



Database tool

PlantMP: a database for moonlighting plant proteins

Bo Su¹, Zhuang Qian¹, Tianshu Li¹, Yuwei Zhou¹ and Aloysius Wong^{2,*} 

¹Department of Computer Science, College of Science and Technology, Wenzhou-Kean University, 88 Daxue Road, Ouhai, Wenzhou, Zhejiang Province 325060, China and ²Department of Biology, College of Science and Technology, Wenzhou-Kean University, 88 Daxue Road, Ouhai, Wenzhou, Zhejiang Province 325060, China

*Corresponding author: Tel.: +86-138-5771-3894; Email: alwong@kean.edu

Citation details: Su,B., Qian,Z., Li,T. *et al.* PlantMP: a database for moonlighting plant proteins. *Database* (2019) Vol. 2019: article ID baz050; doi:10.1093/database/baz050

Received 14 January 2019; Revised 13 March 2019; Accepted 28 March 2019

Abstract

Moonlighting proteins are single polypeptide chains capable of executing two or more distinct biochemical and/or biological functions. Here, we describe the development of PlantMP, which is a manually curated online-based database of plant proteins that are known to ‘moonlight’. The database contains searchable UniProt IDs and names, canonical and moonlighting functions, gene ontology numbers, plant species as well as links to the PubMed indexed articles. Proteins homologous to experimentally confirmed moonlighting proteins from the model plant *Arabidopsis thaliana* are provided as a separate list of ‘likely moonlighters’. Additionally, we also provide a list of predicted *Arabidopsis* moonlighting proteins reported in the literature. Currently, PlantMP contains 110 plant moonlighting proteins, 10 ‘likely moonlighters’ and 27 ‘predicted moonlighters’. Organizing plant moonlighting proteins in one platform enables researchers to conveniently harvest plant-specific raw and processed data such as the molecular functions, biological roles and structural features essential for hypothesis formulation in basic research and for biotechnological innovations.

Database URL: <https://www.plantmp.com>

Introduction

Moonlighting proteins are single polypeptide chains capable of executing two or more distinct biochemical and/or biological functions. Unlike multi-protein complexes, they are not products of gene fusions, RNA splice variants or proteolytic fragments. They also exclude pleiotropic proteins and members of the same protein family that perform

different functions (1). Moonlighting proteins are found in organisms across the tree of life where they assume important secondary functions such as regulating transcription, catalysis and tuning signaling networks (2–8). These alternative functions confer cost- and efficiency-associated benefits by making fewer proteins from a more compact genome and increase control of regulatory and signaling

pathways (9). However, moonlighting functions of proteins are often discovered by serendipity, thus the creation of moonlighting protein databases such as MoonProt (10) and MultitaskProtDB (11) to pool together this class of proteins especially given that they are increasingly common and have been implicated in human diseases (9).

The requirement for spatial and temporal micro-regulations of cellular biochemicals such as signaling molecules, ligands and cofactors becomes more significant considering the relatively crowded space of the plant cell, which is occupied by large central vacuoles (12). Many canonical domains in other organisms, including animals and bacteria that exist as stand-alone proteins, have been incorporated into multi-functional complex plant proteins often as smaller functional centers that are embedded within larger primary domains (13). Indeed, there have been well-characterized examples in receptor protein complexes such as BRI1 from *Arabidopsis thaliana* where externally located hormone recognition domain is connected via a transmembrane region to a cytoplasmic canonical kinase domain that also accommodates a smaller moonlighting guanylate cyclase (GC) center (14). The same is also observed in *Arabidopsis* potassium transporter and channel (KUP5 and GORK) proteins, where moonlighting adenylate cyclase (AC) center and ABA binding site were found to occupy cytosolic sites of these multi-pass membrane proteins (15, 16). This complexity also makes it incredibly difficult to examine precise cellular regulatory functions of individual components and their broader biological roles.

Current databases contain moonlighting proteins from human and other organisms, but plant moonlighters are under-represented. The fundamental differences in domain architecture and cellular organization of plant proteins would require a more focused avenue to aid discovery of similar proteins, data interpretation and annotation of moonlighting proteins in model plants, crops and other economically valuable plants. A comprehensive catalog of plant moonlighting proteins can therefore better serve the plant science community in sequencing, proteomics and structural works, thus the creation of PlantMP.

