

## **Membranome 3.0: database of single-pass membrane proteins with AlphaFold models**

**Andrei L. Lomize<sup>1#</sup>, Kevin A. Schnitzer<sup>2</sup>, Spencer C. Todd<sup>2</sup>, Stanislav Cherepanov<sup>3</sup>,  
Carlos Outeiral<sup>4</sup>, Charlotte M. Deane<sup>4</sup>, Irina D. Pogozeva<sup>1</sup>**

<sup>1</sup>Department of Medicinal Chemistry, College of Pharmacy, University of Michigan, 428 Church St., Ann Arbor, MI, 48109-1065 USA;

<sup>2</sup>Department of Electrical Engineering and Computer Science, College of Engineering, University of Michigan, 1221 Beal Ave, Ann Arbor, MI 48109-2102;

<sup>3</sup>Department of Biophysics, University of Michigan, 930 N. University Avenue, Ann Arbor, MI, 48109-1065, USA;

<sup>4</sup>Department of Statistics, University of Oxford, 24-29 St Giles', OX1 3PB, United Kingdom.

#Corresponding Author: Andrei L. Lomize.

Email: [almz@umich.edu](mailto:almz@umich.edu)

Mailing address: 428 Church St, Ann Arbor, MI 48109-1065, USA

Mobile Phone: +1(734) 925-2370

Running title: Membranome 3.0 database for single-pass membrane proteins

Manuscript: 22 pages total, including 1 Table, and 2 Figures

Supplementary materials: 51 pages total, including 3 Figures and 6 Tables.

**This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as doi: [10.1002/pro.4318](https://doi.org/10.1002/pro.4318)**

This article is protected by copyright. All rights reserved.

## ABSTRACT

The Membranome database provides comprehensive structural information on single-pass (*i.e.*, bitopic) membrane proteins from six evolutionarily distant organisms, including protein-protein interactions, complexes, mutations, experimental structures, and models of transmembrane  $\alpha$ -helical dimers. We present a new version of this database, Membranome 3.0, which was significantly updated by revising the set of 5,759 bitopic proteins and incorporating models generated by AlphaFold 2 in the database. The AlphaFold models were parsed into structural domains located at the different membrane sides, modified to exclude low-confidence unstructured terminal regions and signal sequences, validated through comparison with available experimental structures, and positioned with respect to membrane boundaries. Membranome 3.0 was re-developed to facilitate visualization and comparative analysis of multiple 3D structures of proteins that belong to a specified family, complex, biological pathway, or membrane type. New tools for advanced search and analysis of proteins, their interactions, complexes, and mutations were included. The database is freely accessible at <https://membranome.org>.

**KEYWORDS:** web tool, full-length protein model, visualization, network analysis.

## STATEMENT

The Membranome database of single-pass membrane proteins from six evolutionarily distant organisms was re-developed and improved by refining protein sets and collecting newly available data. 5,690 of AlphaFold 2-generated models of full-length single-pass proteins were modified, reassembled, validated, and positioned in membranes. New tools were added for advanced search, visualization of proteins, and analysis of their interactions, complexes, biological pathways, and mutations.

**Abbreviations:** 3D, three-dimensional; AF2; AlphaFold 2; pLDDT, predicted local distance difference test; RMSD, root-mean-square deviation; TM, transmembrane.

## 1. INTRODUCTION

Single-pass (*i.e.* bitopic) transmembrane (TM) proteins cover up to 50% of mammalian membrane proteins and play key roles in many vital processes, including signal transduction, cell adhesion and communications, immune response, energy conversion, molecular biogenesis and transport, malignant transformation, viral entry, and other cellular processes.<sup>1</sup> Structural characterization of bitopic proteins is a critical step in understanding the molecular mechanisms of their function and regulation, and of the impact of disease-causing mutations. Unfortunately, determining the three-dimensional (3D) structure of full-length bitopic membrane protein remains a challenge for both experimental and computational methods, because these proteins are often composed of multiple structural domains located at different sides of a membrane, and because they exist in many conformational and oligomeric states.

To facilitate structural analysis of bitopic membrane proteins, we previously created the Membranome database,<sup>2;3</sup> a web resource dedicated to this functionally relevant protein class. The Membranome database compiles available structural data for all bitopic proteins of six organisms (*Homo sapiens*, *Arabidopsis thaliana*, *Dictiostelium discoideum*, *Saccharomyces cerevisiae*, *Escherichia coli*, *Methanococcus jannaschii*) representing the six kingdoms of life. The database provides protein classification, amino acid sequences, membrane topologies, intracellular localizations, domain organization (from UniProtKB<sup>4</sup> and Pfam<sup>5</sup>), experimentally supported protein-protein interactions, protein complexes, experimental and computed structures of bitopic protein domains, their static pictures produced by PyMOL,<sup>6</sup> and interactive visualization using 3D viewers. It also includes links to experimental 3D structures of hundreds of water-soluble domains and dozens of TM domains of bitopic proteins from PDB,<sup>7</sup> PDBsum,<sup>8</sup> and OPM.<sup>9</sup> The original version of the database<sup>2</sup> contained models of single TM  $\alpha$ -helices predicted and generated by

FMAP<sup>10</sup> for all included bitopic proteins. In 2018, the version 2.0 of the database<sup>3</sup> was expanded by including more than two thousand models of energetically stable (average helix association energy was about -6 kcal/mol) TM  $\alpha$ -helical homodimers of bitopic proteins from all six species that were predicted and generated by TMDOCK<sup>11</sup> and positioned in membranes by PPM.<sup>9</sup> The Membranome 2.0 database provided a unique point of access to the structural information of bitopic proteins available in 2018.

