

## Nomination for membership of the ISB Executive Committee

September 5<sup>th</sup>, 2016

Jasmine Young, Ph.D.

Biocuration Team Lead, RCSB Protein Data Bank

Global Project Lead, Worldwide Protein Data Bank

Associate Research Professor, Center for Integrative Proteomics Research

Rutgers, The State University of New Jersey

[jasmine.young@rcsb.org](mailto:jasmine.young@rcsb.org)

ORCID: 0000-0001-8896-6878

ResearchGate: [https://www.researchgate.net/profile/Jasmine\\_Young](https://www.researchgate.net/profile/Jasmine_Young)

### Letter of Intent

I have worked in biocuration since 2003, and believe that the overall promotion of the biocuration field and the importance of setting community-approved data standards are critical to enabling data science in biology. Since the ISB was established, my involvement in Conferences has included poster presentations (Semi Finalist in 2012), oral presentations in 2010 and 2015, conference organizer in 2012, and session chair (Protein structure, complexes, interactions) in 2012.

A major focus of my biocuration career has been directed towards the development of standard representations for the Protein Data Bank (PDB) archive. The PDB is the single global archive for experimentally-determined, atomic-level structures of biological macromolecules. The PDB archive is managed by the Worldwide Protein Data Bank organization (wwPDB; <http://wwpdb.org>), which currently includes three founding regional data centers, located in the US (RCSB Protein Data Bank or RCSB PDB; <http://rcsb.org>), Japan (Protein Data Bank Japan or PDBj; <http://pdj.org>), and Europe (Protein Data Bank in Europe or PDBe; <http://pdbe.org>), plus a global NMR specialist data repository BioMagResBank, composed of deposition sites in the US (BMRB; <http://www.bmrwisc.edu>) and Japan (PDBj-BMRB; <http://bmrwisc.edu>). Together, these wwPDB partners collect, annotate, validate, and disseminate standardized PDB data to the public without any limitations on its use.

Over the past decade, the number of polymers and their complexes with small molecules in the PDB has increased significantly. Information about small molecules (bound ligands) and their interactions with proteins and nucleic acids is crucial for our understanding of biochemical processes and vital for structure-based drug design. Therefore, providing standard representation and accurate chemical descriptions of these ligands and their polymers is a primary focus of PDB biocuration. To continue to support scientific advancements and ensure the best quality of data files, I have contributed to the development of new tools (Jasmine

Young *et al.* 2013, Database) and dictionary standards (John Westbrook, *et al.* 2014, Bioinformatics). These standards will be used by the wwPDB organization to standardize new and existing examples in the PDB archive.

Development of tools and processes for biocuration and validation are a key task for the RCSB PDB and wwPDB organizations. I have contributed to the improvement of more uniform and higher quality of the PDB archive by creating controlled vocabularies, defining new data items, setting functional requirements, and standardizing PDB archival data. As a Biocuration Team Lead, I have been involved in validation workshops convened by the wwPDB to collect recommendations and develop community consensus on validation tasks and software. I then oversee the implementation of these recommendations. I also manage wwPDB projects for the development of data deposition, curation and validation tools (OneDep system, Jasmine Young *et al.*, submitted)). There have been 20170 PDB entries deposited and 13628 PDB entries released since the deployment of OneDep system. I am also taking leadership role in the data standardization of the PDB archive. For example, 3,800 X-ray experimental data files have been standardized; Biological assemblies for 5837 PDB entries have been corrected, and 1029 PDB entries containing small polypeptide and antibiotic molecules have been updated with uniform representation (Shuchi Dutta *et al.* 2014, Biopolymers).

As a member of ISB, I would like to contribute my expertise to the biocuration community and to promote the field to the broader community.

