

Visualizing the Impact of a Publication: “The Protein Data Bank” in *Nucleic Acids Research*, 2000

Dr. Talapady Bhat
talapady.bhat@nist.gov

Susan Makar
susan.makar@nist.gov

Amanda Malanowski
amanda.malanowski@nist.gov

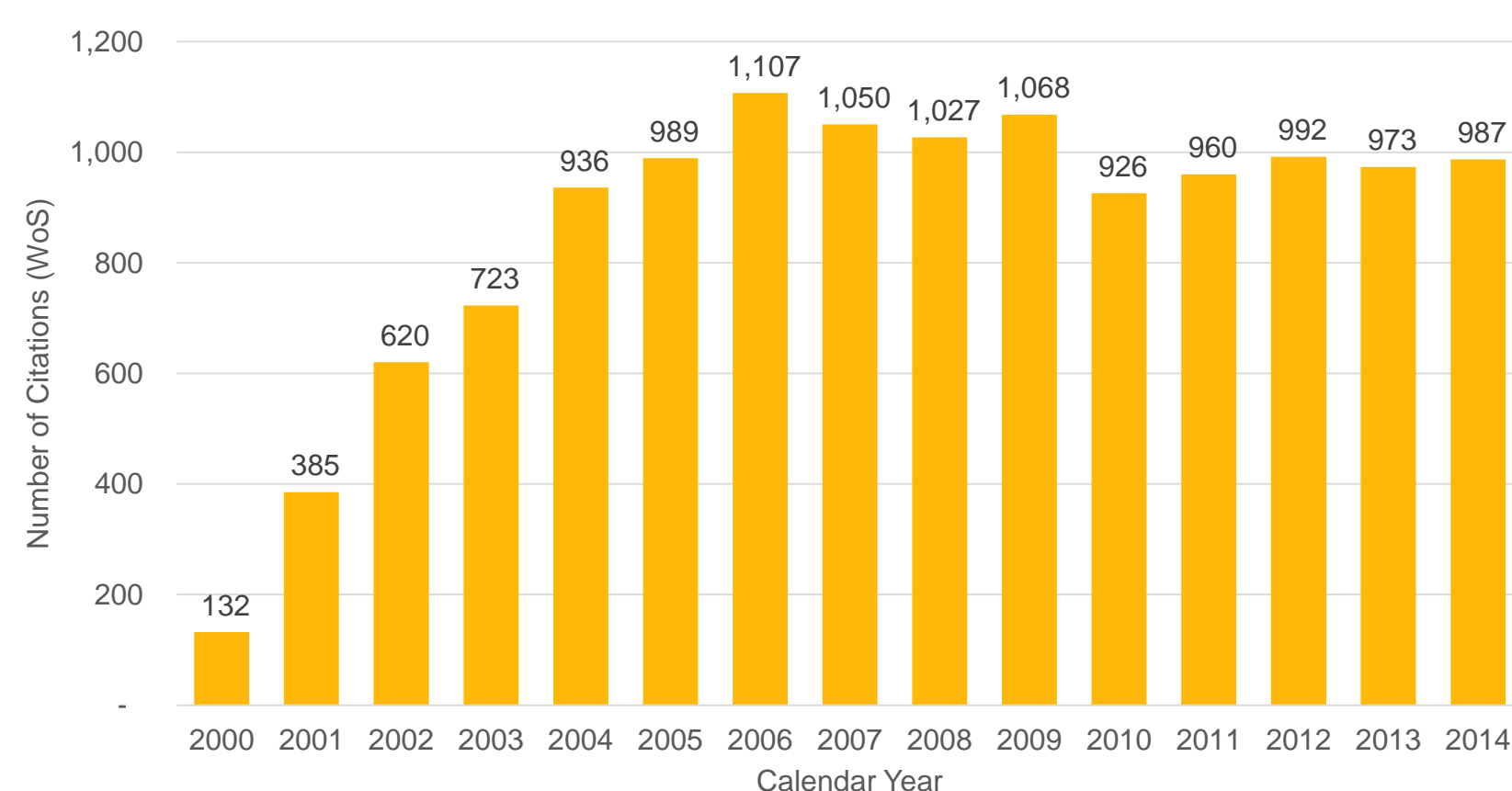
Introduction

The [Protein Data Bank](#) (PDB) was established at Brookhaven National Laboratories in 1971 as an archive for biological macromolecular crystal structures. Originally, it contained only seven structures; today it holds over 109,000 3D structures for large biological molecules, including proteins and nucleic acids. In October 1998, the management of the PDB became the responsibility of the Research Collaboratory for Structural Bioinformatics (RCSB), a consortium composed of Rutgers, the State University of New Jersey; the University of California at San Diego; and the National Institute of Standards and Technology (NIST).