In the past year, advances in protein structure prediction have powered spectacular progress in structural modeling. At the fourteenth Critical Assessment of protein Structure Prediction (CASP), the AlphaFold 2 (AF2) system<sup>12</sup> outperformed all other computational methods, producing models rivaling experimental structures.<sup>12-15</sup> Although AlphaFold 2 has a limited applicability in modeling protein dynamics, multiple conformational states, or effects of mutations,<sup>16; 17</sup> the approach is thought to achieve a significant progress in predicting the structure of a single protein chain.<sup>12</sup> Public availability of the AF2 source code and the recent release of the AlphaFold DataBase<sup>18</sup> with over half a million protein models, including the full proteomes of 16 model organisms and 32 pathogens, are starting to have a transformative impact on structural biology.<sup>13, 15-16</sup> Despite the varying quality of AF2-generated models, they have been added to high-quality authoritative resources for protein sequences, structures, and functional information, such as UniProt,<sup>4</sup> and PDBsum.<sup>19</sup> The availability of high-accuracy predictions for a significant portion of many organisms' proteomes is a novel source of information into bitopic proteins.

In this work, we present an upgraded version of the Membranome database that includes AF2 models of bitopic proteins, along with additional information about protein interactions, complexes, and pathways, and new functionalities for protein analysis. Protein models from the AlphaFold DataBase have been modified and validated by positioning them with respect to

membrane boundaries. To ensure reliability of new models, they have also been compared to available experimental structures of corresponding proteins. These improvements result in a modern database intended to further the study and understanding of bitopic proteins.

## **2. NEW CONTENT OF MEMBRANOME 3.0**

### **2.1 New information for bitopic proteins**

A vast sample of new structural data on bitopic proteins was incorporated to the Membranome 3.0 database from various bioinformatics resources, including 9,078 PDB structures of protein domains, 5,664 direct protein-protein interactions, 1,791 validated complexes, and 1051 mutations in TM domains. Structural and functional information about bitopic proteins and their complexes was filtered to keep the most reliable information supported by publications. For example, only “high-quality” and “level 2” interactions were taken from HINT<sup>20</sup> and APID,<sup>21</sup> respectively. To get additional protein-protein interactions, protein pages were linked to IntAct,<sup>22</sup> BioGrid,<sup>23</sup> and STRING.<sup>24</sup> Bitopic proteins complexes were collected from Reactome,<sup>25</sup> Complex Portal,<sup>26</sup> CORUM,<sup>27</sup> and PDB.<sup>7</sup> Mutations in TM  $\alpha$ -helices of bitopic proteins were taken from MutHTP.<sup>28</sup> Biological pathways were compiled from KEGG,<sup>29</sup> Reactome,<sup>25</sup> BioCyc,<sup>30</sup> WikiPathways,<sup>31</sup> and HMDB<sup>32</sup> databases and classified into classes and sub-classes. Models of individual TM  $\alpha$ -helices were recalculated using the latest version of our FMAP program.<sup>10</sup>

#### **2.1. AlphaFold models of full-length monomers positioned in membranes**

Despite the success of AlphaFold 2 in accurately predicting single-chain protein structures, there is a wide variation in model quality between (and within) proteins. A quality assessment score, the predicted local distance difference test (pLDDT) is natively produced by the AF2 system.<sup>13</sup> High-confidence predictions (pLDDT>70, threshold resulting from benchmark on a test set<sup>13</sup>) cover roughly 62% of the human proteome and 92% of sequences in *E. coli*.<sup>33</sup> Despite progress, there

are many proteins and portions of proteins whose AF2 structural models may not be accurate. Models of single-pass transmembrane proteins are especially problematic, as they include multiple extracellular (EC) and intracellular (IC) domains that can change their relative positions due to the flexibility of connecting loops. One salient problem is that, as AF2 does not consider spatial separation of domains located at opposite membrane sides, it often generates models of large single-pass TM proteins with intertwined TM, EC, and IC domains (**Figure 1A**).

The AlphaFold DataBase<sup>18</sup> contains structural models of almost 99% of the bitopic proteins included in the Membranome database. For each protein with amino acid sequences longer than 2,700 residues, a single model containing TM  $\alpha$ -helix was taken from a set of 1,400-residue models progressively shifted by 200 residues that are provided by the AlphaFold Database. Models of 110 bitopic proteins missing in the AlphaFold Database were calculated by AF2 using the standard protocol, code, parameters and databases available through the CC BY-NC 4.0 license.<sup>13</sup>

To fix models with incorrectly intertwined domains located at opposite membrane sides, we developed an in-house program D-linker that parses EC, IC, and TM parts of a protein, optimizes their spatial positions in membrane using PPM 3.0 and reassembles them at both membrane sides (**Figure 1B**). The D-linker also removes N-terminal signal sequences of bitopic proteins and unstructured regions of low confidence (pLDDT < 70) from N- and C-termini, while keeping all confidently predicted regions (pLDDT  $\geq$  70) and connecting loops. A significant number of bitopic protein models from the AlphaFold Database have distorted or partially unfolded TM  $\alpha$ -helices. To ensure the correct positioning in membrane of these models, their TM segments were automatically superposed by D-linker with TM  $\alpha$ -helices modeled and oriented in membranes by FMAP.<sup>10</sup> Several proteins, such as phospholipid scramblases, BCL-2-like apoptosis regulators, and chloride intracellular channel proteins from the thioredoxin superfamily, are known to adopt

water-soluble and TM forms. Structural models predicted by AF2 for water-soluble forms of these proteins were included to Membranome. Ultimately, 5,690 AF2 models were modified, positioned in membrane and included in the Membranome 3.0 database.

To validate these models, we compared them with subunits of 9,093 experimental 3D structures derived from 947 available bitopic proteins representing all considered organisms, except *D. discoideum*. We excluded 105 bitopic proteins with experimental structures containing less than 45 residues. Experimental structures were superimposed with predicted models by TM-align tool using sequence independent and dependent alignment settings.<sup>34</sup> The average root-mean-square deviation (RMSD) value was less than 2 Å for sets of overlapping residues for single protein domains and semi-rigid multi-domain proteins (**Tables 1, S1-S5**), but it increased to 2.5-6 Å for multi-domain proteins connected by flexible loops or while comparing models with fused protein structures with hybrid amino acid sequences (**Table S6**). In average, about a half of each AF2 model was covered by experimental structures. These results underscore the high reliability of AF2 method for folded protein regions predicted with high confidence.

