New Online Curriculum: The PDB Pipeline & Data Archiving

Catherine L. Lawson[1], Margaret J. Gabanyi[1], John Westbrook[1], Jasmine Young[1], Shuchismita Dutta[1], Ezra Peisach[1], Brian P. Hudson[1], Peter Rose[3], Jose Duarte[2], Amy Sarjeant[4], Stephen K. Burley[1], Helen M. Berman[1]

[1]RCSB Protein Data Bank, Institute for Quantitative Biomedicine, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA

[2]RCSB Protein Data Bank, San Diego Supercomputer Center, University of California, San Diego, La Jolla, CA 92093, USA

[3]Structural Bioinformatics Laboratory, San Diego Supercomputer Center, University of California, San Diego, La Jolla, CA 92093, USA

[4]Cambridge Crystallographic Data Centre, Institute for Quantitative Biomedicine, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA

Drawing from our collective experience as structural biologists, data scientists, and educators developing and managing the RCSB PDB and other related resources, we have created an open access modular educational curriculum covering concepts, approaches and requirements for developing and managing the full data pipeline for a curated public archive of biological experimental data. The online curriculum makes best practices recommendations for data resource management based on the extensive experience accumulated by the RCSB team. The intended audience includes scientists, who can use the materials for self-instruction, as well as librarians and information specialists, who can use the materials to develop training services for students, scientists, and staff at their organizations. In addition, the curriculum is intended to help catalyze formation of federated model and data archives to accelerate progress in Integrative Structural Biology. We have produced eight modules that can either be studied separately or assembled into a complete set as an open online course. Curriculum materials include professionally produced videos, powerpoint slides, and a set of exercises to guide students step-by-step to design, create, and query their own database.