Materials and methods

Selection of proteins for inclusion in PlantMP database

Experimentally validated plant moonlighting proteins were collected from PubMed using ‘plant moonlighting proteins’ and ‘dual/double function/role plant protein’ as the search terms. Retrieved articles were manually screened and information of plant moonlighting proteins were extracted from the articles for inclusion in the

PlantMP database. Homologs of experimentally validated moonlighting proteins from *A. thaliana* were given as a separate list of ‘likely moonlighters’. Additionally, a list of ‘predicted moonlighters’ that consist of *Arabidopsis* proteins, which are not homologs of known moonlighting proteins but have been reported to be candidate GCs or ACs, were included in the database. In both ‘likely moonlighters’ and ‘predicted moonlighters’, their canonical functions were first determined on UniProt (17), and if no primary functions are known, they are excluded from the database. Additionally, we also checked proteins in the list of ‘likely moonlighters’ for the presence of known domains or motifs that could hint moonlighting functions (Figures 1 and 2).

Protein information included in PlantMP database

Moonlighting plant proteins in the PlantMP database contain searchable UniProt IDs and names, canonical and moonlighting functions, gene ontology (GO) numbers, plant species and links associated to the PubMed articles. Included in the canonical and moonlighting functions are their molecular function (MF) and biological process (BP) extracted from UniProt. For *Arabidopsis* proteins, they are also linked to their sources in The *Arabidopsis* Information Resource (TAIR) database (www.arabidopsis.org). The database will be curated and updated continuously to include additional members or new functions as experimental evidence surface in the future.

Database design and interface

The website features a prominent top-aligned banner that houses the name of the database, a text search box, a 3D model of a representative moonlighting protein (BRI1, UniProt ID: O22476) (14) as well as ‘Introduction’ and ‘Browse Database’ section headers. The text search box allows experienced users to quickly query and retrieve proteins of interest from three search options: protein name, UniProt ID and plant species. The search box contains drop down suggestions to aid users identify proteins of interest. Following the search box is a 3D model of BRI1 docked with ATP and GTP and accompanied by clearly labeled primary and moonlighting catalytic activities to illustrate the function of the database. An introduction section that contains a brief description of moonlighting proteins and the database is provided to aid first time users navigate and extract information from the database. Following the introduction is the ‘Browse Database’ section where users can select from one of four lists of proteins: (i) All categories, (ii) Confirmed moonlighters, (iii) Likely moonlighters and



Figure 1. PlantMP database and interface. Information extracted from various databases and articles was manually curated and formatted offline on Excel before importing into MySQL database in the following format: protein name, gene uniprot_id, canonical_function_bp, canonical_function_mf, moonlighting_function_bp, moonlighting_function_mf, go_canonical_bp, go_canonical_mf, go_moonlighting_bp, go_moonlighting_mf, plant_species, pubmed_id, type. 'Type' indicates if the moonlighting function of the protein is predicted, likely or confirmed based on the literature. Database website structure was designed and developed using the front-end and back-end separation, which has the benefit of low-degree coupling. The website front-end will query the database when user invoked either the search function by UniProt ID, protein name or plant species or by selecting one of the four categories provided. At the back-end, data is provided using Restful Web API with JSON format. For example, all moonlighting proteins from *A. thaliana* can be retrieved by querying the URL <https://www.plantmp.com/api/species?species=Arabidopsis> in the web browser. For website development, an online web template provided by Colorlib (<http://www.cssmoban.com/cssthemes/8065.shtml>) was adopted and modulated with feature module for database query. Front-end only requires the setting up of the user query input as a parameter to specify API URL and then displayed in appropriate table HTML format. Users have the option to export all or selected results into Excel format.

