biomolecules

MDPI

Review

Protein Data Bank: A Comprehensive Review of 3D Structure Holdings and Worldwide Utilization by Researchers, Educators, and Students

Stephen K. Burley [1,2,3,4,5,*], Helen M. Berman [1,2,5], Jose M. Duarte [4], Zukang Feng [1,2], Justin W. Flatt [1,2], Brian P. Hudson [1,2], Robert Lowe [1,2], Ezra Peisach [1,2], Dennis W. Piehl [1,2], Yana Rose [4], Andrej Sali [6], Monica Sekharan [1,2], Chenghua Shao [1,2], Brinda Vallat [1,2,3], Maria Voigt [1,2], John D. Westbrook [1,2,3,†], Jasmine Y. Young [1,2] and Christine Zardecki [1,2]

1 Research Collaboratory for Structural Bioinformatics Protein Data Bank, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA
2 Institute for Quantitative Biomedicine, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA
3 Cancer Institute of New Jersey, Rutgers, The State University of New Jersey, New Brunswick, NJ 08901, USA
4 Research Collaboratory for Structural Bioinformatics Protein Data Bank, San Diego Supercomputer Center, University of California San Diego, La Jolla, CA 92093, USA
5 Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA
6 Research Collaboratory for Structural Bioinformatics Protein Data Bank, Department of Bioengineering and Therapeutic Sciences, Department of Pharmaceutical Chemistry, Quantitative Biosciences Institute, University of California San Francisco, San Francisco, CA 94158, USA
* Correspondence: stephen.burley@rcsb.org
† Deceased.

Abstract: The Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB), funded by the United States National Science Foundation, National Institutes of Health, and Department of Energy, supports structural biologists and Protein Data Bank (PDB) data users around the world. The RCSB PDB, a founding member of the Worldwide Protein Data Bank (wwPDB) partnership, serves as the US data center for the global PDB archive housing experimentally-determined three-dimensional (3D) structure data for biological macromolecules. As the wwPDB-designated Archive Keeper, RCSB PDB is also responsible for the security of PDB data and weekly update of the archive. RCSB PDB serves tens of thousands of data depositors (using macromolecular crystallography, nuclear magnetic resonance spectroscopy, electron microscopy, and micro-electron diffraction) annually working on all permanently inhabited continents. RCSB PDB makes PDB data available from its research-focused web portal at no charge and without usage restrictions to many millions of PDB data consumers around the globe. It also provides educators, students, and the general public with an introduction to the PDB and related training materials through its outreach and education-focused web portal. This review article describes growth of the PDB, examines evolution of experimental methods for structure determination viewed through the lens of the PDB archive, and provides a detailed accounting of PDB archival holdings and their utilization by researchers, educators, and students worldwide.

Keywords: Protein Data Bank; Open Access; Worldwide Protein Data Bank; macromolecular crystallography; cryogenic electron microscopy; cryogenic electron tomography; electron crystallography; micro-electron diffraction; nuclear magnetic resonance spectroscopy; biological macromolecules; proteins; nucleic acids; DNA; RNA; carbohydrates; small-molecule ligands

check for updates

Citation: Burley, S.K.; Berman, H.M.; Duarte, J.M.; Feng, Z.; Flatt, J.W.; Hudson, B.P.; Lowe, R.; Peisach, E.; Piehl, D.W.; Rose, Y.; et al. Protein Data Bank: A Comprehensive Review of 3D Structure Holdings and Worldwide Utilization by Researchers, Educators, and Students. *Biomolecules* **2022**, *12*, 1425. https://doi.org/10.3390/biom12101425

Academic Editors: Cameron Mura and Lei Xie

Received: 30 August 2022
Accepted: 26 September 2022
Published: 4 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

The Protein Data Bank (PDB) is now in its 51st year of continuous operations. As the first open-access digital data resource in biology, it was established in 1971 with just

seven protein structures [1]. At the time of writing, PDB holdings numbered nearly 200,000 experimentally-determined three-dimensional (3D) structures of proteins and nucleic acids (DNA and RNA) and their complexes with one another and small-molecule ligands (e.g., enzyme co-factors, drugs, investigational agents). Since 2003, the PDB archive has been jointly managed by the Worldwide Protein Data Bank (wwPDB, wwpdb.org, accessed on 28 August 2022) partnership [2,3]. wwPDB Full Members include the US-funded Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB, RCSB.org, [4–7]); Protein Data Bank in Europe (PDBe, PDBe.org, [8]); Protein Data Bank Japan (PDBj, PDBj.org, accessed on 28 August 2022 [9]); the Electron Microscopy Data Bank (EMDB, emdb-empiar.org, accessed on 28 August 2022 [10,11]); and the Biological Magnetic Resonance Bank (BMRB, bmrb.io, accessed on 28 August 2022 [12,13]). The activities of the wwPDB are governed by a charter, which was last renewed in 2021 on the occasion of the accession of EMDB (www.wwpdb.org/about/agreement, accessed on 28 August 2022). The RCSB PDB is headquartered at Rutgers, The State University of New Jersey with smaller teams based at the University of California San Diego (UCSD) and the University of California San Francisco (UCSF). Within the wwPDB, RCSB PDB serves as the designated Archive Keeper for the PDB, responsible for safeguarding both digital information and a physical archive of correspondence. A conservative estimate of USD 100,000 for the average replacement cost of each individual PDB structure translates to a replacement cost of the structures in the entire archive of nearly USD 20 billion (as of mid-2022).

wwPDB partners are committed to the FAIR (Findability, Accessibility, Interoperability, and Reusability [14]) and FACT (Fairness, Accuracy, Confidentiality, and Transparency [15]) Principles emblematic of responsible data stewardship in the modern era. The PDB archive has been accredited by CoreTrustSeal (coretrustseal.org accessed on 28 August 2022). Since its inception, the PDB has been regarded as a pioneer in the open-access data movement. More than 60,000 structural biologists working on every inhabited continent have generously deposited 3D structure information (atomic coordinates, experimental data, and related metadata) to the archive over more than fifty years. Today, many millions of PDB data consumers worldwide working in fundamental biology, biomedicine, bioengineering, biotechnology, and energy sciences enjoy no-cost access to 3D biostructure information with no limitations on data usage. Many scientific research areas have been profoundly impacted by the creation and availability of the PDB archive [16–42].