This analysis examines the impact of the most highly cited NIST-authored paper, “The Protein Data Bank,” published in *Nucleic Acids Research* in 2000. This poster illustrates the impact of this article using various data visualizations.

Overview of Citations to “The Protein Data Bank”

The analysis, based on a *Web of Science* (WoS) search on August 5, 2015, yielded 13,657 citations to “The Protein Data Bank” since it was published in 2000. A Google Scholar search December 10, 2015 yielded 19,814 citations.



Methodology

A *Web of Science* search identified papers citing “The Protein Data Bank” since 2000. WoS was selected for this study because of its interdisciplinary coverage, its indexing of the peer-reviewed literature, and because it easily exports publication data for analysis. The citing papers were then analyzed using Tableau to create this poster’s graphics.

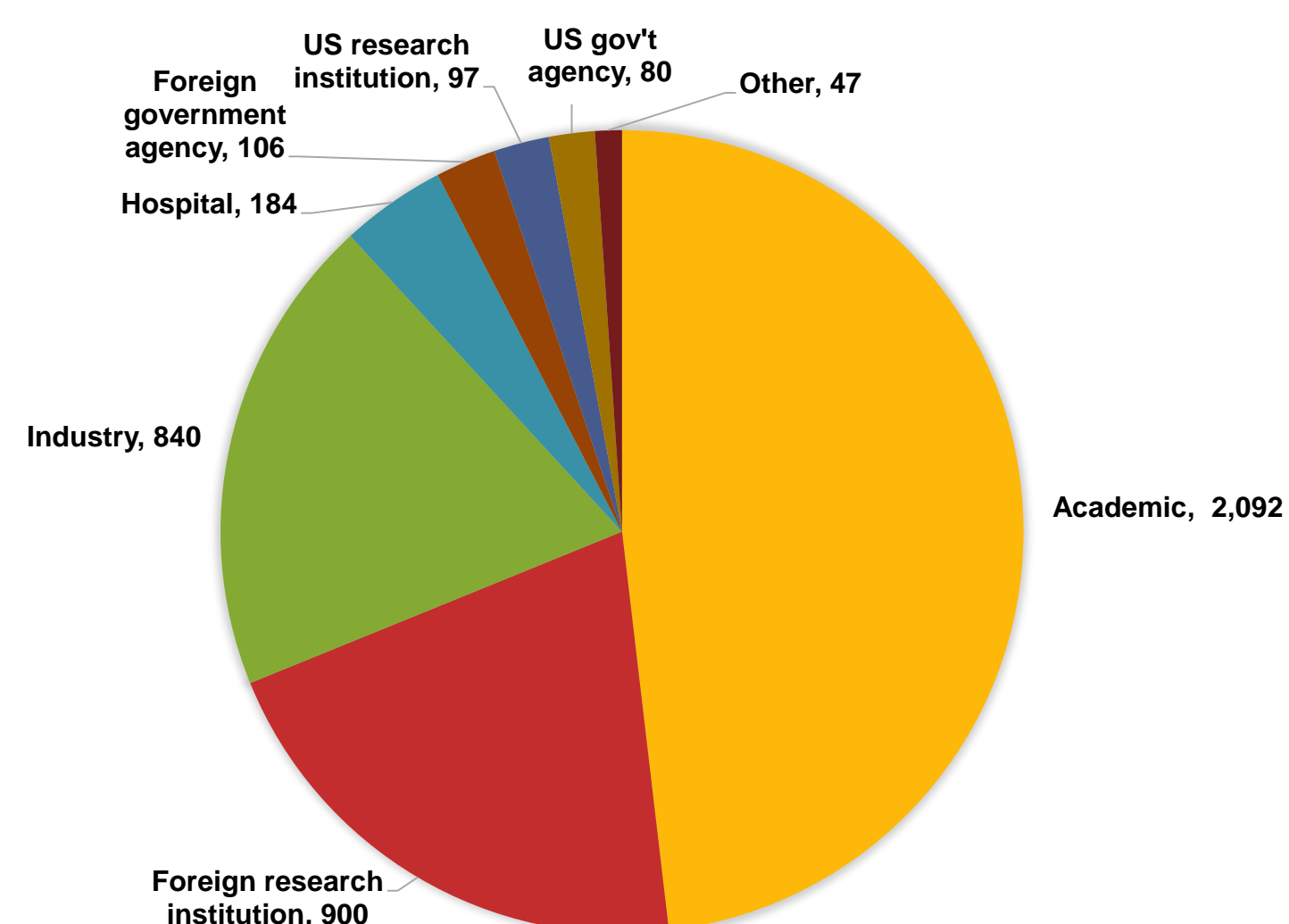
Findings

Top Papers Citing “The Protein Data Bank”

“The Protein Data Bank” has been cited by a variety of articles, some of them very highly cited themselves. Its citing by “Coot: model-building tools for molecular graphics,” with 12,176 citations, further increases the visibility and impact of “The Protein Data Bank.”