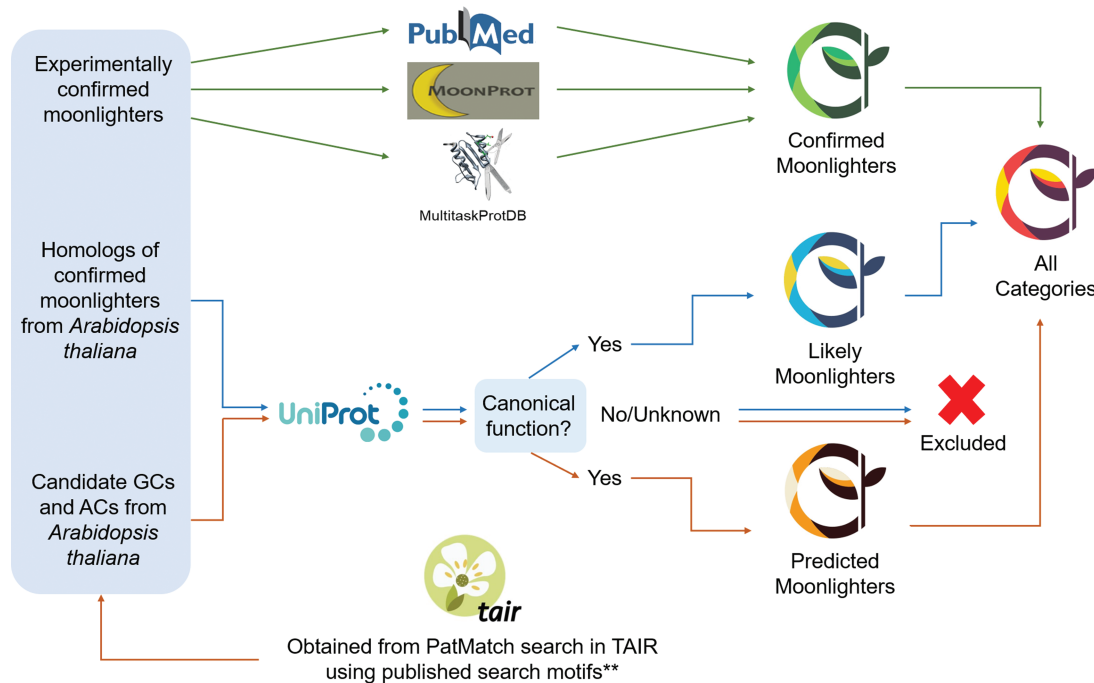


Figure 2. Selection of moonlighting plant proteins for inclusion in PlantMP database. Proteins with experimentally validated moonlighting functions retrieved from PubMed and other databases such as MoonProt (<http://moonlightingproteins.org>) (10) and MultiTaskProtDB (<http://wallace.uab.es/multitaskII>) (11) were included in the list of 'confirmed moonlighters' (indicated as green arrows). Homologs of confirmed moonlighters from *A. thaliana* were included in the list of 'likely moonlighters' (indicated as blue arrows), while reported candidate GC and AC (which are typically moonlighting in multi-domain plant proteins) from *A. thaliana* (13) were included in the list of 'predicted moonlighters' (indicated as orange arrows). Both 'likely' and 'predicted' moonlighters were only included if they have known primary functions and harbor domains or motifs that could hint moonlighting functions based on UniProt annotations (17). The candidate GCs and ACs were retrieved from a PatMatch search on TAIR (<https://www.arabidopsis.org/cgi-bin/patmatch/nph-patmatch.pl>) (28) using search motifs **reported by (13, 20).

(iv) Predicted moonlighters. The website allows users to select desired proteins and export their results as Excel. Currently, PlantMP contains 110 plant moonlighting proteins, 10 'likely moonlighters' and 27 'predicted moonlighters' (Figure 3).

Results and discussion

Plant MP is available at <https://www.plantmp.com> without registration or license. While moonlighting proteins from

human and other animal systems are well-characterized, those from plants are, however, ambiguous and under-represented in existing databases. PlantMP is to-date the most comprehensive catalog of plant moonlighting proteins featuring proteins from *A. thaliana*, crops and other non-model plants. Moonlighting proteins not found in existing databases such as those from *Pharbitis nil* and *Hippeastrum hybridum* whose catalytic activities are linked to phytochrome controlled photoperiodic flower induction and pathogen infection respectively (18, 19), are included here.