This review article is published in a Special Issue of *Biomolecules* honoring Professor Phil Bourne, who served as Associate Director of the RCSB PDB from 1998–2014. Phil led the UCSD site, where he focused on database development, integration with the scientific literature, and PDB search and data visualization tools. Bourne and Helge Weissig played critical roles in developing the inaugural version of the RCSB PDB data-delivery web portal at RCSB.org [4,43]. Access to PDB data and development of tools for query, visualization, and analysis as supported by the wwPDB partnership have helped drive the growth of structural and computational biology. PDB data and its usage by researchers, educators, and students over more than five decades is presented to highlight the evolution of these scientific fields and inform the next fifty years of successful PDB operations.

2. Results

2.1. PDB Data Metrics and Trends

Since 1971, PDB structures have been contributed freely by more than sixty thousand structural biologists (depositors) working on every permanently inhabited continent (Figure 1). Structural biologists in 53 countries, territories, etc. recognized by the United Nations deposited data to PDB during 2021. All used the wwPDB OneDep software system (deposit.wwpdb.org) that enables complete structure data deposition [44], rigorous validation [45,46], and expert biocuration [47]. OneDep currently supports 3D macromolecular structures determined using the following experimental methods: macromolecular crystallography (MX), 3D electron microscopy (3DEM), nuclear magnetic resonance (NMR)

spectroscopy, electron crystallography (EC), and micro-electron diffraction (microED). Currently, newly deposited structures are processed at RCSB PDB (Americas, Oceania), PDBe (Europe, Africa), or PDBj (Asia, Middle East), allocated based on the depositor's IP address location.

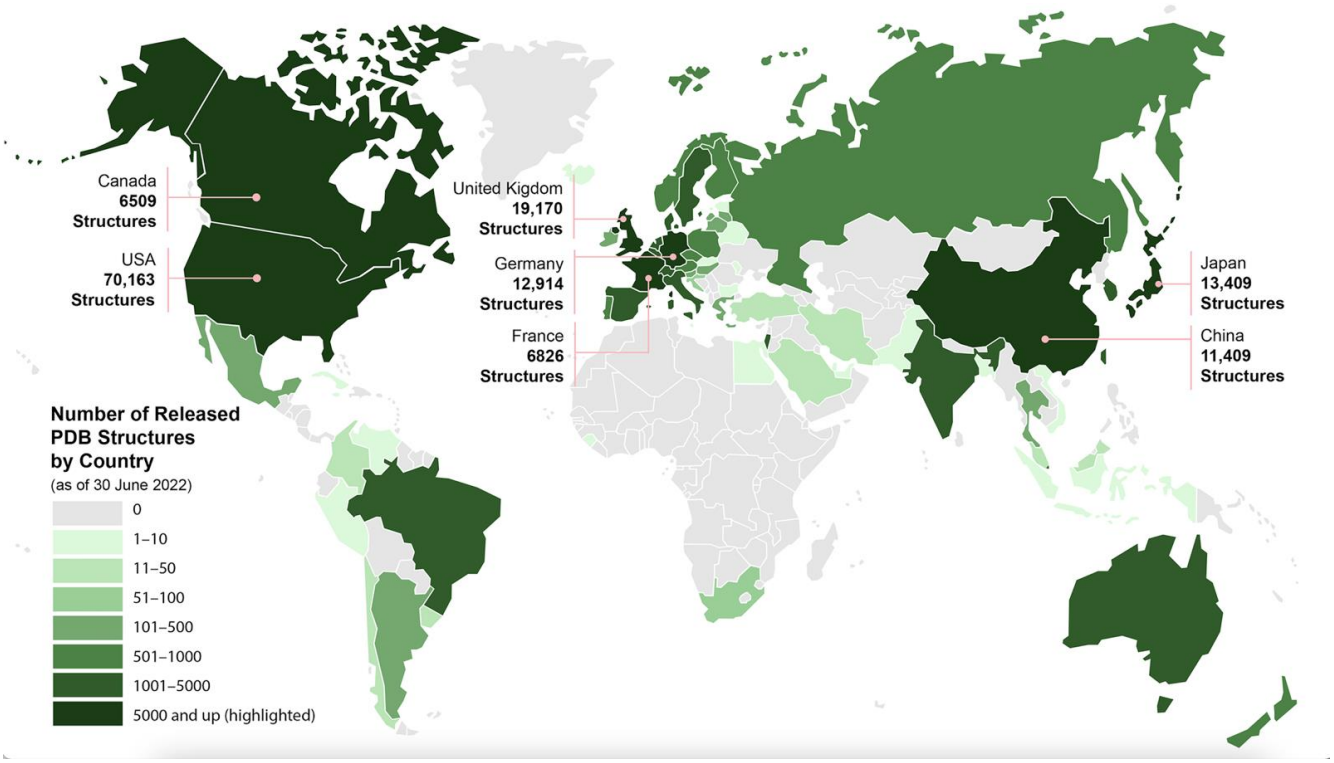

Figure 1. Geographic distribution of PDB depositions from 1971 to mid-2022.

Figure 2A illustrates growth of the PDB archive over the past 50+ years. Since the first X-ray crystal structure of a protein (sperm whale myoglobin) was determined by Sir John Kendrew and his colleagues [48], the discipline has become central to molecular and cellular biology. Figure 2B documents the impact of MX, 3DEM, and NMR on annual PDB data releases. Since 2016, annual releases of PDB MX structures have plateaued at ~10,000, with the exception of substantial spike in 2020 driven by the pandemic lockdown and various MX-based fragment screening campaigns against SARS-CoV-2 proteins thought to represent good drug discovery targets. During the same period, NMR structure releases declined, and 3DEM structure releases grew exponentially (increasing ~6-fold in only 4 years). As of mid-2022, the archive contained 166,894 MX structures, 11,294 3DEM structures, and 13,738 NMR structures. Given current deposition metrics, aggregate 3DEM structure holdings are expected to surpass those of NMR in late 2022 or early 2023. Of immediate importance to those working to combat the COVID-19 pandemic, the PDB archive currently holds >2600 SARS-CoV-2 related structures (~800 released in 2020, and ~900 released in 2021). Figure 2C shows the number of PDB MX and 3DEM structures broken down as a function of resolution (median value ~2.0 Å). While nearly all PDB structures determined at better than 2.5 Å resolution came from MX (~99.6%), 3DEM is now capable of delivering structures to nearly 1Å resolution (e.g., 1.15 Å resolution structure of apoferritin, PDB ID 7a6a [49]).