## **Biography**

Jasmine Young received a Ph.D. in Chemistry from University of Florida. She is trained in structure biology and bioinformatics. Since 2003, her work has focused on the data deposition, biocuration, and representation of biological macromolecules. As a member of the RCSB Protein Data Bank ([rcsb.org](http://rcsb.org)), she has worked as a Biocurator responsible for the annotation of data submitted to the Protein Data Bank archive and a Biocuration Trainer of biocurators at RCSB PDB and PDBj. Since 2008, in her role as the RCSB PDB Biocuration Team Lead, she oversees the practice, management, and tool developments for data curation, validation, and standardization of the PDB archive with the goal of ensuring data correctness and uniformity. Since 2015 she has been appointed as Global Project Lead managing various international projects for the development of deposition and curation tools and for the archival data standardization.

She leads on-going data analysis and strategic planning for improved PDB data representation. She coordinates with software development team to set requirements that meet new representations, to carry out data standardization, to review data consistency, and to schedule release to the PDB archive. These efforts have enabled a more effective representation of the data.

As part of the leadership team at RCSB PDB, she has been involved in the development of tools and resources for data query, analysis, and visualization that are driven by improvements in data annotation and representation. The improved uniformity of data representation and better data quality of the PDB archive has made the data more accessible to the user community. To achieve this, she has been involved in the design of the website resource for data query, reporting and visualization. She has also contributed to new feature enhancement for the support of incoming structures from newly evolved science and methodologies (Peter Rose, *et al.* 2015, NAR).

In her role of representing the Worldwide Protein Data Bank (wwPDB) globally, she collaborates with the wwPDB partners to resolve differences in biocuration practices and to set standard procedures and requirements. She leads and manages international wwPDB projects including the development of unified deposition and biocuration system and the data representation and standardization in the PDB archive. She has been involved in creating validation standards at method-specific Validation Task Force Workshops (Paul Adam, *et al.* 2016, Structure; R. Henderson, *et al.* 2012, Structure). These efforts have had significant impact on increasing the understanding of quality measures and requirements for PDB archive going forward and will greatly enhance the ability of the scientific community in general to use our data effectively.

#### Honors

- A co-author of many peer-reviewed publications. Full Bibliography: <http://www.ncbi.nlm.nih.gov/myncbi/browse/collection/49198469/?sort=date&direction=ascending>
- Served as reviewer for *PLoS Computational Biology* and for *Database*.
- Invited speaker/presenter at various scientific meetings:
  - Presentation at PSI Protein Modeling Workshop: "Information Captured in the PDB for Experimental Structures" in 2013
  - Representative of RCSB PDB at NSF workshop: "Curating for Quality: Ensuring Data Quality to Enable New Science" at DC in 2012
  - Presentation at U.S. Government Chemical Databases and Open Chemistry: "Management and Distribution of Chemical Data in the PDB" in 2011
  - Panel discussion at U.S. Government Chemical Databases and Open Chemistry: "How do we get chemistry information in a computer-readable format into primary chemical literature?" in 2011

#### Conflict of interest

Person/organization	Nature of relationship and/or nature of conflict of interest
Person: Stephen Burley	PI
Person: John Westbrook	Co-author and co-investigator

Person: Mike Gilson	Collaborator
Person: Rommie Amaro	Collaborator
Person: Helen Berman	Collaborator
Person: Robert Woods	Collaborator
Organization: RCSB Protein Data Bank	Current employer
Organization: Protein Databank at Europe (PDBe)	Collaborator
Organization: Protein Databank at Japan (PDBj)	Collaborator
Organization: Biological Magnetic Resonance Bank (BMRB)	Collaborator
Activity: Enabling Data Science in Biology (eDSB)	Co-PI
Activity: wwPDB OneDep project	Project Lead
Activity: GroupDep in D3R project	Project Lead
Activity: wwPDB File Version project	Project Lead
Activity: wwPDB carbohydrate project	Project Lead
Activity: PDBx/mmCIF Working Group	Collaborator
Activity: Validation Task Forces	Collaborator
Activity: CASP target	Collaborator
Activity: RCSB PDB Advisory Committee	Advisor
Activity: wwPDB Advisory Committee	Advisor
Activity: D3R project	Collaborator
Journal: Database	Reviewer
Journal: PLoS	Reviewer