D368–D372 *Nucleic Acids Research, 2021, Vol. 49, Database issue* *Published online 27 November 2020*
doi: 10.1093/nar/gkaa1101

MoonProt 3.0: an update of the moonlighting proteins database

Chang Chen[1,2,†], Haipeng Liu[3,†], Shadi Zabad[4], Nina Rivera[5], Emily Rowin[5], Maheen Hassan[5], Stephanie M. Gomez De Jesus[6], Paola S. Llinás Santos[6], Karyna Kravchenko[7], Mariia Mikhova[8], Sophia Ketterer[9], Annabel Shen[9], Sophia Shen[9], Erin Navas[10], Bryan Horan[10], Jaak Raudsepp[9] and Constance Jeffery[1,5,*]

[1]Department of Bioengineering, University of Illinois at Chicago, Chicago, IL 60607, USA, [2]Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL 60611, USA, [3]Center for Biomolecular Sciences, College of Pharmacy, University of Illinois at Chicago, Chicago, IL 60612, USA, [4]Department of Computer Science, McGill University, Montreal, QC, Canada, [5]Department of Biological Sciences, University of Illinois at Chicago, Chicago, IL 60607, USA, [6]Universidad de Puerto Rico en Cayey, Cayey, PR 00736, USA, [7]Department of Biotechnology and Bioengineering, V. N. Karazin Kharkiv National University, IL 61002, Ukraine, [8]Minot State University, Minot, ND 58707, USA, [9]Cold Spring Harbor High School, Cold Spring Harbor, NY 11724, USA and [10]Northport High School, Northport, NY 11768, USA

Received September 15, 2020; Revised October 21, 2020; Editorial Decision October 22, 2020; Accepted October 31, 2020

