a protein to interact with different protein partners.

Tompa et al. [34] reported earlier that some known moonlighting proteins have disordered regions with which they bind the same partner in different conformations and at different binding sites, resulting in opposite effects of inhibiting or activating their interaction partners.

Although some moonlighting proteins exhibit dual function due to disordered regions, this is not the case in the majority of moonlighting proteins. Hernández et al. [13] investigated whether moonlighting proteins tend to have intrinsically disordered regions. Twenty-eight known moonlighting proteins were analysed. Disordered regions of these proteins were predicted by four programs, PrDos, DisEMBL, Disopred and IUpred. It turned out that most of the moonlighting proteins do not have long disordered regions and are not considered as members of the intrinsic disordered protein (IDP) class, which is defined as proteins that have more than 40 residues in disordered regions [35]. Most of the predicted disordered regions for these moonlighting proteins were quite short and in many cases were located at the N- or C-terminal regions of the proteins. On the basis of these results, the authors concluded that most moonlighting proteins do not fall into the IDP class.

Database of moonlighting proteins

Currently, there exist three databases of moonlighting proteins. One of them, MultitaskProtDB (http://wallace.uab.es/multitask/) [14], has compiled 288 multitasking/moonlighting proteins at the time of this writing. This database lists known moonlighting proteins extracted from ten review articles. In addition, the authors performed text mining on articles in PubMed to identify moonlighting proteins using the following keywords: moonlight/moonlighting proteins/enzymes, multitask/multitasking proteins/enzymes and gene sharing. The database holds 288 moonlighting proteins from ∼100 difference organisms, among which 91 are from human (32 %), 23 from yeast (8 %), 23 from *Arabidopsis* (8 %) and 20 from *Escherichia coli* (7 %).

For each protein, users can retrieve its NCBI code, UniProt accession number, species information, canonical and moonlighting functions, PDB codes (if available), oligomeric state (if available) and reference to the corresponding literature. Interestingly, from the database the authors found that the most prevalent canonical/moonlighting GO pair is enzyme/nucleic acid-binding proteins (74 out of 288). For example, proteins that has 'transcription factor' as their secondary function belong to this set. The second most prevalent pair is enzyme/adhesion protein for pathogens (48 out of 288).

MOONPROT (http://moonlightingproteins.org/) is a database compiled by the Jeffery laboratory that stores information about moonlighting proteins for which there exists biochemical or biophysical evidence [36]. It contains 291 proteins. MoonDB (http://tagc.univ-mrs.fr/MoonDB/) contains human moonlighting proteins recovered from the literature and candidates predicted by a protein–protein network-based approach (C.E. Chapple, B. Robisson, C. Herrmann and C. Brun, unpublished work). These databases provide platforms for systematic analysis of multifunctional/moonlighting proteins.

Gene Ontology annotations of moonlighting proteins

Most moonlighting proteins are found serendipitously by experiments. Consequently, the majority of these proteins are best known for their primary function. Partly owing to this fact, annotation in UniProt often lacks GO terms related to their moonlighting functions. In the present paper, we show two such examples of experimentally known moonlighting proteins (Table 1). PFK1 (UniProt ID: Q92448) is an ATP-dependent phosphofructokinase that phosphorylates D-fructose 6-phosphate in the first committed step of the glycolysis pathway. Additionally, this protein has been found to be involved in rapid and selective degradation of peroxisomes by microautophagy [37]. In a PFK1-knockout mutant, peroxisomes are observed to remain outside of vacuole and degradation is disabled. The existing GO annotations for this protein includes 14 GO terms describing its ATP-dependent catalytic activity in glycolysis [5 in BP, 7 in MF and 2 in cellular component (CC)], but lacks

Table 1 | **GO annotations of PFK1 and murL**

The GO terms describing their primary function were from UniProt. GO terms for the moonlighting functions are those we have added. No GO terms were found for the moonlighting functions in UniProt.