The incorporation of AF2-generated models has several advantages for improving database content and advancing protein analysis. First, these structural models provide atomic representation of the complex domain architecture of full-length bitopic proteins (**Figure 2**). Second, comparisons of collections of bitopic membrane proteins in Membranome and the AlphaFold DataBase help refining the sets of all bitopic proteins from six organisms by removing mispredicted polytopic and monotopic proteins from Membranome and by revising and recalculating bitopic proteins that are missing in the AlphaFold DataBase. After corrections, the final bitopic protein sets for *H. sapiens*, *A. thaliana*, *D. discoideum*, *S. cerevisiae*, *E. coli*, *M. jannaschii* contain 2,384, 2,106, 605, 383, 205, and 77 entries, respectively (**Table 1**). Third,



comparison of multiple static pictures of protein models during database search or browsing helps to find errors in membrane topology and to revise the structure-based classification of proteins. The visual comparison of full-length models of proteins from the same functional category (family, complex, biological pathway, or membrane type) for evolutionarily distant organisms allows exposing their structural features that could be essential for specific protein functions and protein-protein interactions.

The database provides downloadable coordinates only for AF2 models of protein monomers, even though many bitopic proteins are known to form functional dimers or higher order oligomers. Experimental structures of such dimers and multimers available from PDB and OPM databases are linked to Membranome pages for individual bitopic proteins. Besides, Membranome 3.0 includes more than 2,000 downloadable coordinate files of TM  $\alpha$ -helical homodimers that were modeled by TMDOCK.<sup>11</sup> The full-length protein dimers could be generated by the AlphaFold-Multimer (AFM) program.<sup>35</sup> However, the quality and the reliability of modeling protein dimers by AFM remains rather limited. For a set of 4,433 protein complexes, the AFM successfully predicted dimer interface ( $\text{DockQ} \geq 0.23$ ) in 67% cases and produced high accuracy models ( $\text{DockQ} \geq 0.8$ ) in 23% cases.<sup>35</sup> Modeling of dimers of full-length multidomain bitopic proteins remains a challenging problem. Therefore, we did not include any low-reliability predictions of full-length protein dimers in the Membranome 3.0 database.

### **3. NEW FUNCTIONALITIES OF MEMBRANOME 3.0**

The new web tool, 1TMnet, was created for analysis of structural and functional (pathways) interaction networks of bitopic proteins (<https://membranome.org/1tmnet>). For a set of user-selected proteins from the database, 1TMnet generates interactive tables and graphs that show structural and functional relations between these proteins based on experimentally proven

interactions, known complexes, and associations in biological pathways. For example, while selecting *Homo sapiens* and “EGFR” in search boxes (**Figure S1**), the user gets a Table of 29 proteins associated with the human EGFR protein. By selecting all 29 proteins from the Table, the user gets graphs and Tables for 10 direct (**Figure S2**) and 42 indirect (**Figure S3**) interactions between selected bitopic proteins, together with Tables presenting 15 complexes with selected proteins and 115 pathways related to these proteins (only complexes and pathways with at least two proteins from the set are included). Direct interactions and protein complexes obtained using 1TMnet facilitates analysis of bitopic protein partners participating in formation of functional protein heteromers. Collecting structural information for proteins from evolutionarily distant organisms in the databases and providing the tools for advanced search, protein network analysis, and interactive 3D visualization facilitates comparative structural and evolutionary analysis of bitopic proteins. To facilitate the interactive visualization of protein structures, iCn3D<sup>36</sup> and GLMol<sup>37</sup> web-based 3D viewers were included, in addition to JMol.<sup>38</sup>

#### 4. DATABASE IMPLEMENTATION

The Membranome database was re-developed using the Ruby on Rails server-side web application framework and the PostgreSQL database management system for the back-end. The front-end application was developed using ReactJS. The database is hosted on the Heroku Cloud platform with assets (static protein images and PDB files) stored on the Google Cloud platform. Firebase hosting was used for the front-end of the website. The Membranome website also provides access to the FMAP,<sup>10</sup> TMDOCK,<sup>11</sup> and 1TMnet web tools. 1TMnet is written in Ruby language on the back-end and uses the ActiveRecord library to generate SQL to query the membrane database for results based on the inputs. The front-end uses the TypeScript programming language and the Cytoscape.js library.<sup>39</sup>

## **SUPPLEMENTARY MATERIAL**

Tables 1-6. Verification of AF2-generated models of bitopic proteins from five organisms (*H. sapiens*, *A. thaliana*, *S. cerevisiae*, *E. coli*, *M. jannaschii*) by comparison with experimental structures from PDB. Figures 1-2. Direct (1) and indirect (2) interactions for EGFR\_HUMAN produced by 1TMnet.

## **ACKNOWLEDGMENTS**

This work was supported by the Division of Biological Infrastructure of the National Science Foundation [Award # **1855425** to A.L. and I.P.].

## **CONFLICT OF INTEREST**

The authors declare no competing financial interests.