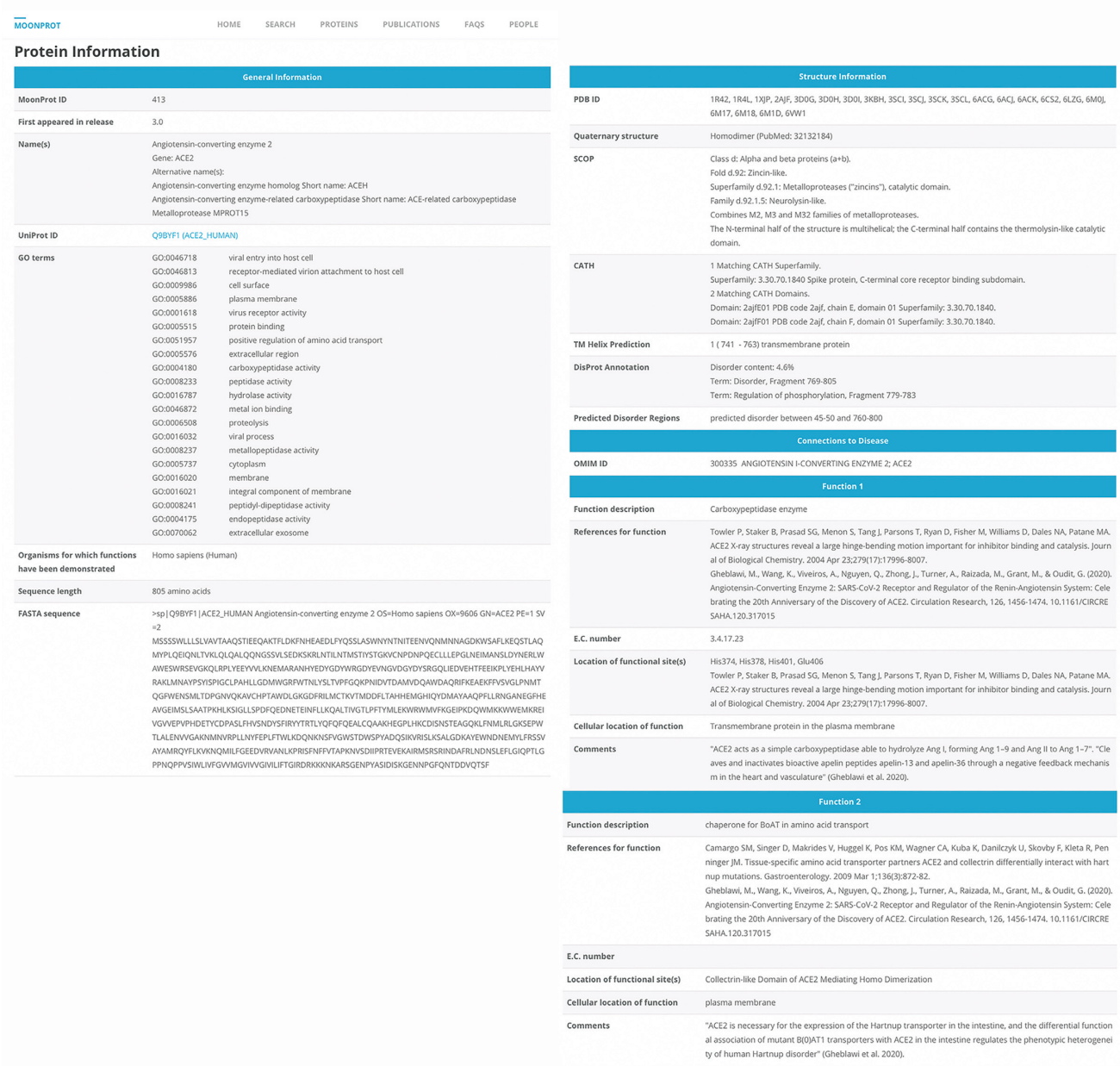
ABSTRACT

MoonProt 3.0 (http://moonlightingproteins.org) is an updated open-access database storing expert-curated annotations for moonlighting proteins. Moonlighting proteins have two or more physiologically relevant distinct biochemical or biophysical functions performed by a single polypeptide chain. Here, we describe an expansion in the database since our previous report in the Database Issue of Nucleic Acids Research in 2018. For this release, the number of proteins annotated has been expanded to over 500 proteins and dozens of protein annotations have been updated with additional information, including more structures in the Protein Data Bank, compared with version 2.0. The new entries include more examples from humans, plants and archaea, more proteins involved in disease and proteins with different combinations of functions. More kinds of information about the proteins and the species in which they have multiple functions has been added, including CATH and SCOP classification of structure, known and predicted disorder, predicted transmembrane helices, type of organism, relationship of the protein to disease, and relationship of organism to cause of disease.

INTRODUCTION

MoonProt is an online resource of information about moonlighting proteins that is manually curated by experts. Moonlighting proteins are proteins in which more than one physiologically relevant discrete biochemical or biophysical function is performed by a single polypeptide chain (1–3). Some of the first moonlighting proteins to be identified were the taxon specific crystallins. These crystallins are metabolic enzymes and chaperones that have been adopted to perform a second, noncatalytic function in the lens or cornea the eyes of a few species (4). Since then hundreds of moonlighting proteins with many different functions have been found throughout the evolutionary tree (1–12). Because a large variety of proteins have evolved to perform multiple functions, there is no shared sequence or structural feature that can be used to indicate or predict whether or not a protein is a moonlighting protein. Instead, second functions are often found through serendipity, and information about these proteins is scattered in diverse publications. Our MoonProt Database brings this information together to help researchers learn about proteins with multiple functions and provides a quick method to find out if a protein of interest is a known moonlighting protein or related to a known moonlighting protein (Figure 1). In addition, this collection of information about the sequences, structures and functions of known moonlighting proteins can aid in several current areas of research, including the analysis of protein structure and function, interpreting genome sequences and the results of proteomics studies, elucidating the evolution of

*To whom correspondence should be addressed. Tel: +1 312 996 3168; Fax: +1 312 413 2691; Email: cjeffery@uic.edu
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.