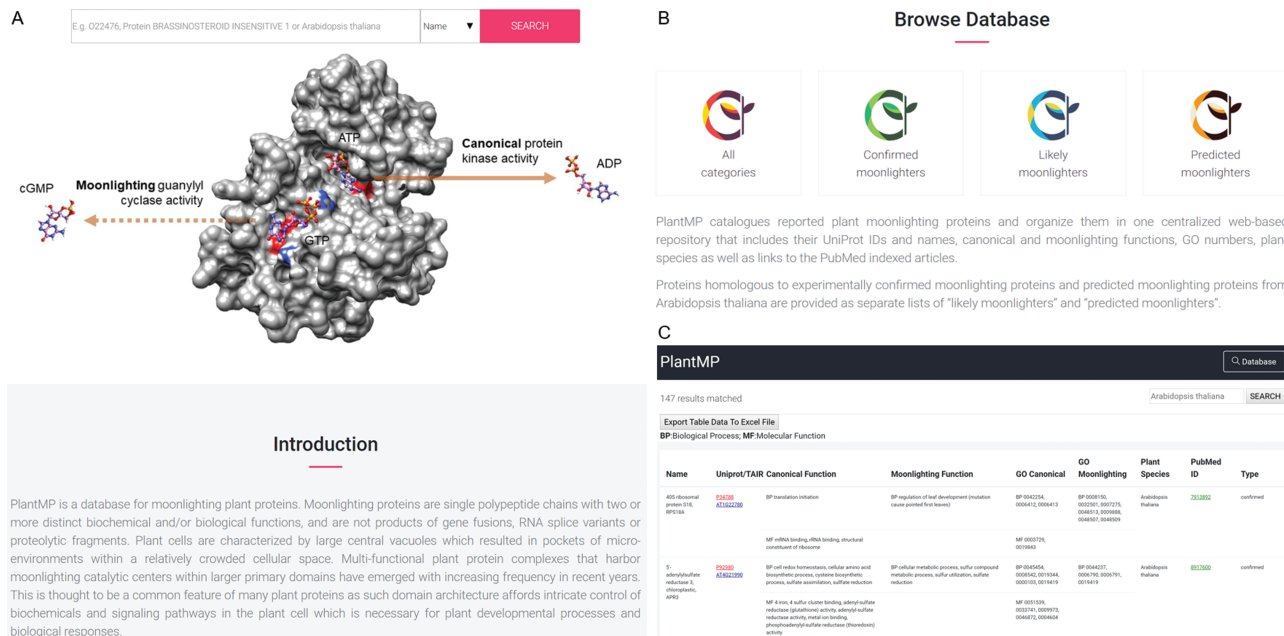


Figure 3. Layout of the PlantMP web server. Screenshot of the home page (A) beginning with a text search box that allows experienced users to quickly query and retrieve proteins of interest from three search options (protein name, UniProt ID and plant species) and then followed by a 3D model of a representative moonlighting protein (BRI1, UniProt ID: O22476) (14) docked with ATP and GTP and accompanied by clearly labeled primary and moonlighting catalytic activities to illustrate the function of the database. An introduction section that contains a brief description of moonlighting proteins and the database is provided to aid first time users navigate and extract information from the database. Following the Introduction is the ‘Browse Database’ section (B) where users can select from one of four lists of proteins: (1) All categories, (2) Confirmed moonlighters, (3) Likely moonlighters and (4) Predicted moonlighters. In the result page (C), moonlighting proteins in the PlantMP database contain UniProt IDs and names, canonical and moonlighting functions, GO numbers, plant species and links associated to the PubMed articles. Included in the canonical and moonlighting functions are their MF and BP extracted from UniProt (17). The website allows users to select desired proteins and export their results as Excel.

Also featured are moonlighting proteins from crops such as tomato, carrot and rice. PlantMP is an initial effort to catalog plant moonlighters and will require a community-driven effort to make this database as comprehensive as possible. As such, we welcome users to reach us through the email provided at the bottom of the webpage to suggest plant proteins that warrant inclusion in our database, and our team will validate the request and update the database accordingly. Additionally, we are also exploring other repositories to identify plant moonlighters and conduct periodic updates on the database as new experimental evidence surface. A more focused plant-specific database allows less prominent moonlighting proteins or those from non-model plants to be highlighted especially since they may have potential for horticulture innovations and, consequently, added commercial value.