## **ORCID**

Andrei L. Lomize <https://orcid.org/0000-0002-3044-7597>

Kevin A. Schnitzer <https://orcid.org/0000-0002-6490-9323>

Spencer C. Todd <https://orcid.org/0000-0002-0120-7674>

Stanislav Cherepanov <https://orcid.org/0000-0001-9459-1849>

Carlos Outeiral <https://orcid.org/0000-0003-1408-5554>

Charlotte M Deane <https://orcid.org/0000-0003-1388-2252>

Irina D. Pogozheva <https://orcid.org/0000-0002-3024-9574>

## REFERENCES

1. Pogozeva ID, Lomize AL (2018) Evolution and adaptation of single-pass transmembrane proteins. *Biochim Biophys Acta.* 1860:364-377. PMID: 29129605 {Medline}
2. Lomize AL, Lomize MA, Krolicki SR, Pogozeva ID (2017) Membranome: A database for proteome-wide analysis of single-pass membrane proteins. *Nucleic Acids Res.* 45:D250-D255. PMID: 27510400 {Medline}
3. Lomize AL, Hage JM, Pogozeva ID (2018) Membranome 2.0: Database for proteome-wide profiling of bitopic proteins and their dimers. *Bioinformatics.* 34:1061-1062. PMID: 29126305 {Medline}
4. Consortium The UniProt (2017) UniProt: The universal protein knowledgebase. *Nucleic Acids Res.* 45:D158-D169. PMID: 27899622 {Medline}
5. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, Finn RD (2019) The Pfam protein families database in 2019. *Nucleic Acids Res.* 47:D427-D432. PMID: 30357350 {Medline}
6. DeLano WL (2003) The PyMOL molecular graphics system. Delano Scientific LLC, San Carlos, CA, USA. <http://www.pymol.org/>
7. Burley SK, Berman HM, Kleywegt GJ, Markley JL, Nakamura H, Velankar S (2017) Protein Data Bank (PDB): The single global macromolecular structure archive. *Methods Mol Biol.* 1607:627-641. PMID: 28573592 {Medline}

8. Laskowski RA, Jablonska J, Pravda L, Varekova RS, Thornton JM (2018) PDBsum: Structural summaries of PDB entries. *Protein Sci.* 27:129-134. PMID: 28875543  
{Medline}
9. Lomize MA, Pogozheva ID, Joo H, Mosberg HI, Lomize AL (2012) OPM database and PPM web server: Resources for positioning of proteins in membranes. *Nucleic Acids Res.* 40:D370-376. PMID: 21890895 {Medline}
10. Lomize AL, Schnitzer KA, Todd SC, Pogozheva ID (2021) Thermodynamics-based molecular modeling of  $\alpha$ -helices in membranes and micelles. *J Chem Inf Model.* 61:2884-2896.
11. Lomize AL, Pogozheva ID (2017) TMDOCK: An energy-based method for modeling alpha-helical dimers in membranes. *J Mol Biol.* 429:390-398. PMID: 27622289  
{Medline}
12. Kryshtafovych A, Schwede T, Topf M, Fidelis K, Moult J (2021) Critical assessment of methods of protein structure prediction (CASP)-round XIV. *Proteins.* 89:1607-1617. PMID: 34533838 {Medline}
13. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Židek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D (2021) Highly accurate protein structure prediction with AlphaFold. *Nature.* 596:583-589. PMID: 34265844  
{Medline}

14. Tunyasuvunakool K, Adler J, Wu Z, Green T, Zielinski M, Židek A, Bridgland A, Cowie A, Meyer C, Laydon A, Velankar S, Kleywegt GJ, Bateman A, Evans R, Pritzel A, Figurnov M, Ronneberger O, Bates R, Kohl SAA, Potapenko A, Ballard AJ, Romera-Paredes B, Nikolov S, Jain R, Clancy E, Reiman D, Petersen S, Senior AW, Kavukcuoglu K, Birney E, Kohli P, Jumper J, Hassabis D (2021) Highly accurate protein structure prediction for the human proteome. *Nature*. 596:590-596.
15. Akdel M, Pires DEV, Porta Pardo E, Jänes J, Zalevsky AO, Mészáros B, Bryant P, Good LL, Laskowski RA, Pozzati G, Shenoy A, Zhu W, Kundrotas P, Ruiz Serra V, Rodrigues CHM, Dunham AS, Burke D, Borkakoti N, Velankar S, Frost A, Lindorff-Larsen K, Valencia A, Ovchinnikov S, Durairaj J, Ascher DB, Thornton JM, Davey NE, Stein A, Elofsson A, Croll TI, Beltrao P (2021) A structural biology community assessment of AlphaFold 2 applications. *bioRxiv:2021.2009.2026.461876*.
16. Outeiral C, Nissley DA, Deane CM (2022) Current structure predictors are not learning the physics of protein folding. *Bioinformatics*. PMID: 35099504 {Medline}
17. Pak MA, Markhieva KA, Novikova MS, Petrov DS, Vorobyev IS, Maksimova ES, Kondrashov FA, Ivankov DN (2021) Using AlphaFold to predict the impact of single mutations on protein stability and function. *bioRxiv:2021.2009.2019.460937*.
18. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A, Židek A, Green T, Tunyasuvunakool K, Petersen S, Jumper J, Clancy E, Green R, Vora A, Lutfi M, Figurnov M, Cowie A, Hobbs N, Kohli P, Kleywegt G, Birney E, Hassabis D, Velankar S (2021) AlphaFold protein structure database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res*. PMID: 34791371 {Medline}

19. Laskowski RA, Thornton JM (2021) PDBsum extras: SARS-Cov-2 and AlphaFold models. *Protein Sci.* PMID: 34779073 {Medline}
20. Das J, Yu H (2012) HINT: High-quality protein interactomes and their applications in understanding human disease. *BMC Syst Biol.* 6:92. PMID: 22846459 {Medline}
21. Alonso-López D, Campos-Laborie FJ, Gutiérrez MA, Lambourne L, Calderwood MA, Vidal M, De Las Rivas J (2019) APID database: Redefining protein-protein interaction experimental evidences and binary interactomes. *Database (Oxford)* 2019. PMID: 30715274 {Medline}
22. Orchard S, Ammari M, Aranda B, Breuza L, Briganti L, Broackes-Carter F, Campbell NH, Chavali G, Chen C, del-Toro N, Duesbury M, Dumousseau M, Galeota E, Hinz U, Iannuccelli M, Jagannathan S, Jimenez R, Khadake J, Lagreid A, Licata L, Lovering RC, Meldal B, Melidoni AN, Milagros M, Peluso D, Perfetto L, Porras P, Raghunath A, Ricard-Blum S, Roechert B, Stutz A, Tognolli M, van Roey K, Cesareni G, Hermjakob H (2014) The MIntAct project--intact as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.* 42:D358-D363. PMID: 24234451 {Medline}
23. Oughtred R, Stark C, Breitkreutz BJ, Rust J, Boucher L, Chang C, Kolas N, O'Donnell L, Leung G, McAdam R, Zhang F, Dolma S, Willems A, Coulombe-Huntington J, Chatr-Aryamontri A, Dolinski K, Tyers M (2019) The bioGRID interaction database: 2019 update. *Nucleic Acids Res.* 47:D529-D541. PMID: 30476227 {Medline}
24. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, Mering CV (2019) STRING V11: Protein-protein association networks with increased coverage, supporting functional discovery in