MOONPROT HOME SEARCH PROTEINS PUBLICATIONS FAQS PEOPLE

Protein Information

General Information	
MoonProt ID	413
First appeared in release	3.0
Name(s)	Angiotensin-converting enzyme 2 Gene: ACE2 Alternative name(s): Angiotensin-converting enzyme homolog Short name: ACEH Angiotensin-converting enzyme-related carboxypeptidase Short name: ACE-related carboxypeptidase Metalloprotease MPROT15
UniProt ID	Q9BYF1 (ACE2_HUMAN)
GO terms	GO:0046718 viral entry into host cell GO:0046813 receptor-mediated virion attachment to host cell GO:0009986 cell surface GO:0005886 plasma membrane GO:0001618 virus receptor activity GO:0005515 protein binding GO:0051957 positive regulation of amino acid transport GO:0005576 extracellular region GO:0004180 carboxypeptidase activity GO:0008233 peptidase activity GO:0016787 hydrolase activity GO:0046872 metal ion binding GO:0006508 proteolysis GO:0016032 viral process GO:0008237 metallopeptidase activity GO:0005737 cytoplasm GO:0016020 membrane GO:0016021 integral component of membrane GO:0008241 peptidyl-dipeptidase activity GO:0004175 endopeptidase activity GO:0070062 extracellular exosome
Organisms for which functions have been demonstrated	Homo sapiens (Human)
Sequence length	805 amino acids
FASTA sequence	>sp|Q9BYF1|ACE2_HUMAN Angiotensin-converting enzyme 2 OS=Homo sapiens OX=9606 GN=ACE2 PE=1 SV=2 MSSSSWLLLSLVAVTAAQSTIEEQAKTFLDKFNHEAEDLFYQSSLASWNYNTNITEENVQNMNNAGDKWSAFLKEQSTLAQMYPLQEIQNLTVKLQLQALQQNGSSVLSEDKSKRLNTILNTMSTIYSTGKVCNPDNPQECLLLEPGLNEIMANSLDYNERLWAWESWRSEVGKQLRPLYEEYVVLKNEMARANHYEDYGDYWRGDYEVNGVDGYDYSRGQLIEDVEHTFEEIKPLYEHLHAYVRAKLMNAYPSYISPIGCLPAHLLGDMWGRFWTNLYSLTVPFGQKPNIDVTDAMVDQAWDAQRIFKEAEKFFVSVGLPNMTQGFWENSMLTDPGNVQKAVCHPTAWDLGKGDFRILMCTKVTMDDFLTAHHEMGHIQYDMAYAAQPFLLRNGANEGFHEAVGEIMSLSAATPKHLKSIGLLSPDFQEDNETEINFLLKQALTIVGTLPFTYMLEKWRWMVFKGEIPKDQWMKKWWEMKREIVGVVEPVPHDETYCDPASLFHVSNDYSFIRYYTRTLYQFQFQEALCQAAKHEGPLHKCDISNSTEAGQKLFNMLRLGKSEPWTLALENVVGAKNMNVRPLLNYFEPLFTWLKDQNKNSFVGWSTDWSPYADQSIKVRISLKSALGDKAYEWNDNEMYLFRSSVAYAMRQYFLKVKNQMILFGEEDVRVANLKPRISFNFFVTAPKNVSDIIPRTEVEKAIRMSRSRINDAFRLNDNSLEFLGIQPTLGPPNQPPVSIWLIVFGVVMGVIVVGIVILIFTGIRDRKKKNKARSGENPYASIDISKGENNPGFQNTDDVQTSF