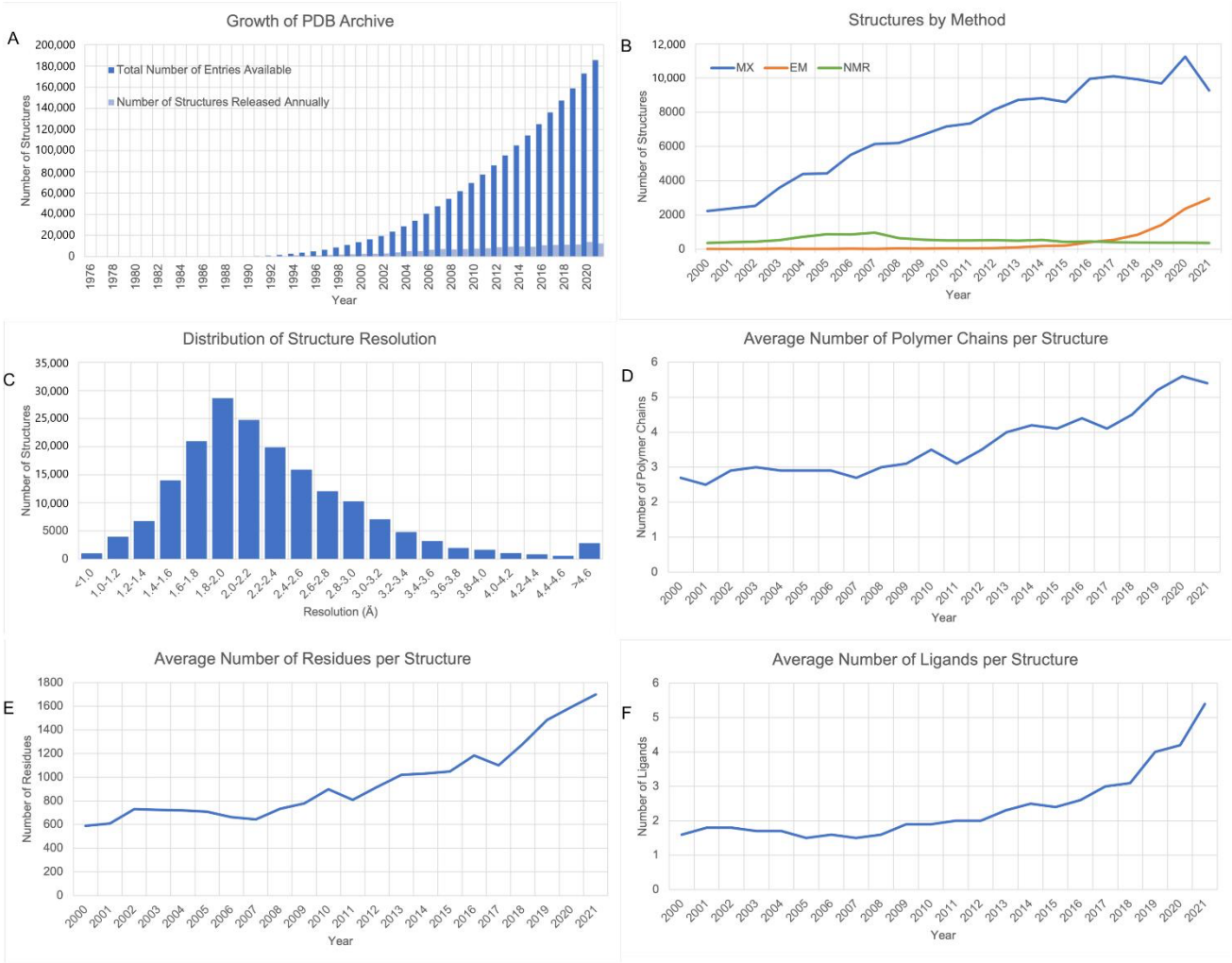

Figure 2. PDB archive metrics. (**A**). Growth 1976–2021. (**B**). New MX, 3DEM, and NMR structures released annually (2000–2021). (**C**). MX and 3DEM structure counts vs. resolution (Å). (**D**). Average number of residues per structure for structures released annually (2000–2021). (**E**). Average number of polymer chains per structure for structures released annually (2000–2021). (**F**). Average number of non-polymer ligands per structure for structures released annually (2000–2021).

While the total number of PDB structures continues to grow, their complexity is increasing year-on-year. Figure 2D illustrates structure complexity as a function of time as judged by the average number of amino acid and/or nucleotide residues per PDB ID. As of mid-2022, the total number of residues (proteins and nucleic acid) in the archive exceeded 200 million and the total number of atoms exceeded 1.5 billion. Figure 2E,F show similar trends for the average number of polymer chains per PDB ID and average number of ligands per PDB ID (excluding bound water molecules, other solvents, salts, ions, common buffers, crystallization and cryoprotection agents as specified in Shao et al. [50]), respectively.

2.2. Evolution of Structural Biology Methods Viewed through the Lens of the PDB

As evidenced in Figure 2A, growth of the PDB has been much faster than linear. This section examines the evolution of structural biology as a discipline viewed through the lens of PDB archival holdings. Technical innovations in MX, 3DEM, and NMR are discussed in some detail, followed by a brief account of the emergence of microED as an exciting new diffraction method for structure determination of biological macromolecules.