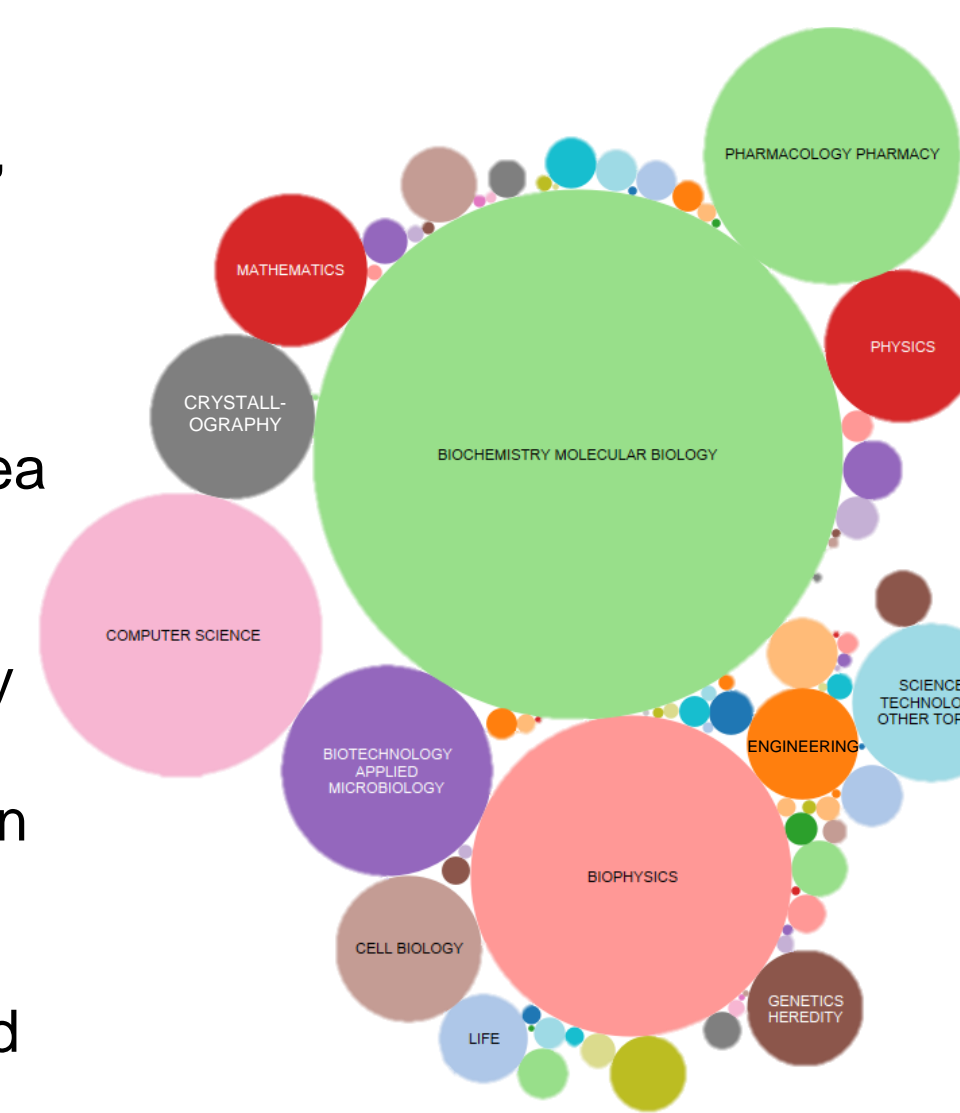
Institutional Impacts

“The Protein Data Bank” is cited by a variety of institution types. It is cited most frequently by academic institutions with 2,092 institutions citing it a total of 21,344 times.