Protein name/UniProt ID	GO terms for primary function	Added GO terms for moonlighting function
PFK1/Q92448	GO:0006002: fructose 6-phosphate metabolic process GO:0006096: glycolytic process GO:0008152: metabolic process GO:0016310: phosphorylation GO:0046835: carbohydrate phosphorylation GO:0000166: nucleotide binding GO:0003824: catalytic activity GO:0003872: 6-phosphofructokinase activity GO:0005524: ATP binding GO:0016301: kinase activity GO:0016740: transferase activity GO:0046872: metal ion binding GO:0005945: 6-phosphofructokinase complex GO:0005737: cytoplasm	GO:0016237: microautophagy GO:0010508: positive regulation of autophagy GO:0030242: peroxisome degradation
murL/D3FPC2	GO:0006807: nitrogen compound metabolic process GO:0008152: metabolic process GO:0008360: regulation of cell shape GO:0009252: peptidoglycan biosynthetic process GO:0008881: glutamate racemase activity GO:0016853: isomerase activity GO:0016855: racemase and epimerase activity, acting on amino acids and derivatives GO:0036361: racemase activity, acting on amino acids and derivatives	GO:0008657: DNA topoisomerase (ATP-hydrolysing) inhibitor activity GO:2000372: negative regulation of DNA topoisomerase (ATP-hydrolysing) activity GO:0004857: enzyme inhibitor activity GO:0090143: nucleoid organization

GO terms describing the moonlighting function, 'autophagy peroxisomes'. The second example is glutamate racemase (UniProt ID: D3FPC2). It is an essential enzyme in the cell-wall biosynthesis-pathway in bacteria because it converts D-glutamate into L-glutamate, an important building block for peptidoglycan synthesis. Independent of the enzymatic function, this protein in *Mycobacterium tuberculosis* is shown to have a role as an inhibitor of DNA gyrase [38]. The UniProt entry of this protein has eight GO terms that clearly describe its racemase activity in cell-wall biosynthesis (four in BP and four in MF), but no GO terms regarding the moonlighting function ('DNA gyrase inhibitor'). In Table 1, we listed the GO terms for these two proteins from UniProt as well as GO terms we have chosen that describe the secondary function. As illustrated in these two examples, it is not rare that moonlighting proteins are well-annotated in terms of their primary functions but under-annotated regarding moonlighting functions.

Discussion

We have reviewed existing computational works on moonlighting proteins. These papers analysed moonlighting proteins from several different perspectives, i.e. sequence-based function prediction, PPI and structural properties.

Generally speaking, one advantage of computational analysis is that it can provide a big picture of biological phenomena. However, because the number of known moonlighting proteins is still small, the existing works were based on datasets of limited size. Moreover, annotations in the database often do not reflect the moonlighting functions of these proteins. To enable large-scale computational characterization of moonlighting proteins, a comprehensive online repository with consistent functional annotations is the foremost requirement. In this regard, the three databases, which are currently available and under continuous development, are a good resource for future studies.

Structural analysis can provide a physical concrete picture of moonlighting proteins. Although a drawback of structural analysis is that it is only applicable to proteins that have experimentally solved tertiary structures, it is noteworthy that computationally modelled structures could be used because structure prediction methods have matured in the last few years [39–41]. To aid in finding binding sites of moonlighting proteins, methods for detecting binding-pocket sites in protein structures [42] and predicting binding ligands [43,44] can be useful.

The mechanisms by which moonlighting proteins exhibit multiple functions differ from case to case. Ultimately, an integrative approach will be needed for comprehensive

understanding and classification of moonlighting proteins, which combines various types of data, such as proteomics, phenotypes, genomics and other biochemical data. Investigation of moonlighting proteins is still in its early stage. We foresee that moonlighting proteins will be more systematically studied in the near future and anticipate that computational work will play important roles there.

Acknowledgements

We are grateful to Lenna X. Peterson for critical comments and to Lyman Monroe for proofreading this article before submission.

Funding

This work was supported partly by the National Institutes of Health [grant number R01GM097528], the National Science Foundation [grant numbers IIS1319551, DBI1262189 and IOS1127027], and National Research Foundation of Korea [grant number NRF-2011-220-C00004].