2.3. Macromolecular Crystallography (MX)

Structures determined using the MX method were the first to be deposited into the PDB. All of these early structures were determined using isomorphous replacement (IR) [51] to solve the crystallographic phase problem. Slow but steady growth of the PDB archive during the 1980s combined with development of the molecular replacement (MR) method for structure determination by Michael Rossmann [52] helped to accelerate MX. In 2001, after PDB first began systematic collection of phasing method information, it was already apparent that most 3D structures being deposited to the archive were determined using MR. Figure 3 also shows that by 2001 IR had been largely abandoned as a de novo structure determination method in favor of multiple-wavelength anomalous dispersion (MAD, to be supplanted by single-wavelength anomalous dispersion or SAD) for new structure determinations for which MR was not feasible. Analyses across the entire archive revealed that MR was used to determine ~85% of all PDB MX structures as of mid-2022. This method depends critically on the parsimony of macromolecular evolution. Protein domain folds (3D structures) are reused repeatedly within biomolecules carrying out similar biochemical or biological functions. According to generally accepted estimates, ~10,000 distinct polypeptide chain folds account for the vast majority of naturally occurring proteins.

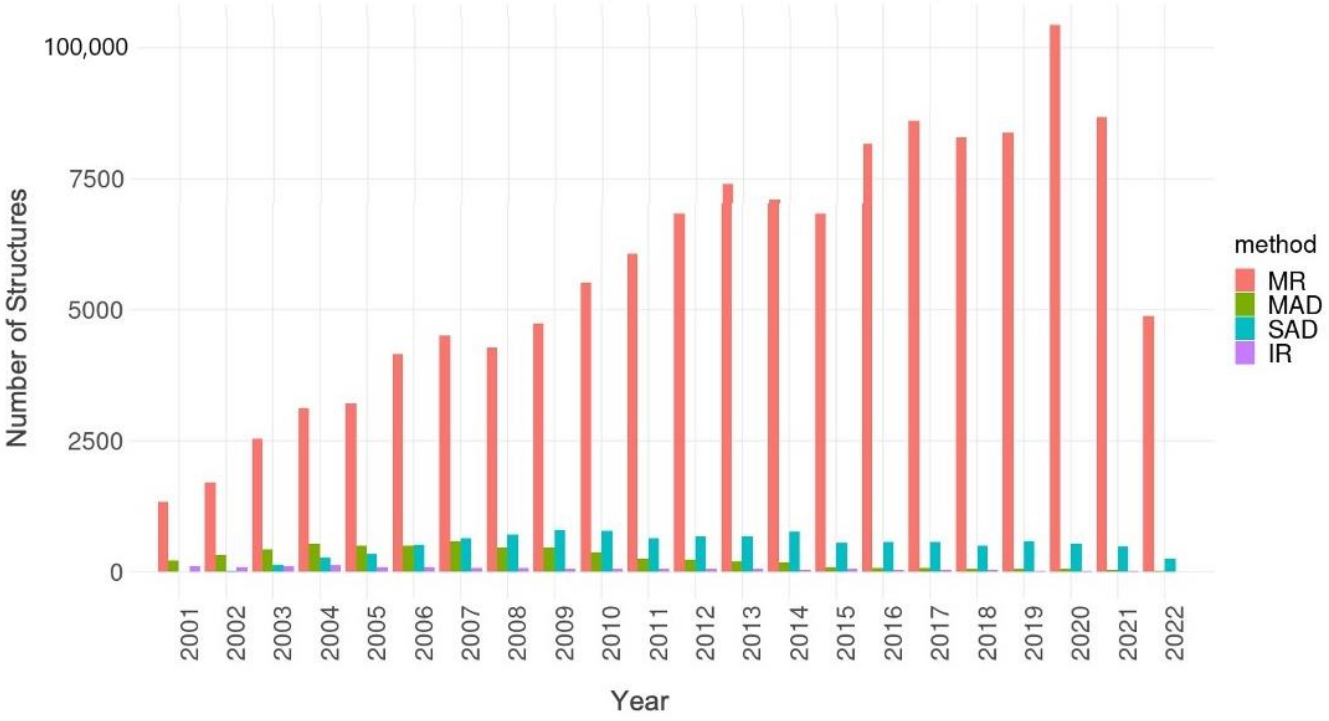

Figure 3. PDB MX structure phasing method trends vs. year of structure release from 2001–2021 (MR: molecular replacement; MAD: multi-wavelength anomalous dispersion; SAD: single-wavelength anomalous dispersion; IR: isomorphous replacement).

The other important trend in MX structure determination practices evident from historical PDB data concerns X-ray sources. Widespread availability of MX beamlines at synchrotron radiation sources transformed how protein crystallographers work. As of mid-2022, ~85% of PDB MX structures relied on diffraction data collected at synchrotrons vs. ~15% that used home X-ray sources. Before 2000, most PDB MX structures released annually were the products of home sources. In contrast, only ~7% of new PDB MX structures came from home sources during the period of 2017 through 2021. Among global synchrotron sources worldwide, the top five contributors of PDB MX structures in rank order as of mid-2022 were the Advanced Photon Source (APS, ~21% of all PDB MX structures), the European Synchrotron Research Facility (ESRF, ~12%), Diamond (~9%), the Advanced

Light Source (ALS, ~7%), and the National Synchrotron Light Source (NSLS, ~6%). Three of these top five biostructure-producing synchrotrons (APS, ALS, and NSLS) and others operated by the US Department of Energy contributed ~41% of all PDB MX structures worldwide as of mid-2022.

Given the critical roles played by synchrotron radiation sources in MX structure studies, one could reasonably expect that bright X-ray sources combined with cryogenic data collection would have contributed to ongoing improvements in structure resolution throughout the history of the PDB. Figure 4 tells an entirely different story. As of 1990, well before access to synchrotron beamlines and cryo-cooling of protein crystals became routine, median resolution of new MX structures released by the PDB annually plateaued at ~2.0 Å. Since then, median resolution of PDB MX structures has not changed appreciably. This reality almost certainly reflects limitations due to the degree of order (or disorder) typical of crystalline preparations of biological macromolecules. Absent new crystallization strategies that markedly increase the order of protein crystals or modeling methods that deconvolute this disorder into multiple structural states, it appears unlikely that median resolution of MX structures in PDB will improve substantially, if at all. Fortunately for most PDB data consumers, 2 Å resolution usually suffices to reveal features of macromolecules relevant for understanding biological phenomena in 3D. In contrast, higher resolution studies may be required to understand fully biochemical functions of proteins and nucleic acids (e.g., reactions catalyzed by protein enzymes and ribozymes).