Structure Information	
PDB ID	1R42, 1R4L, 1XJP, 2AJF, 3D0G, 3D0H, 3D0I, 3KBH, 3SCI, 3SCJ, 3SCK, 3SCL, 6ACG, 6ACJ, 6ACK, 6CS2, 6LZG, 6M0J, 6M17, 6M18, 6M1D, 6VW1
Quaternary structure	Homodimer (PubMed: 32132184)
SCOP	Class d: Alpha and beta proteins (a+b). Fold d.92: Zincin-like. Superfamily d.92.1: Metalloproteases ("zincins"), catalytic domain. Family d.92.1.5: Neurolysin-like. Combines M2, M3 and M32 families of metalloproteases. The N-terminal half of the structure is multihelical; the C-terminal half contains the thermolysin-like catalytic domain.
CATH	1 Matching CATH Superfamily. Superfamily: 3.30.70.1840 Spike protein, C-terminal core receptor binding subdomain. 2 Matching CATH Domains. Domain: 2ajfE01 PDB code 2ajf, chain E, domain 01 Superfamily: 3.30.70.1840. Domain: 2ajfF01 PDB code 2ajf, chain F, domain 01 Superfamily: 3.30.70.1840.
TM Helix Prediction	1 (741 - 763) transmembrane protein
DisProt Annotation	Disorder content: 4.6% Term: Disorder, Fragment 769-805 Term: Regulation of phosphorylation, Fragment 779-783
Predicted Disorder Regions	predicted disorder between 45-50 and 760-800
Connections to Disease	
OMIM ID	300335 ANGIOTENSIN I-CONVERTING ENZYME 2; ACE2
Function 1	
Function description	Carboxypeptidase enzyme
References for function	Towler P, Staker B, Prasad SG, Menon S, Tang J, Parsons T, Ryan D, Fisher M, Williams D, Dales NA, Patane MA. ACE2 X-ray structures reveal a large hinge-bending motion important for inhibitor binding and catalysis. Journal of Biological Chemistry. 2004 Apr 23;279(17):17996-8007. Gheblawi, M., Wang, K., Viveiros, A., Nguyen, Q., Zhong, J., Turner, A., Raizada, M., Grant, M., & Oudit, G. (2020). Angiotensin-Converting Enzyme 2: SARS-CoV-2 Receptor and Regulator of the Renin-Angiotensin System: Celebrating the 20th Anniversary of the Discovery of ACE2. Circulation Research, 126, 1456-1474. 10.1161/CIRCRESAHA.120.317015
E.C. number	3.4.17.23
Location of functional site(s)	His374, His378, His401, Glu406 Towler P, Staker B, Prasad SG, Menon S, Tang J, Parsons T, Ryan D, Fisher M, Williams D, Dales NA, Patane MA. ACE2 X-ray structures reveal a large hinge-bending motion important for inhibitor binding and catalysis. Journal of Biological Chemistry. 2004 Apr 23;279(17):17996-8007.
Cellular location of function	Transmembrane protein in the plasma membrane
Comments	"ACE2 acts as a simple carboxypeptidase able to hydrolyze Ang I, forming Ang 1–9 and Ang II to Ang 1–7". "Cleaves and inactivates bioactive apelin peptides apelin-13 and apelin-36 through a negative feedback mechanism in the heart and vasculature" (Gheblawi et al. 2020).
Function 2	
Function description	chaperone for BoAT in amino acid transport
References for function	Camargo SM, Singer D, Makrides V, Huggel K, Pos KM, Wagner CA, Kuba K, Danilczyk U, Skovby F, Kleta R, Penninger JM. Tissue-specific amino acid transporter partners ACE2 and collectrin differentially interact with hartnup mutations. Gastroenterology. 2009 Mar 1;136(3):872-82. Gheblawi, M., Wang, K., Viveiros, A., Nguyen, Q., Zhong, J., Turner, A., Raizada, M., Grant, M., & Oudit, G. (2020). Angiotensin-Converting Enzyme 2: SARS-CoV-2 Receptor and Regulator of the Renin-Angiotensin System: Celebrating the 20th Anniversary of the Discovery of ACE2. Circulation Research, 126, 1456-1474. 10.1161/CIRCRESAHA.120.317015
E.C. number	
Location of functional site(s)	Collectrin-like Domain of ACE2 Mediating Homo Dimerization
Cellular location of function	plasma membrane
Comments	"ACE2 is necessary for the expression of the Hartnup transporter in the intestine, and the differential functional association of mutant B(0)AT1 transporters with ACE2 in the intestine regulates the phenotypic heterogeneity of human Hartnup disorder" (Gheblawi et al. 2020).