- genome-wide experimental datasets. *Nucleic Acids Res.* 47:D607-D613. PMID: 30476243 {Medline}
25. Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fabregat A, Sidiropoulos K, Cook J, Gillespie M, Haw R, Loney F, May B, Milacic M, Rothfels K, Sevilla C, Shamovsky V, Shorser S, Varusai T, Weiser J, Wu G, Stein L, Hermjakob H, D'Eustachio P (2020) The Reactome pathway knowledgebase. *Nucleic Acids Res.* 48:D498-D503. PMID: 31691815 {Medline}
26. Meldal BHM, Bye AJH, Gajdoš L, Hammerová Z, Horácková A, Melicher F, Perfetto L, Pokorný D, Lopez MR, Türková A, Wong ED, Xie Z, Casanova EB, Del-Toro N, Koch M, Porras P, Hermjakob H, Orchard S (2019) Complex portal 2018: Extended content and enhanced visualization tools for macromolecular complexes. *Nucleic Acids Res.* 47:D550-D558. PMID: 30357405 {Medline}
27. Giurgiu M, Reinhard J, Brauner B, Dunger-Kaltenbach I, Fobo G, Frishman G, Montrone C, Ruepp A (2019) CORUM: The comprehensive resource of mammalian protein complexes-2019. *Nucleic Acids Res.* 47:D559-D563. PMID: 30357367 {Medline}
28. Kulandaisamy A, Binny Priya S, Sakthivel R, Tarnovskaya S, Bizin I, Hönigsmid P, Frishman D, Gromiha MM (2018) MutHTP: Mutations in human transmembrane proteins. *Bioinformatics.* 34:2325-2326. PMID: 29401218 {Medline}
29. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K (2017) KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45:D353-D361. PMID: 27899662 {Medline}
30. Paley S, Karp PD (2021) The BioCyc metabolic network explorer. *BMC Bioinformatics.* 22:208. PMID: 33882841 {Medline}



31. Martens M, Ammar A, Riutta A, Waagmeester A, Slenter DN, Hanspers K, R AM, Digles D, Lopes EN, Ehrhart F, Dupuis LJ, Winckers LA, Coort SL, Willighagen EL, Evelo CT, Pico AR, Kutmon M (2021) WikiPathways: Connecting communities. *Nucleic Acids Res.* 49:D613-D621. PMID: 33211851 {Medline}
32. Wishart DS, Guo A, Oler E, Wang F, Anjum A, Peters H, Dizon R, Sayeeda Z, Tian S, Lee BL, Berjanskii M, Mah R, Yamamoto M, Jovel J, Torres-Calzada C, Hiebert-Giesbrecht M, Lui VW, Varshavi D, Varshavi D, Allen D, Arndt D, Khetarpal N, Sivakumaran A, Harford K, Sanford S, Yee K, Cao X, Budinski Z, Liigand J, Zhang L, Zheng J, Mandal R, Karu N, Dambrova M, Schiöth HB, Greiner R, Gautam V (2022) HMDB 5.0: The Human Metabolome Database for 2022. *Nucleic Acids Res.* 50:D622-D631. PMID: 34986597 {Medline}
33. Thornton JM, Laskowski RA, Borkakoti N (2021) AlphaFold heralds a data-driven revolution in biology and medicine. *Nat Med.* 27:1666-1669. PMID: 34642488 {Medline}
34. Zhang Y, Skolnick J (2005) TM-align: A protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* 33:2302-2309. PMID: 15849316 {Medline}
35. Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, Židek A, Bates R, Blackwell S, Yim J, Ronneberger O, Bodenstern S, Zielinski M, Bridgland A, Potapenko A, Cowie A, Tunyasuvunakool K, Jain R, Clancy E, Kohli P, Jumper J, Hassabis D (2021) Protein complex prediction with AlphaFold-Multimer. [bioRxiv:2021.10.20.463034](https://doi.org/10.1101/2021.10.20.463034).
36. Wang J, Youkharibache P, Zhang D, Lanczycki CJ, Geer RC, Madej T, Phan L, Ward M, Lu S, Marchler GH, Wang Y, Bryant SH, Geer LY, Marchler-Bauer A (2020) iCn3D, a

web-based 3D viewer for sharing 1D/2D/3D representations of biomolecular structures. *Bioinformatics*. 36:131-135. PMID: 31218344 {Medline}

37. Virag I, Stoicu-Tivadar L, Crişan-Vida M (2016) Gesture interaction browser-based 3D molecular viewer. *Stud Health Technol Inform*. 226:17-20. PMID: 27350455 {Medline}
38. Herráez A (2006) Biomolecules in the computer: Jmol to the rescue. *Biochem Mol Biol Educ*. 34:255-261. PMID: 21638687 {Medline}
39. Franz M, Lopes CT, Huck G, Dong Y, Sumer O, Bader GD (2016) Cytoscape.js: A graph theory library for visualisation and analysis. *Bioinformatics*. 32:309-311. PMID: 26415722 {Medline}

**Table 1.** Verification of AF2-generated models by comparison with experimental structures

Bitopic proteins in Membranome 3.0			Superposition of AF2 models with PDB structures <sup>3</sup>				
Species	N <sub>prot</sub> <sup>1</sup>	N <sub>AF2</sub> <sup>2</sup>	N <sub>AF2</sub> <sup>2</sup>	R <sub>over-AF2</sub> <sup>4</sup>	R <sub>over-AF2</sub> , <sup>5</sup> %	RMSD, <sup>6</sup> Å	Identity, <sup>7</sup> %
<i>H. sapiens</i>	2384	2368	788	271±184	53±22	1.5±0.9	97±10
<i>A. thaliana</i>	2106	2069	52	215±106	51±24	1.0±1.0	98±4
<i>S. cerevisiae</i>	383	374	69	104±21	26±17	1.1±0.5	100±1
<i>D. discoideum</i>	605	598	0	0	0	0	0
<i>E. coli</i>	205	204	34	476±358	83±10	2.8±1.8	99±1
<i>M. jannaschii</i>	77	77	4	131±29	77±9	0.6±0.5	99±1

<sup>1</sup> N<sub>prot</sub>, number of bitopic proteins

<sup>2</sup> N<sub>AF2</sub>, number of AF2 models included in Membranome database

<sup>3</sup> Average values with standard deviations. One PDB entry with the largest number of overlapped residues was selected for each protein.