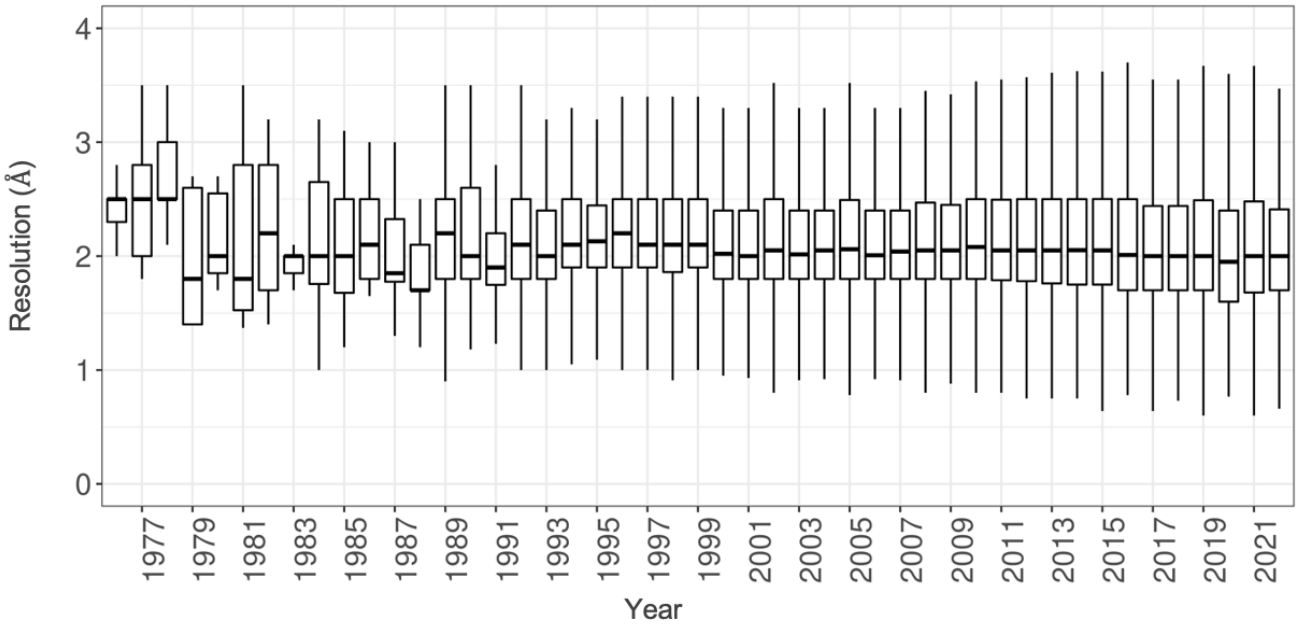

Figure 4. Box plot display of PDB MX structure resolution vs. time. The bold solid bar within each box corresponds to the median value for structures publicly released that year. (N.B.: Small numbers of extreme outliers with resolution > 4 Å were excluded from this analysis for clarity).

Geometric validation of atomic coordinates deposited to the PDB was introduced in the 1990s. Validation of 3D structures vs. experimental structure factors was not routinely performed until 2008, when deposition of experimental structure factor data became mandatory at the behest of the MX community. Stakeholder recommendations regarding some additional means of validating MX structures were subsequently provided in 2011 by the wwPDB X-ray Validation Task Force [53] and implemented in wwPDB legacy deposition systems in 2013 before the wwPDB global OneDep system was launched in 2014 [44]. Availability of experimental data has enabled systematic validation of atomic structures and contributed to development of better validation tools [45] and improved quality of the archived data [54].

Notwithstanding numerous aspects of 3D structure validation initially implemented within the wwPDB OneDep software system validation module, ligand validation was somewhat limited at the outset. The 2016 wwPDB/CCDC/D3R Ligand Validation Workshop recommended best practices for validation of MX co-crystal structures [55]. These recommendations were subsequently incorporated into the OneDep validation module to provide "Buster-like" 2D geometry quality and 3D electron density graphical overlays with small-molecule ligands [46]. Validation of PDB MX structures was further enhanced with introduction of uniform representation for carbohydrates [56].

Arguably, one of the most exciting new methods for measuring diffraction data at the time of writing is serial crystallography [57–59]. This approach is being used to probe dynamic properties of proteins and nucleic acids and visualize progress of chemical reactions in 3D (e.g., *M. tuberculosis* β-lactamase (BlaC) inactivating the β-lactam antibiotic ceftriaxone: PDB IDs 6b5x, 6b5y, 6b6a-6b6f, 6b68, and 6b69 [60]). Both X-ray free-electron lasers (XFELs) and 3rd generation synchrotron sources are being used to conduct such experiments. As of mid-2022, PDB archival holdings included 587 serial crystallography structures, with 343 (~58%) coming from XFELs and 244 (~42%) based on data collected from synchrotrons. Additionally, 217 PDB MX structures were determined using XFEL data without recourse to serial methods (e.g., PDB ID 3pcq [61]).

2.4. 3D Electron Microscopy (3DEM)

Over the last decade, resolution of 3DEM PDB structures has improved dramatically. Since 2013, average resolution of a 3DEM PDB structure has improved from worse than 14 Å to better than ~4 Å (Figure 5A). These overall statistics, however, obscure some of the most impressive recent developments in 3DEM. Between the beginning of 2019 and mid-2022, 40 3DEM structures with resolution better than 2.0 Å were publicly released by the PDB.

Technical breakthroughs in four critical areas were responsible for this "Resolution Revolution" [62,63]. First, improvements in electron optics, driven by the needs of materials scientists and the semiconductor industry, ensure that state-of-the-art transmission electron microscopes (TEM, e.g., Thermo-Fisher Titan Krios, Waltham, MA, USA) preserve phase information at atomic resolution. Second, vitrification of biological samples and imaging under cryogenic conditions is now routine [64]. Third, direct electron detectors (DEDs) have revolutionized how we collect TEM data for single particles arrayed on EM grids. The move away from charge-coupled device (CCD) detectors to DEDs has been nothing short of a stampede. Figure 5B illustrates the trend. In 2013, only ~5% of new 3DEM PDB structures relied on DEDs. By 2017, the fraction relying on DEDs exceeded 90%, and in 2021 the fraction was ~99%. In aggregate, DEDs have been used to collect data for 10,406 3DEM PDB structures released as of mid-2022 (vs. 11,309 total 3DEM PDB structures). Finally, the other key contributor to the rapid rise of 3DEM has been advances made in data processing software. Key software engineering developments include beam-induced motion correction [65–67] and use of Bayesian maximum-likelihood statistics [68]. Figure 5C shows that the most popular 3DEM reconstruction software package at the time of writing is RELION [69], which has been used for determination of more than 4000 3DEM PDB structures since 2013.