Figure 1. Example of information included in the annotation for the human Angiotensin-Converting Enzyme 2 (ACE2), which is the Sars-Cov2 coronavirus target. ACE2 is a moonlighting protein because it is both (1) an enzyme that cleaves angiotensin to produce bioactive peptides and (2) a chaperone that helps in the proper folding and plasma membrane targeting of the BoAT amino acid transporter. As for many moonlighting proteins, several names of the protein are included in the entry so that the corresponding entries can be found by users using different names. UniProtKB and PDB accession numbers are included to link to additional resources. The GO terms illustrate the protein's functions, cellular locations, and processes it is involved in. An Enzyme Commission number provides information about the type of catalytic activity and substrates. The species of organism for which the protein has been shown to have more than one function and the amino acid sequence in FASTA format are provided to clarify which version of a protein has been shown to have two functions. One or more peer-reviewed publications describing experiments that demonstrated the protein functions are also included. For some proteins and organisms, information about connections to disease have been added. For ACE2, both of the protein's functions are involved in diseases in humans. In addition, the ACE2 protein is a 'receptor' used to invade host cells by the Sars-Cov2 virus, which caused the coronavirus pandemic of 2020. Although binding of a pathogen's protein to a protein on a host cell is not a function that arose during evolution of the host protein but instead is a function of the pathogen's protein, ACE2 still qualifies as a moonlighting protein because it has a catalytic function and a chaperone function.

protein function, and design of proteins or nanomachines with novel functions.

Our lab constructed MoonProt, the open-access online Moonlighting Proteins Database (http://www.moonlightingproteins.org/) in 2014 and updated its features and expanded the number of proteins included in 2018 (13,14). In this paper, we present MoonProt version 3.0. Since its previous update, the database has grown to include annotation for over 500 proteins and the information about individual moonlighting proteins has been expanded and updated.

MATERIALS AND METHODS

Selection of moonlighting proteins included in the database

For inclusion of a protein in the MoonProt Database, peer-reviewed published biochemical, biophysical, mutagenic or other data to support the presence of multiple physiologically-relevant functions was required and was critically reviewed by the PI. Proteins were not included if the 'multiple functions' described in publications are due to different RNA splice variants, the same function performed in two different cellular or tissue locations, pleiotropic effects on multiple pathways or multiple physiological processes, or a family of proteins in which the different functions are performed by different proteins. Proteins were not included if the 'multiple functions' are simply different aspects of the same function (i.e. 'membrane protein' and 'transmembrane transporter').

Information included about individual proteins

As described for versions 1.0 and 2.0 (13,14), information about each protein was manually curated from published journal articles and online resources. The entry for each protein includes a description of each function and one or more references for publications providing experimental evidence of each function. The specific organism and protein sequence that corresponds to the protein that has two or more functions was identified and included because a homologous protein might or might not have both functions. Amino acid sequences were identified using the NCBI (National Center for Biotechnology Information) Protein Database [https://www.ncbi.nlm.nih.gov/protein/] or the Universal Protein Resource Protein Knowledgebase (UniProtKB) (15) resources and are included in the FASTA text-based single letter format. Those sequences were used in the BLAST Basic Local Alignment Search Tool (16) (http://blast.ncbi.nlm.nih.gov/Blast.cgi) to search the Protein Data Bank (17) for structures corresponding to the amino acid sequence. In some cases, the structure of the moonlighting protein is not available, so an indication of the presence of a related protein is included with a note about the percent amino acid sequence identity. Gene ontology (GO) terms, which are defined as evidence-based statements relating a specific gene product to a specific ontology term (18) were identified from the UniProtKB (15), and Enzyme Commission (EC) numbers, which are part of a classification scheme for enzymes based on the chemical reactions they catalyze, are included in order to illustrate the different types of protein functions, their cellular locations, and processes in which they are involved. UniProtKB and Protein Data Bank IDs are included for locating more information in these external resources.