PlantMP is also inclusive as it accommodates ambiguous proteins or those whose functions/roles are unclear in separate categories of ‘likely’ and ‘predicted’ moonlighting proteins. Proteins in the ‘predicted’ list were identified using previously published search motifs: [KS][YF][GCS][VIL][VILFG][DVIL][VILADG][EPVIL][DVIL][TVIL][WST][PD RG][KEG][KR]_{x{2,3}}[DHSE] for GC (13) and [R]X_{5,20}

[RKS][YFW][DE][VIL]X_{4}[VIL]X_{4}[KR]X_{1,3}[DE] for AC (20). As one of the first search terms created for the discovery of plant GCs/ACs, these manually curated motifs were made stringent to reduce false positives, hence the inclusion of their identified candidates in our database. Other more relaxed versions of the search motifs have since been reported and applied (12, 21–23). Recently created prediction tools, GCPred (<http://gcpred.com>) (24) and ACPred (<http://gcpred.com/acpred>) (25), have considered more relaxed versions of the search motifs in their algorithms and used scores assigned to individual hits that are based on the physicochemical parameters of experimentally validated proteins instead, as measure of confidence. New hits obtained from predictions by currently available web servers are not included here because they have not been published nor validated experimentally or computationally. However, if users find new hits that are not experimentally proven but received good computational support that warrant their inclusion as ‘predicted moonlighters’, they can submit their requests to us for inclusion in our database. Given the increasing number of reported moonlighting plant proteins (12, 26, 27), an exclusive database is therefore necessary in order

to reveal the features and nature of such proteins many of which are unique to the plant cell. Organizing plant moonlighting proteins in one platform enables researchers to conveniently harvest plant-specific data such as the MFs, biological roles and structural features essential for hypothesis formulation in basic research and for biotechnological innovations.

Authors contributions

A.W. conceived and supervised this research. B.S. coordinated the project. B.S., Z.Q., T.L. and Y.Z. collected, compiled and analyzed the data and developed the web interface of the database. A.W. wrote the manuscript. All authors read and approved the final version of the manuscript.

Acknowledgements

The authors would like to thank Ms Yi Mei Ng (King Abdullah University of Science and Technology, Saudi Arabia) and Ms Cynthia Wong (University of Toronto, Canada) for testing the server and providing valuable feedback.

Funding

National Natural Science Foundation of China (grant no. 31850410470); Zhejiang Provincial Natural Science Foundation of China (grant no. LQ19C130001).

Conflict of interest. None declared.