Year-on-year growth of 3DEM PDB structure depositions evident in Figure 3B was driven by the single-particle method, which is revealing structures of ever more complex macromolecular assemblies and illuminating important areas of biology (e.g., ion channels, transcription–translation expressome complexes, nuclear pore complexes). Arguably even more exciting advances are yet to be made using cryo-electron tomography (cryo-ET) combined with sub-tomogram averaging [70]. One of the earliest cryo-ET structures in the archive is PDB ID 4bzj (40 Å resolution COPII Transport-Vesicle Coat Assembled on Membranes [71]). As of mid-2022, the highest resolution cryo-ET structure in the archive was PDB ID 7zbt (3.3 Å resolution RuBisCO visualized within native *Halothiobacillus neapolitanus* carboxysomes [72]). At better than 3.5 Å resolution, both α-helix and β-strand secondary

structural elements and bulky amino acid sidechains are discernible in experimental 3DEM density maps (deposited to EMDB) revealing molecular details in 3D important for understanding biochemical and biological function.

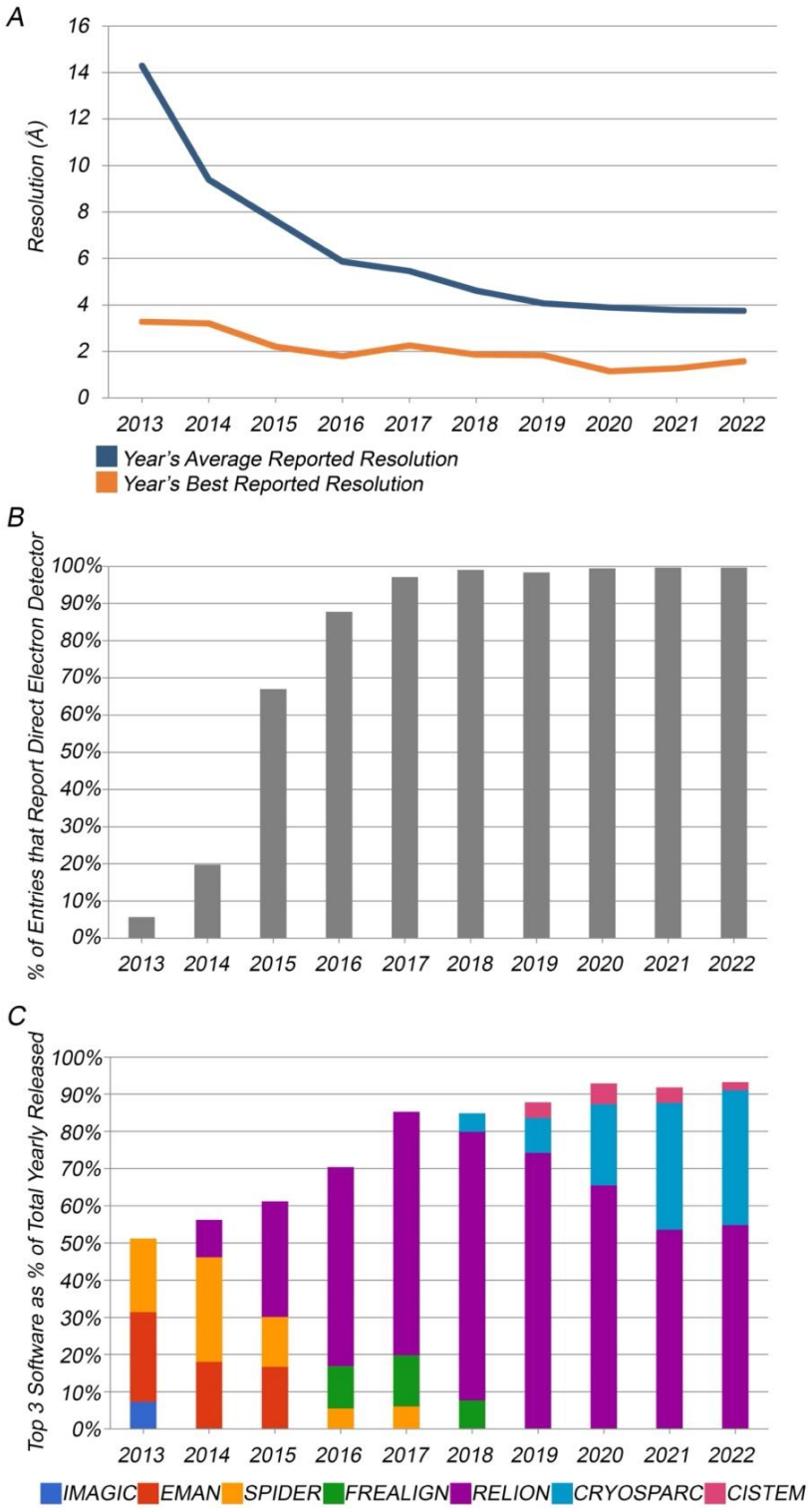

Figure 5. (**A**). Annual average reported resolution (blue) and annual best reported resolution (orange) for 3DEM PDB structures released 2013–2022. (**B**). Percentage of 3DEM PDB structures released per year reporting use of direct electron detectors. (**C**). Top-three reported image reconstruction software packages per year shown as a percentage of 3DEM PDB structures reporting reconstruction software.