New information included in Version 3.0

SCOP (19) and CATH (20) classifications of structural domains were retrieved from the corresponding Protein Data Bank entry when available. The four number classification for CATH includes the Class, Architecture, Topology and Homology information. The Fold information was retrieved from the SCOP classification.

Information about known disordered proteins and disordered regions within proteins was retrieved from the DisProt Database of Protein Disorder (21–23), which contains manually curated information about proteins that have been experimentally shown to contain disordered regions. Predicted transmembrane helices were calculated using TMHMM (24).

Amino acid sequence analysis for regions of low complexity was performed using Protein DisOrder prediction System (PrDOS) and IUPred (25,26). These regions are likely to have higher flexibility and disorder than regions with higher sequence complexity.

Information about type of organism (mammal, plant, etc.) has been added, and, for nonhuman organisms, whether or not the organism is known to cause disease in humans or other animals or plants was also added. Connections of human proteins to disease were retrieved from the Online Mendelian Inheritance in Man online database of human genes and genetic disorders (OMIM) (27).

Database architecture and web interface

The database is based on MySQL (http://www.mysql.com) for data storage, together with PHP 7.1 (http://www.php.net), HTML (HyperText Markup Language), and CSS (Cascading Style Sheets). A Content Management System (CMS): WordPress, was used to update the software. On the homepage, a Search link leads to a page with two search options, a text search and a 'BLAST' sequence similarity search. The Search box enables a text search of all the annotated information in the database, and the search returns a list of protein entries containing that term. The BLAST box enables use of the NCBI-blast-2.6.0+ algorithm (Basic Local Alignment Search Tool) (16) to search the database for proteins that share sequence similarity with a query sequence.

RESULTS

New developments in MoonProt

Additional proteins and updated annotations. The MoonProt Database version 3.0 is now available at www.moonlightingproteins.org. The database has grown by the addition of over one hundred proteins compared to version 2.0. The database now includes information for over 500 moonlighting proteins for which experimental evidence is available confirming the presence of more than one function, and more entries are in progress.

As in versions 1.0 and 2.0, most of the new entries have a catalytic activity as at least one of their functions. The relatively large number of proteins that are enzymes or chaperones inside the cell and have a second function on the cell surface or when secreted to the extracellular media also continues to increase (28). Many of these proteins act as receptors for nutrients or as cell surface adhesins to enable pathogens or probiotic 'good' bacteria to bind to host cells, and others act as secreted signaling molecules to modulate the host immune system. Several dozen proteins have been added that function as transcription or translation factors and have been found to have a second function during mitosis, for example, in binding to the mitotic spindle (12). The number of known or predicted transmembrane proteins was very small in versions 1 and 2, and has increased in version 3.0.

Along with adding more proteins to the database, the annotations for many of the proteins have been updated. As the number of protein structures that are available in the Protein Data Bank grows, more structures of moonlighting proteins have become available, so more PDB identification numbers (PDB IDs) have been added for proteins previously in the MoonProt Database. For some proteins that were included in previous versions of MoonProt, additional functions and references have also been included.

Additional types of information about individual proteins. As the number of moonlighting proteins continues to grow, the examples in MoonProt provide information about protein folds, features or other characteristics that can enable a protein to have two or more biochemical or biophysical functions. To aid in research on how new functions evolve or how to design proteins with additional functions, version 3.0 includes more information about known or predicted structural features of the moonlighting proteins. Protein fold information has been included by adding the SCOP (19) and CATH (20) structural classification for many proteins that have structures in the Protein Data Bank. While some protein folds might enable evolution of a second function, regions of flexibility or disorder can also be important in performing multiple functions, so information about proteins experimentally shown to contain disordered regions was retrieved from the DisProt Database (21–23). Previous versions of the MoonProt Database included very few transmembrane proteins, which raised the question if soluble globular proteins might be more amenable to evolving additional functions. In preparing version 3.0, an emphasis was placed on finding additional examples in the literature of transmembrane moonlighting proteins. Information about the number and location of predicted transmembrane helices was included based on calculations using TMHMM (24), and this information helps identify the twenty proteins in version 3.0 of the database that contain transmembrane helices.