<sup>4</sup> R<sub>over-AF2</sub>, number of overlapping residues between AF2 model and PDB structure

<sup>5</sup> Percentage of overlapping residues between AF2 model and PDB structure

<sup>6</sup> RMSD for C $\alpha$ -atoms

<sup>7</sup> sequence identity between overlapping residues in AF2 models and PDB structures

## FIGURE LEGENDS

**Figure 1.** Modeling of full-length bitopic membrane proteins. (A) AF2-generated model of receptor tyrosine phosphatase F (PTPRF). (B) AF2 model of PTPRF parsed into the extracellular domain (ECD), TM domain (TMD), and the intracellular domain (ICD) by D-linker and positioned in membranes by PPM 3.0. Rainbow colored cartoon representations were produced by PyMOL. PPM-calculated membrane boundaries are shown by gray dots.

**Figure 2.** Membranome 3.0 page for integrin alpha-10 (UniProt ID: ITA10\_HUMAN).

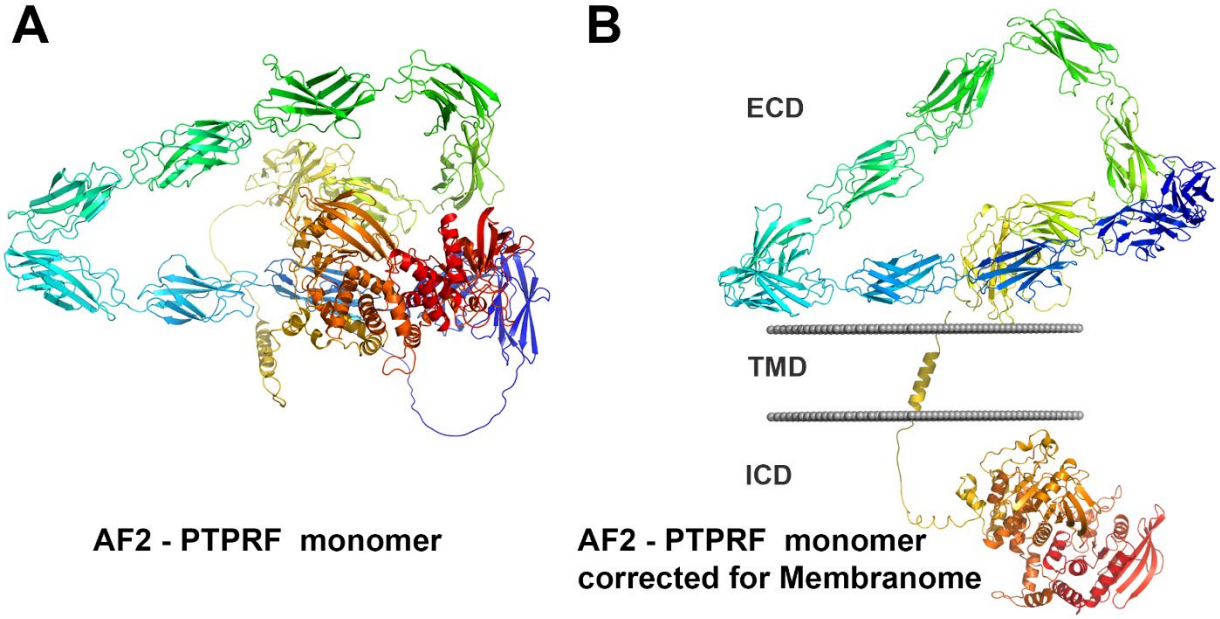


Figure 1.