The *H. neapolitanus* RuBisCO cryo-ET structure employed a relatively new sample preparation technique that relies on cryogenic dual-beam focused ion beam/scanning electron microscopes (cryo-FIB/SEM) to generate 10–20 nm thickness *lamellae* of vitrified samples using the focused ion beam to "mill" away unwanted parts of the sample. This tool allows researchers to isolate thin wafer-like volumes from inside frozen cells for subsequent cryo-ET imaging and sub-tomogram averaging. Immediate-term prospects for cryo-ET plus cryo-FIB/SEM milling with sub-tomogram averaging brightened considerable with the advent of AlphaFold2 [73–75] and RoseTTAFold [76]. For example, in 2021, computed structure models of human nuclear pore complex (NPC) proteins from AlphaFoldDB were combined with cellular cryo-ET and molecular dynamics simulations, to generate composite 3DEM density maps of the human NPC in both dilated and constricted conformations (PDB IDs 7r5k, 7tbl, 7tbm, 7tbj, 7tbk, and 7tbi [77]). Combining cryo-FIB/SEM with correlative light microscopy prior to cryo-ET imaging of *lamellae* holds the promise of improving the efficiency of the method by maximizing the number of molecular assemblies of interest present in a given wafer-like sample for imaging and subsequent sub-tomogram averaging [78].

At the time of writing, wwPDB validation reports for 3DEM structures included: (a) assessment of model geometry similar to that used for all MX and NMR structures (ClashScore, Ramachandran outliers, Sidechain outliers, nucleic acid polymer backbone); (b) orthogonal projections of map and map-model overlays; (c) half-map FSC plot based on mandatory half-maps collected at deposition; (d) voxel-value distribution and volume-estimation graph; (e) evaluation of map-model fit via atom-inclusion plot and residue inclusion analysis; and (f) finer evaluation of map-model fit incorporating both overall and per residue Q-scores [79]. EMDB also provides 3DEM density map and structure quality assessments on its website, including Q-scores [80]. (For more details regarding the history of 3DEM validation in the PDB, see [81]).

2.5. Nuclear Magnetic Resonance (NMR) Spectroscopy

Solution nuclear magnetic resonance (NMR) spectroscopy can be used to determine 3D structures of biomolecules (e.g., [82,83]). The first NMR structure of a protein was deposited to the PDB in 1988 and released publicly in 1989 (PDB ID 1bds [84]). By the end of the 1980s, solution NMR structures of 10 proteins had been determined, for which no crystallographic data were previously available [85]. At the same time, heteronuclear 3D and 4D NMR experiments were introduced to overcome limitations of spectral complexity and increased molecular weight (polypeptide chains longer than 150 amino acid residues, hereafter residues) [86]. At the beginning of the 1990s, the first NMR data file that included NMR restraints used to determine the 3D structure of Interleukin-8 (IL-8/NAP) was deposited to the archive (PDB ID 1il8 [87]). At the end of the 1990s, the first chemical shift file (containing a total of 179 chemical shifts) was deposited as part of PDB ID 1qlo [88]. Upon the recommendation of the wwPDB NMR Validation Task Force (NMR-VTF), NMR PDB structure depositions were required to include NMR restraint data and chemical shift data, in 2008 and 2010, respectively [89].

The number of new NMR structures released to the public annually from the PDB peaked in 2007 at 965, when NMR structures accounted for ~17% of the entire archive. Annual depositions have been trending downward ever since (362 NMR structures released publicly in 2021), and NMR structures now account for only ~7% of PDB holdings. As of mid-2022, the archive housed 13,733 NMR structures, 13,602 solution plus 131 solid-state. Figure 6 provides a breakdown of NMR PDB structures as a function of biomolecule sample type.

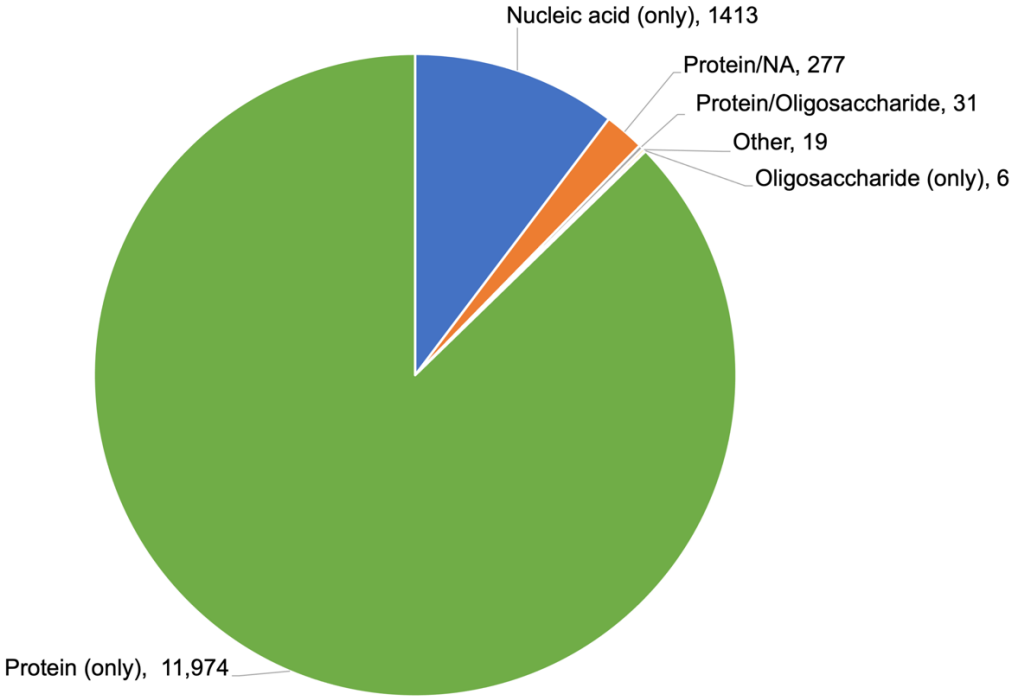

Figure 6. Breakdown of NMR PDB structure holdings by sample type.

Historically, NMR structural studies of biomolecules were size-limited. Most NMR PDB structures are those of smaller proteins or isolated protein domains (polymer entities < 8.5 kDa). Both solution and solid-state NMR (SSNMR) can, however, be used to study larger, more complex structures. SSNMR has been utilized to overcome some of the obstacles restricting the purview of solution NMR (e.g., relatively insoluble proteins). Both techniques can be deployed in tandem to overcome respective limitations. As of mid-2022, the PDB archive housed at least six structures determined using a combination of solution and SSNMR (e.g., *O. cuniculus* phosphorylated phospholamban homopentamer PDB ID 2m3b [90]).