References

1 Jeffery, C. (1999) Moonlighting proteins. Trends Biochem. Sci. **24**, 8–11 CrossRef PubMed
2 Piatigorsky, J. (1998) Multifunctional lens crystallins and corneal enzymes. More than meets the eye. Ann. N.Y. Acad. Sci. **842**, 7–15 CrossRef PubMed
3 Jeffery, C. (2009) Moonlighting proteins: an update. Mol. Biosyst. **5**, 345–350 CrossRef PubMed
4 Jeffery, C.J. (2004) Moonlighting proteins: complications and implications for proteomics research. Drug Discov. Today: Targets **3**, 71–78 CrossRef
5 Huberts, D.H. and van der Klei, I.J. (2010) Moonlighting proteins: an intriguing mode of multitasking. Biochim. Biophys. Acta **1803**, 520–525 CrossRef PubMed
6 Jeffery, C.J. (2005) Mass spectrometry and the search for moonlighting proteins. Mass Spectrom. Rev. **24**, 772–782 CrossRef PubMed
7 Jeffery, C.J. (2011) Proteins with neomorphic moonlighting functions in disease. IUBMB Life **63**, 489–494 CrossRef PubMed
8 Sriram, G., Martinez, J.A., McCabe, E.R., Liao, J.C. and Dipple, K.M. (2005) Single-gene disorders: what role could moonlighting enzymes play? Am. J. Hum. Genet. **76**, 911–924 CrossRef PubMed
9 Ovádi, J. (2011) Moonlighting proteins in neurological disorders. IUBMB Life **63**, 453–457 CrossRef PubMed
10 Khan, I., Chitale, M., Rayon, C. and Kihara, D. (2012) Evaluation of function predictions by PFP, ESG, and PSI-BLAST for moonlighting proteins. BMC Proc. **6** (Suppl. 7), S5 CrossRef PubMed
11 Gómez, A., Domedel, N., Cedano, J., Piñol, J. and Querol, E. (2003) Do current sequence analysis algorithms disclose multifunctional (moonlighting) proteins? Bioinformatics **19**, 895–896 CrossRef PubMed
12 Gómez, A., Hernández, S., Amela, I., Piñol, J., Cedano, J. and Querol, E. (2011) Do protein–protein interaction databases identify moonlighting proteins? Mol. Biosyst. **7**, 2379–2382 CrossRef PubMed
13 Hernández, S., Amela, I., Cedano, J., Piñol, J., Perez-Pons, J., Mozo-Villarias, A. and Querol, E. (2011) Do moonlighting proteins belong to the intrinsically disordered protein class? J. Proteomics Bioinform. **5**, 262–264
14 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 CrossRef PubMed
15 Ashburner, M., Ball, C., Blake, J., Botstein, D., Butler, H., Cherry, J., Davis, A., Dolinski, K., Dwight, S., Eppig, J. et al. (2000) Gene ontology: tool for the unification of biology. Nat. Genet. **25**, 25–34 CrossRef PubMed
16 UniProt Consortium (2014) Activities at the universal protein resource (UniProt). Nucleic Acids Res. **42**, D191–D198 CrossRef PubMed
17 Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. J. Mol. Biol. **215**, 403–410 CrossRef PubMed
18 Pearson, W.R. (1990) Rapid and sensitive sequence comparison with FASTP and FASTA. Methods Enzymol. **183**, 63–98 CrossRef PubMed
19 Bru, C., Courcelle, E., Carrère, S., Beausse, Y., Dalmar, S. and Kahn, D. (2005) The ProDom database of protein domain families: more emphasis on 3D. Nucleic Acids Res. **33**, 212–215 CrossRef
20 Finn, R.D., Mistry, J., Schuster-Böckler, B., Griffiths-Jones, S., Hollich, V., Lassmann, T., Moxon, S., Marshall, M., Khanna, A., Durbin, R. et al. (2006) Pfam: clans, web tools and services. Nucleic Acids Res. **34**, 247–251 CrossRef
21 Hunter, S., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Das, U., Daugherty, L., Duquenne, L. et al. (2009) InterPro: the integrative protein signature database. Nucleic Acids Res. **37**, 211–215 CrossRef
22 Hawkins, T., Luban, S. and Kihara, D. (2006) Enhanced automated function prediction using distantly related sequences and contextual association by PFP. Protein Sci. **15**, 1550–1556 CrossRef PubMed
23 Hawkins, T., Chitale, M., Luban, S. and Kihara, D. (2009) PFP: automated prediction of gene ontology functional annotations with confidence scores using protein sequence data. Proteins **74**, 566–582 CrossRef PubMed
24 Chitale, M., Hawkins, T., Park, C. and Kihara, D. (2009) ESG: extended similarity group method for automated protein function prediction. Bioinformatics **25**, 1739–1745 CrossRef PubMed
25 Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**, 3389–3402 CrossRef PubMed
26 Brun, C., Chevenet, F., Martin, D., Wojcik, J., Guénoche, A. and Jacq, B. (2003) Functional classification of proteins for the prediction of cellular function from a protein–protein interaction network. Genome Biol. **5**, R6 CrossRef PubMed
27 Chua, H.N., Sung, W.K. and Wong, L. (2006) Exploiting indirect neighbours and topological weight to predict protein function from protein–protein interactions. Bioinformatics **22**, 1623–1630 CrossRef PubMed
28 Letovsky, S. and Kasif, S. (2003) Predicting protein function from protein/protein interaction data: a probabilistic approach. Bioinformatics **19**, i197–i204 CrossRef PubMed
29 Nariai, N., Kolaczyk, E.D. and Kasif, S. (2007) Probabilistic protein function prediction from heterogeneous genome-wide data. PLoS ONE **2**, e337 CrossRef PubMed
30 Sharan, R., Ulitsky, I. and Shamir, R. (2007) Network-based prediction of protein function. Mol. Syst. Biol. **3**, 88–100 CrossRef PubMed
31 Deng, M., Tu, Z, Sun, F. and Chen, T. (2004) Mapping gene ontology to proteins based on protein–protein interaction data. Bioinformatics **20**, 895–902 CrossRef PubMed
32 Prieto, C. and De Las Rivas, J. (2006) APID: Agile Protein Interaction DataAnalyzer. Nucleic Acids Res. **34**, W298–W302 CrossRef PubMed
33 Dunker, A.K., Brown, C.J., Lawson, J.D., Iakoucheva, L.M. and Obradović, Z. (2002) Intrinsic disorder and protein function. Biochemistry **41**, 6573–6582 CrossRef PubMed
34 Tompa, P., Szász, C. and Buday, L. (2005) Structural disorder throws new light on moonlighting. Trends Biochem. Sci. **30**, 484–489 CrossRef PubMed
35 Dyson, H.J. (2011) Expanding the proteome: disordered and alternatively folded proteins. Q. Rev. Biophys. **44**, 467–518 CrossRef PubMed
36 Mani, M.K. (2014) Moonlighting protein database (MoonProt): a database for proteins that are known to moonlight. M.S. Dissertation. University of Illinois, Chicago, IL, U.S.A.
37 Yuan, W., Tuttle, D.L., Shi, Y.J., Ralph, G.S. and Dunn, Jr, W.A. (1997) Glucose-induced microautophagy in *Pichia pastoris* requires the α-subunit of phosphofructokinase. J. Cell Sci. **110**, 1935–1945 PubMed
38 Sengupta, S., Ghosh, S. and Nagaraja, V. (2008) Moonlighting function of glutamate racemase from *Mycobacterium tuberculosis*: racemization and DNA gyrase inhibition are two independent activities of the enzyme. Microbiology **154**, 2796–2803 CrossRef PubMed
39 Chen, H. and Kihara, D. (2011) Effect of using suboptimal alignments in template-based protein structure prediction. Proteins **79**, 315–334 CrossRef PubMed