References

- Jeffery, C.J. (1999) Moonlighting proteins. *Trends Biochem. Sci.*, **24**, 8–11.
- Jeffery, C.J. (2003) Moonlighting proteins: old proteins learning new tricks. *Trends Genet.*, **19**, 415–417.
- Irving, H., Cahill, D.M. and Gehring, C. (2018) Moonlighting proteins and their role in the control of signaling microenvironments, as exemplified by cGMP and phytosulfokine receptor 1 (PSKR1). *Front. Plant Sci.*, **9**, 415.
- Swiezawska, B., Duszyn, M., Jaworski, K. *et al.* (2018) Downstream targets of cyclic nucleotides in plants. *Front. Plant Sci.*, **9**, 1428.
- Kwezi, L., Wheeler, J.I., Marondedze, C. *et al.* (2018) Intramolecular crosstalk between catalytic activities of receptor kinases. *Plant Signal. Behav.*, **13**, e1430544. doi: [10.1080/15592324.2018.1430544](https://doi.org/10.1080/15592324.2018.1430544)
- Freihat, L., Muleya, V., Manallack, D.T. *et al.* (2014) Comparison of moonlighting guanylate cyclases: roles in signal direction? *Biochem. Soc. Trans.*, **42**, 1773–1779.
- Muleya, V., Wheeler, J.I., Ruzvidzo, O. *et al.* (2014) Calcium is the switch in the moonlighting dual function of the ligand-activated receptor kinase phytosulfokine receptor 1. *Cell Commun. Signal.*, **12**, 60.
- Irving, H.R., Kwezi, L., Wheeler, J. *et al.* (2012) Moonlighting kinases with guanylate cyclase activity can tune regulatory signal networks. *Plant Signal. Behav.*, **7**, 201–204.
- Franco-Serrano, L., Hernández, S., Calvo, A. *et al.* (2017) MultitaskProtDB-II: an update of a database of multitasking/moonlighting proteins. *Nucleic Acids Res.*, **46**, D645–D648.
- Mani, M., Chen, C., Amblee, V. *et al.* (2015) MoonProt: a database for proteins that are known to moonlight. *Nucleic Acids Res.*, **43**, D277–D282.
- Hernández, S., Ferragut, G., Amela, I. *et al.* (2014) MultitaskProtDB: a database of multitasking proteins. *Nucleic Acids Res.*, **42**, D517–D520.
- Wong, A., Tian, X., Gehring, C. *et al.* (2018) Discovery of novel functional centers with rationally designed amino acid motifs. *Comput. Struct. Biotechnol. J.*, **16**, 70–76.
- Wong, A. and Gehring, C. (2013) The *Arabidopsis thaliana* proteome harbors undiscovered multi-domain molecules with functional guanylyl cyclase catalytic centers. *Cell Commun. Signal.*, **11**, 48.
- Wheeler, J.I., Wong, A., Marondedze, C. *et al.* (2017) The brassinosteroid receptor BRI1 can generate cGMP enabling cGMP-dependent downstream signaling. *Plant J.*, **91**, 590–600.
- Ooi, A., Lemtiri-Chlieh, F., Wong, A. *et al.* (2017) Direct modulation of the guard cell outward-rectifying potassium channel (GORK) by abscisic acid. *Mol. Plant*, **10**, 1469–1472.
- Al-Younis, I., Wong, A., Lemtiri-Chlieh, F. *et al.* (2018) The *Arabidopsis thaliana* K⁺-uptake Permease 5 (AtKUP5) contains a functional cytosolic adenylate cyclase essential for K⁺ transport. *Front. Plant Sci.*, **9**, 1645.
- The UniProt Consortium (2017) UniProt: the universal protein knowledgebase. *Nucleic Acids Res.*, **45**, D158–D169.
- Szmidt-Jaworska, A., Jaworski, K., Paweleket, A. *et al.* (2009) Molecular cloning and characterization of a guanylyl cyclase, PnGC-1, involved in light signaling in *Pharbitis nil*. *J. Plant Growth Regul.*, **28**, 367–380.
- Swiezawska, B., Jaworski, K., Duszyn, M. *et al.* (2017) The *Hippeastrum hybridum* PepR1 gene (HpPepR1) encodes a functional guanylyl cyclase and is involved in early response to fungal infection. *J. Plant Physiol.*, **216**, 100–107.
- Al-Younis, I., Wong, A. and Gehring, C. (2015) The *Arabidopsis thaliana* K⁺-uptake permease 7 (AtKUP7) contains a functional cytosolic adenylate cyclase catalytic Centre. *FEBS Lett.*, **589**, 3848–3852.
- Gehring, C. (2010) Adenyl cyclases and cAMP in plant signaling-past and present. *Cell Commun. Signal.*, **8**, 15.
- Marondedze, C., Wong, A., Thomas, L. *et al.* (2017) Cyclic nucleotide monophosphates in plants and plant signaling. *Handb. Exp. Pharmacol.*, **238**, 87–103 doi: [10.1007/164_2015_35](https://doi.org/10.1007/164_2015_35).
- Wong, A. and Gehring, C. (2013) Computational identification of candidate nucleotide cyclases in higher plants. *Methods Mol. Biol.*, **1016**, 195–205.
- Xu, N., Zhang, C., Lim, L.L. *et al.* (2018) Bioinformatic analysis of nucleotide cyclase functional centers and development of ACPreD webserver. In: *Proceedings of the ACM-BCB'18: 9th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*. Association for Computing Machinery (ACM), New York, NY, USA. doi: [10.1145/3233547.3233549](https://doi.org/10.1145/3233547.3233549).

25. Xu,N., Fu,D., Li,S. *et al.* (2018) GCPred: a web tool for guanylyl cyclase functional center prediction from amino acid sequence. *Bioinformatics*, **34**, 2134–2135.
26. Wong,A., Gehring,C. and Irving,H.R. (2015) Conserved functional motifs and homology modeling to predict hidden moonlighting functional sites. *Front. Bioeng. Biotechnol.*, **3**, 82.
27. Gehring,C. and Turek,I.S. (2017) Cyclic nucleotide monophosphates and their cyclases in plant signaling. *Front. Plant Sci.*, **8**, 1704.
28. Yan,T., Yoo,D., Berardini,T.Z. *et al.* (2005) PatMatch: a program for finding patterns in peptide and nucleotide sequences. *Nucleic Acids Res.*, **33**, W262–W266.