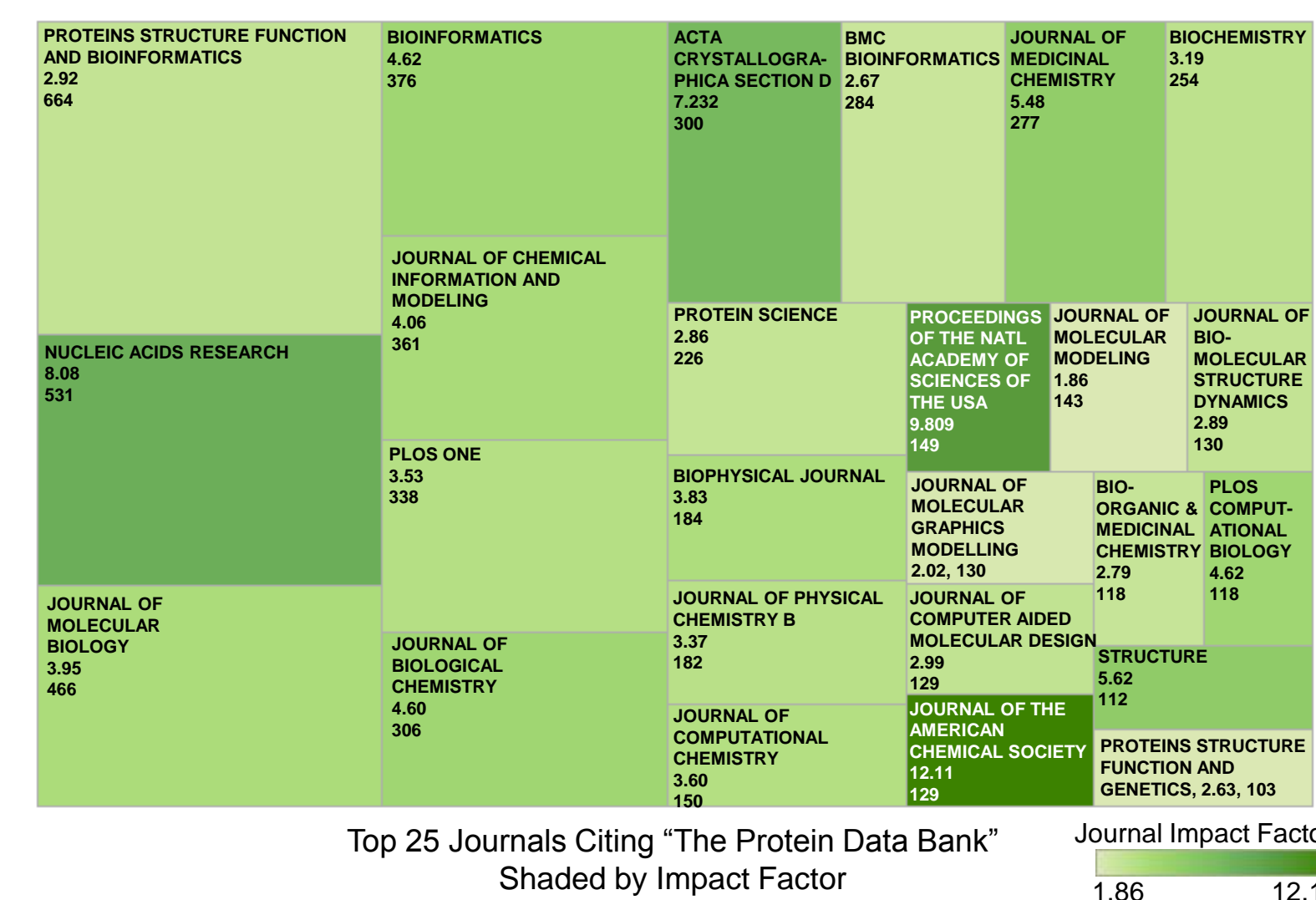
Citing Research Areas

“The Protein Data Bank” has far-reaching impact with papers from 95 different research areas citing the article. The area showing the highest impact, with 7,079 citations, is Biochemistry and Molecular Biology. The article has also been cited by papers in areas ranging from Computer Science to Dentistry, and Biophysics to Geology.



Citing Journals

“The Protein Data Bank” has been cited by over 1,800 journals. *Proteins: Structure, Function, and Bioinformatics*, with an Thomson Reuters impact factor of 2.921, cited “The Protein Data Bank” most often with 664 citations.



International Impacts

“The Protein Data Bank” has been cited by authors in 163 different countries. It has been cited most frequently by authors in the United States (4,998 authors), England (1,257 authors), Germany (1,240 authors), and India (926 authors).



Conclusions and Recommendations

As the most highly cited NIST-authored paper, “The Protein Data Bank,” with over 13,000 citations (>19,000 according to Google Scholar), is considered a classic in its field. There are many ways to measure the impact of this paper, as demonstrated through the various graphical representations shown in this poster. “The Protein Data Bank” has been cited across 95 different research areas in over 1,800 journals by authors from over 4,000 institutions in 163 countries.

Too often, impact is measured simply by the number of citations, when in fact, a much richer story can be told through a close look at the citation data.



The Protein Data Bank
Nucleic Acids Research
Volume 28, Issue 1, Pages 235-242, 2000