40 Kihara, D., Lu, H., Kolinski, A. and Skolnick, J. (2003) TOUCHSTONE: an *ab initio* protein structure prediction method that uses threading-based tertiary restraints. Proc. Natl. Acad. Sci. U.S.A. **98**, 10125–10130 CrossRef
41 Kihara, D., Zhang, Y., Lu, H, Kolinski, A. and Skolnick, J. (2002) *Ab initio* protein structure prediction on a genomic scale: application to the *Mycoplasma genitalium* genome. Proc. Natl. Acad. Sci. U.S.A. **99**, 5993–5998 CrossRef PubMed
42 Li, B., Turuvekere, S., Agrawal, M., La, D., Ramani, K. and Kihara, D. (2008) Characterization of local geometry of protein surfaces with the visibility criterion. Proteins **71**, 670–683 CrossRef PubMed
43 Chikhi, R., Sael, L. and Kihara, D. (2010) Real-time ligand binding pocket database search using local surface descriptors. Proteins **78**, 2007–2028 CrossRef PubMed
44 Sael, L. and Kihara, D. (2012) Detecting local ligand-binding site similarity in nonhomologous proteins by surface patch comparison. Proteins **80**, 1177–1195 CrossRef PubMed

Received 4 August 2014
doi:10.1042/BST20140214