DISCUSSION

The MoonProt Database version 3.0 is now available at www.moonlightingproteins.org The database provides a centralized, organized, searchable, online resource containing information about over 500 moonlighting proteins for which experimental evidence is available for more than one function. Both the number of proteins and the amount of annotation per protein are continuing to grow as new peer-reviewed publications about moonlighting proteins become available and as new protein structures are solved and deposited in the Protein Data Bank.

The wide variety of proteins that moonlight inhibits identification of one general characteristic that could be used to identify all moonlighting proteins and all their functions. However, the numbers of proteins within some subsets of moonlighting proteins are increasing. Two examples are proteins that have one function in the cell cytoplasm and another function when displayed on the cell surface (28) and cytoplasmic proteins with a second function in the nucleus, for example as a transcription factor that regulates gene expression (9). Having collections of the sequences and structures for proteins in these subsets might enable identification of sequence or structural motifs or algorithms that can be used to identify additional moonlighting proteins within these subsets. The database also serves as a resource for labs interested in developing computational methods for predicting protein functions based on sequence, structure, protein–protein interactions or other characteristics.

Information about X-ray crystal structures available in the Protein Data Bank has been included and updated, but a crystal structure might represent only one structure out of several for a moonlighting protein. Many moonlighting proteins undergo changes in structure when they change functions, including changes in conformation, tertiary fold, and/or multimeric assembly. For example, some proteins have one function while part of the ribosome or another multiprotein complex and a different function as a monomer or homomultimer. A small number of proteins, called metamorphic and morpheein proteins, can undergo changes in the tertiary folds of protein subunits, and, in some cases, the change in tertiary fold is correlated with a change in function (29,30). Some information about the different structures correlating with different functions has been included in the MoonProt Database, but adding the information is still in progress, and the structure performing each function is not always known. Nevertheless, having a database of proteins that change function could be a starting point for identifying additional proteins that undergo significant changes in structure. Changes in the structures of moonlighting proteins that correlate with changes in function could be valuable leads for the development of bioinspired nanoswitches and nanomachines.

We also note that updating the MoonProt Database has been one method by which the Jeffery lab has continued its research activities during the coronavirus pandemic of 2020. While other research opportunities for undergraduates and high school students were not available during social distancing, we developed opportunities for three UIC undergraduate students, two Summer Research Opportunity Program undergraduate students in Puerto Rico, two additional undergraduate volunteers from within the USA or internationally (Ukraine) and four high school students to be involved in lab activities and learn about protein sequence and structural analysis.

DATA AVAILABILITY AND LICENSE

The MoonProt Database is freely available via a user-friendly graphical user interface (GUI) at the web address www.moonlightingproteins.org. The interface enables text search for a protein name, species, or a UniProtKB or PDB identifier and a BLAST search using an amino acid sequence in the one letter code. The user can also browse a list of all the proteins in the database. The database is 'read and search only' by the public, but additional information about the known moonlighting proteins and suggestions of other proteins that might also be moonlighting are welcome and can be sent to the curators for possible inclusion in the database.

FUNDING

UIC Cancer Center Award (to C.J.J.). Funding for open access charge was from Jeffery lab funds, a UIC Cancer Center award (to C.J.J.), and the Research Open Access Publishing (ROAAP) Fund of the University of Illinois at Chicago.
Conflict of interest statement. None declared.