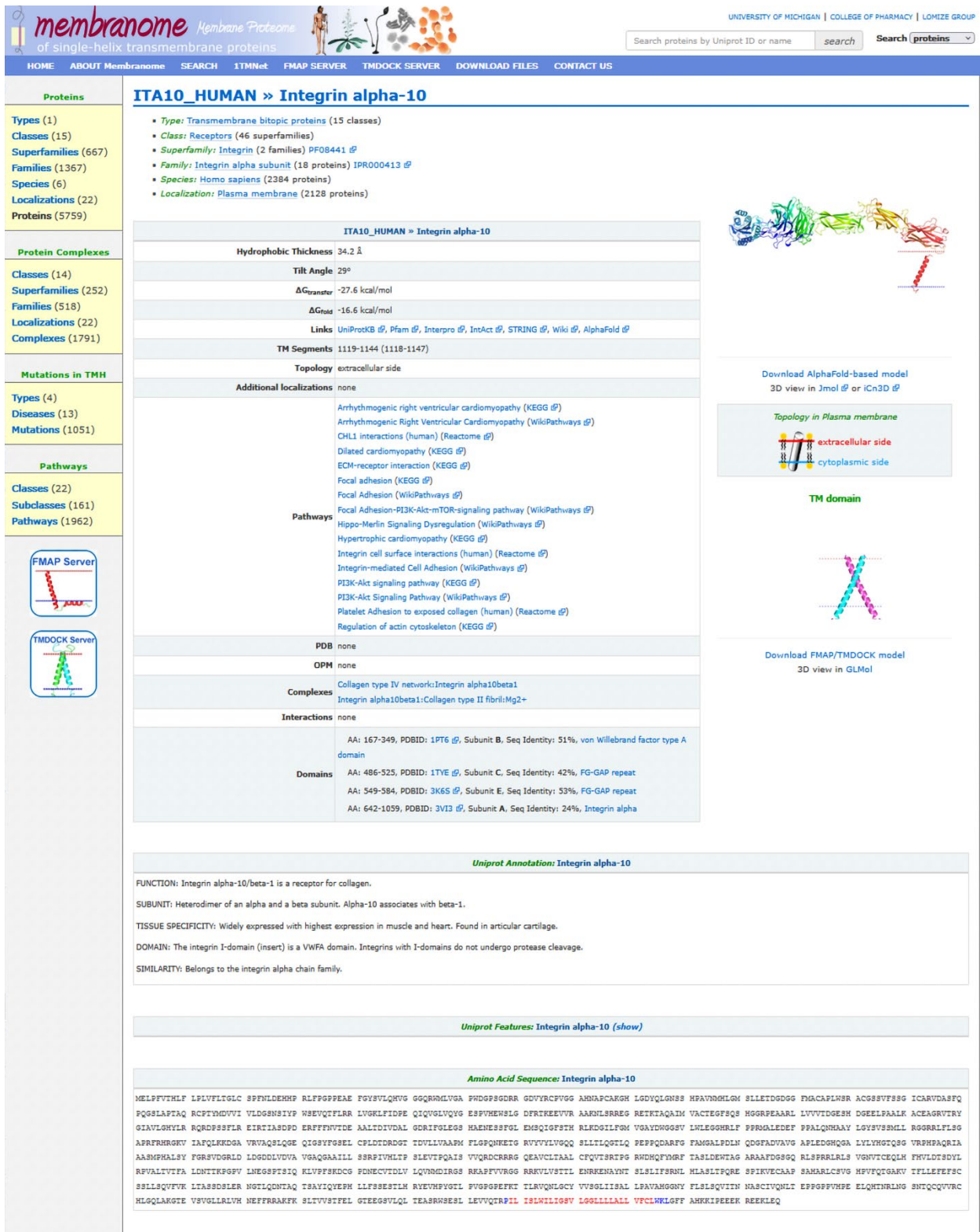
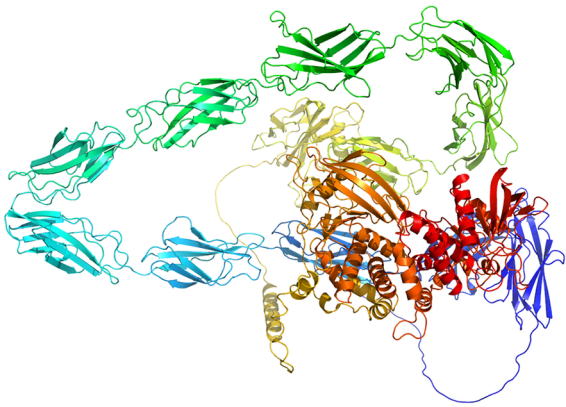


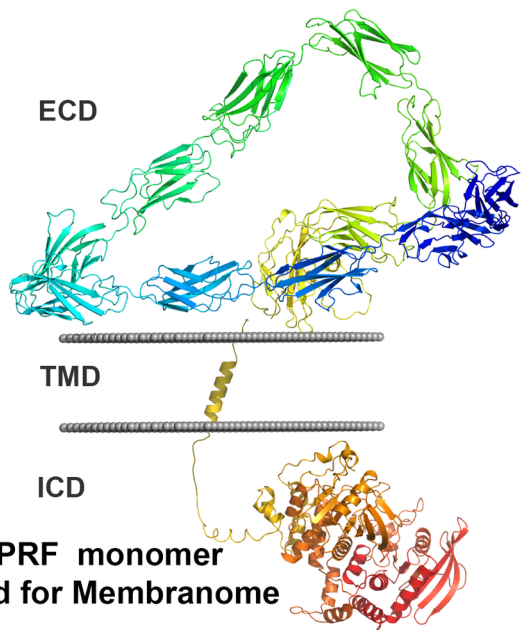
Figure 2.

**A**




**AF2 - PTPRF monomer**

**B**




**AF2 - PTPRF monomer  
corrected for Membrane**

PRO\_4318\_Fig1.tif



Membrane Proteome  
of single-helix transmembrane proteins



UNIVERSITY OF MICHIGAN | COLLEGE OF PHARMACY | LOMIZE GROUP

Search proteins by Uniprot ID or name  Search **proteins** ▾

---

**Proteins**

Types (1)  
Classes (15)  
Superfamilies (667)  
Families (1367)  
Species (6)  
Localizations (22)  
Proteins (5759)

**Protein Complexes**



Classes (14)  
Superfamilies (252)  
Families (518)  
Localizations (22)  
Complexes (1791)

**Mutations in TMH**

Types (4)  
Diseases (13)  
Mutations (1051)

**Pathways**

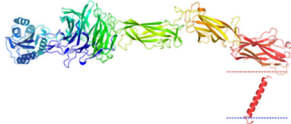
Classes (22)  
Subclasses (161)  
Pathways (1962)

### ITA10\_HUMAN » Integrin alpha-10


- **Type:** Transmembrane bitopic proteins (15 classes)
- **Class:** Receptors (46 superfamilies)
- **Superfamily:** Integrin (2 families) PF08441 [↗](#)
- **Family:** Integrin alpha subunit (18 proteins) IPR000413 [↗](#)
- **Species:** Homo sapiens (2384 proteins)
- **Localization:** Plasma membrane (2128 proteins)

ITA10_HUMAN » Integrin alpha-10	
<b>Hydrophobic Thickness</b>	34.2 Å
<b>Tilt Angle</b>	29°
<b>ΔG<sub>transfer</sub></b>	-27.6 kcal/mol
<b>ΔG<sub>fold</sub></b>	-16.6 kcal/mol
<b>Links</b>	UniProtKB <a href="#">↗</a> , Pfam <a href="#">↗</a> , Interpro <a href="#">↗</a> , Interact <a href="#">↗</a> , STRING <a href="#">↗</a> , Wiki <a href="#">↗</a> , AlphaFold <a href="#">↗</a>
<b>TM Segments</b>	1119-1144 (1118-1147)
<b>Topology</b>	extracellular side
<b>Additional localizations</b>	none
<b>Pathways</b>	Arrhythmic right ventricular cardiomyopathy (KEGG <a href="#">↗</a> ) Arrhythmic Right Ventricular Cardiomyopathy (WikiPathways <a href="#">↗</a> ) CHL1 interactions (human) (Reactome <a href="#">↗</a> ) Dilated cardiomyopathy (KEGG <a href="#">↗</a> ) ECM-receptor interaction (KEGG <a href="#">↗</a> ) Focal adhesion (KEGG <a href="#">↗</a> ) Focal Adhesion (WikiPathways <a href="#">↗</a> ) Focal Adhesion-PI3K-Akt-mTOR-signaling pathway (WikiPathways <a href="#">↗</a> ) Hippo-Merlin Signaling Dysregulation (WikiPathways <a href="#">↗</a> ) Hypertrophic cardiomyopathy (KEGG <a href="#">↗</a> ) Integrin cell surface interactions (human) (Reactome <a href="#">↗</a> ) Integrin-mediated Cell Adhesion (WikiPathways <a href="#">↗</a> ) PI3K-Akt signaling pathway (KEGG <a href="#">↗</a> ) PI3K-Akt Signaling Pathway (WikiPathways <a href="#">↗</a> ) Platelet Adhesion to exposed collagen (human) (Reactome <a href="#">↗</a> ) Regulation of actin cytoskeleton (KEGG <a href="#">↗</a> )
<b>PDB</b>	none
<b>OPM</b>	none
<b>Complexes</b>	Collagen type IV network: Integrin alpha10beta1 Integrin alpha10beta1: Collagen type II fibril: Mg2+
<b>Interactions</b>	none
<b>Domains</b>	AA: 167-349, PDBID: 1PT6 <a href="#">↗</a> , Subunit B, Seq Identity: 51%, von Willebrand factor type A domain AA: 486-525, PDBID: 1TYE <a href="#">↗</a> , Subunit C, Seq Identity: 42%, FG-GAP repeat AA: 549-584, PDBID: 3K6S <a href="#">↗</a> , Subunit E, Seq Identity: 53%, FG-GAP repeat AA: 642-1059, PDBID: 3V13 <a href="#">↗</a> , Subunit A, Seq Identity: 24%, Integrin alpha




Download AlphaFold-based model  
3D view in Jmol [↗](#) or iCn3D [↗](#)

Topology in Plasma membrane



TM domain



Download FMAP/TMDOCK model  
3D view in GLMol [↗](#)

Uniprot Annotation: Integrin alpha-10

FUNCTION: Integrin alpha-10/beta-1 is a receptor for collagen.  
SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-10 associates with beta-1.  
TISSUE SPECIFICITY: Widely expressed with highest expression in muscle and heart. Found in articular cartilage.  
DOMAIN: The integrin I-domain (insert) is a VWFA domain. Integrins with I-domains do not undergo protease cleavage.  
SIMILARITY: Belongs to the integrin alpha chain family.

Uniprot Features: Integrin alpha-10 [\(show\)](#)

Amino Acid Sequence: Integrin alpha-10

```

MELSPFTHLF LPLVFLTGLC SPFNLDKHPH RLPFGPPEAE FQVSVLQHWG GQQRMLVGA EWDGPGSDRR GDUVRCFVGG AHNAPCANGH LGDVQLGNSS HEATNMHLM SILETDGDSG FMCAPLMSR ACSSSVFSSG ICARDVDSFQ
FQGSIAFTAQ RCFYMDVVI VLDGSSNIYP NBEVQFLLR LVGKLFIDPE QIQVGLVQVG ESPVHEWSLG DFRTRREEVVR AAKNLSRRRG RETKTAQAIM VACTEGFSQS HGRRFEAARL LUVVVDGESH DGEELFAALK ACEAGRVTYR
GIAVLGHYLR RQRDPSSFLR EIRTIASDFD ERFFNFVDE AALTDIVDAL GDRIFGLEGS HAENESSFGL EMSQIGFSTH RLKDGILFGM VGAYDWGGSV LWLGGHRLF PFRMALEDEF FFALQNHAY LGYSVSSMLL RGRRLFLSG
APFRHRGKV IAFQLKDDGA VRVAQSLQGE QIGSYFQSEL CPDIDTDRDGT TDVLLVAAM FLGPNKRETS RVVYLVVGGQ SILLTQSTLQ EEPFDARFG FMGALFDLN QDGFADVAVG APLDGHQGA LYLVTGQSG VRHPAQRIA
AAMPHALSY FGRSVDGRLD LDGDDLVDA VQAQGAAILL SSRPIVHLIF SLEVIPQAIS VVQRDCRRGG QEAUCLTAA CFQVTSRIFG RWDHQFMRF TASLDEWTAG ARAAFDSSGQ RLSFRRLRLS VGVVTCQLM FHLVDTSDVL
RPAVLTIFA LDNITKPGV LNEGSPFSIQ KLVFSPKDCG PDNECVIDLV LQVNDMIRGS RKAPFVVRGG RRKVLVSTL ENRRENAVNT SLSLIFSRL HLASLTPQRE SPIRVECAAP SAHARLCSVG HFFVQIGARKV TFLFEFESC
SLLSQVFKV LIASDSDSLR NQTLQNTAQ TSAYIQHEPH LLPSSBSTLM RVEVHPXGL FVSGPPEFKI IRLVQMLGCV VVSGLIISAL LPAVHGGNY FLSLQVITH NASCIQMLI EPPGFFVHPE ELQHTNRLNG SHTQQCVRC
HLGQLARGTE VSVGLRLVH NEFFRAKFK SLIVVSTFEL GTEEGSVLQL TEASRWSEL LEVVQTRPIL ISLWILIGSV LGGLLILALL VFCLWLGFF AHKRIPEER REEKLEQ
                    
```

PRO\_4318\_Fig2.tif