REFERENCES

1. Jeffery,C.J. (1999) Moonlighting proteins. *Trends Biochem. Sci.*, **24**, 8–11.
2. Jeffery,C.J. (2003) Moonlighting proteins: old proteins learning new tricks. *Trends Genet.*, **19**, 415–417.
3. Jeffery,C.J. (2017) Moonlighting proteins–nature's Swiss army knives. *Sci. Prog.*, **100**, 363–373.
4. Wistow,G. and Piatigorsky,J. (1987) Recruitment of enzymes as lens structural proteins. *Science*, **236**, 1554–1556.
5. Guo,M. and Schimmel,P. (2013) Essential nontranslational functions of tRNA synthetases. *Nat. Chem. Biol.*, **9**, 145–153.
6. 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.
7. Hernandez,S., Ferragut,G., Amela,I., Perez-Pons,J., Pinol,J., Mozo-Villarias,A., Cedano,J. and Querol,E. (2014) MultitaskProtDB: a database of multitasking proteins. *Nucleic Acids Res.*, **42**, D517–D520.
8. Gancedo,C. and Flores,C.-L. (2008) Moonlighting proteins in yeasts. *Microbiol. Mol. Biol. Rev.*, **72**, 197–210.
9. Commichau,F.M. and Stülke,J. (2008) Trigger enzymes: bifunctional proteins active in metabolism and in controlling gene expression. *Mol. Microbiol.*, **67**, 692–702.
10. Piatigorsky,J. (2007) Harvard University Press, Cambridge, MA.
11. Wool,I.G. (1996) Extraribosomal functions of ribosomal proteins. *Trends Biochem. Sci.*, **21**, 164–165.
12. Somma,M.P., Andreyeva,E.N., Pavlova,G.A., Pellacani,C., Bucciarelli,E., Popova,J.V., Bonaccorsi,S., Pindyurin,A.V. and Gatti,M. (2020) Moonlighting in mitosis: analysis of the mitotic functions of transcription and splicing factors. *Cells*, **9**, 1554.
13. 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.
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. The UniProt Consortium (2017) UniProt: the universal protein knowledgebase. *Nucleic Acids Res.*, **45**, D158–D169.
16. McGinnis,S. and Madden,T.L. (2004) BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res.*, **32**, W20–W25.
17. Rose,P.W., Prlić,A., Altunkaya,A., Bi,C., Bradley,A.R., Christie,C.H., Costanzo,L.D., Duarte,J.M., Dutta,S. and Feng,Z. (2016) The RCSB protein data bank: integrative view of protein, gene and 3D structural information. *Nucleic Acids Res.*, **45**, D271–D281.
18. GO Consortium (2017) Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res.*, **45**, D331–D338.
19. Murzin,A.G., Brenner,S.E., Hubbard,T. and Chothia,C. (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.*, **247**, 536–540.
20. Pearl,F.M., Bennett,C., Bray,J.E., Harrison,A.P., Martin,N., Shepherd,A., Sillitoe,I., Thornton,J. and Orengo,C.A. (2003) The CATH database: an extended protein family resource for structural and functional genomics. *Nucleic Acids Res.*, **31**, 452–455.
21. Uversky,V.N. (2002) Natively unfolded proteins: a point where biology waits for physics. *Protein Sci.*, **11**, 739–756.
22. Tompa,P. (2002) Intrinsically unstructured proteins. *Trends Biochem. Sci.*, **27**, 527–533.
23. 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.
24. Sonnhammer,E.L., Von Heijne,G. and Krogh,A. (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, **6**, 175–182.
25. Ishida,T. and Kinoshita,K. (2007) PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res.*, **35**, W460–W464.
26. Dosztányi,Z., Csizmok,V., Tompa,P. and Simon,I. (2005) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics*, **21**, 3433–3434.
27. Amberger,J.S., Bocchini,C.A., Schiettecatte,F., Scott,A.F. and Hamosh,A. (2015) OMIM. org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Res.*, **43**, D789–D798.
28. Amblee,V. and Jeffery,C.J. (2015) Physical features of intracellular proteins that moonlight on the cell surface. *PLoS One*, **10**, e0130575.
29. Jaffe,E.K. (2005) Morpheeins–a new structural paradigm for allosteric regulation. *Trends Biochem. Sci.*, **30**, 490–497.
30. Dishman,A.F. and Volkman,B.F. (2018) Unfolding the mysteries of protein metamorphosis. *ACS Chem. Biol.*, **13**, 1438–1446.