Advances in technology for both solution and SSNMR have allowed for larger structures to be determined. For example, the largest solution NMR structure in the archive (as judged by total number of residues) is the Box C/D enzyme, a multimeric complex consisting of four instances of three unique proteins totaling 3044 residues (PDB ID 4by9 [91]). Additionally, use of magic angle spinning (MAS) SSNMR has enabled determination of structures with no inherent molecular size limitation, overcoming obstacles faced by solution NMR and MX. Exploiting these capabilities, SSNMR has been used to elucidate structures of complex assemblies similar in size to those studied by cryo-EM while in their native state, without the need for cryogenic preservation. As of mid-2022, the largest macromolecular structure determined by MAS SSNMR is the HIV-1 Capsid Tube, containing 378 repeats of a 231-residue subunit for a total of 87,318 residues (PDB ID 6x63 [92]). Larger structures have also been determined using integrative or hybrid methods, including that of a 484.61 kDa, 24mer αB-crystallin oligomer (4200 residues), incorporating experimental data from solution NMR, solution scattering, and 3DEM (PDB ID 3j07 [93]), and that of the 470.42 kDa tetrahedral aminopeptidase TET2 (4236 residues total), incorporating data from SSNMR and 3DEM (PDB ID 6r8n [94]).

With use of membrane-mimicking systems (e.g., micelles, bicelles, and nanodiscs), it is possible to study integral membrane proteins in their near-native environments using NMR [95]. A structure of the 7.77 kDa transmembrane domain of bacterioopsin (residues 1–71) was determined using solution NMR by solubilizing the protein in methanol/chloroform and SDS micelles, and deposited into PDB in 1993 (PDB IDs 1bha and 1bhb [96]). At the time of writing, the largest membrane protein structure determined via solution NMR deposited to the PDB is that of 149.16 kDa, 1360 residue human α7 nicotinic acetylcholine receptor, deter-

mined by a combination of solution NMR, electron spin resonance spectroscopy, and Rosetta calculations (PDB ID 7rpm [97]). As of mid-2022, the largest membrane protein structure determined by SSNMR in the PDB is that of 183.51 kDa, 1750 residue M13 bacteriophage capsid (PDB ID 2mjz [98]).

In addition to the study of 3D structures of biological macromolecules, examination of dynamics is often important for understanding function. Insights into a biomolecule's local dynamic behavior can be used to identify parts of structures important for ligand binding, protein–protein or protein–nucleic acid interactions, allostery, or conformational changes (e.g., integral membrane proteins). NMR spectroscopy is uniquely capable of studying macromolecular movement because of its ability to study samples spanning a wide range of solvent/solute conditions at atomic resolution over relevant timescales (i.e., picoseconds to seconds). Such studies are also possible using MAS SSNMR, which can be used to interrogate dynamics of the protein backbone atoms and sidechains (both globally and locally). As of mid-2022, the PDB archive housed results of dynamics studies of both small proteins (e.g., 8.58 kDa ubiquitin, PDB ID 2k39 [99]) and large biological nanomachines (e.g., 181.87 kDa proteasome subunit alpha heptamer, PDB ID 2ku1 [100]).

As is the case for MX and 3DEM, validation standards for NMR structures archived in the PDB are being developed collaboratively by the wwPDB and independent experts. Following implementation of chemical shift validation in 2015 at the behest of community stakeholders, the NMR Data Exchange Format (NEF) Working Group, which includes developers of NMR structure determination and refinement software packages, recommended use of a common exchange format to represent NMR chemical shifts, restraints, and related metadata [101]). NMR structure validation utilizing this unified exchange format was incorporated within the wwPDB OneDep software system and wwPDB validation reports in 2020. At the time of writing, archive-wide regeneration of extant NMR structure validation reports to enable restraint validation was underway. Completion of this remediation project and public release of regenerated wwPDB validation reports for all NMR structures archived in the PDB is anticipated in 2023. Additional improvements in wwPDB validation of NMR structures is expected to encompass data representation and validation of multiple conformers (e.g., pro-islet amyloid polypeptide open conformer (PDB ID 6ucj) and pro-islet amyloid polypeptide bent conformer (PDB ID 6uck [102]) and validation of structures determined using NMR combined with other experimental methods (e.g., PDB ID 3j07 [93]).

2.6. Electron Crystallography (EC) and Micro-Electron Diffraction (microED)

Electron diffraction or electron crystallography (EC) has also been used to determine 3D structures of biological macromolecules. The method employs 2D crystals, beginning with those of bacteriorhodopsin, the first integral membrane protein structure to be deposited into the archive (PDB ID 1brd [103], resolution 3.5 Å). Prior to 2013, a total of 37 biostructures determined using EC were deposited to PDB. With the advent of modern electron microscopes, a new electron diffraction method using miniscule 3D crystals (micro-electron diffraction or microED) has been developed [104]. The first microED structure of a globular protein (hen egg white lysozyme, PDB ID 3j4g [105], resolution 2.9 Å) was deposited to the PDB in late 2013. As of mid-2022, the PDB housed 137 microED structures of biomolecules, the largest two of which are human adenosine receptor A2a/cytochrome b562 chimeric protein (PDB ID 7rm5, 50 kDa, resolution 2.8Å [106]) and bovine catalase (PDB ID 3j7b, 60 kDa, resolution 3.2Å [107]). Unlike most EC structures archived in PDB, microED structures are typically determined at very high resolution. As of mid-2022, the highest resolution microED structure in PDB was that of hen egg white lysozyme (PDB ID 7skw [108], resolution 0.87 Å).

2.7. PDB Archive Management and Weekly Update/Release

The PDB data standard is defined by the PDBx/mmCIF dictionary [109–111]. It is the macromolecular extension of an earlier community data standard, the Crystallization Information Framework (cif.iucr.org, accessed on 28 August 2022), developed for small molecules by the International Union of Crystallography [112]. The macromolecular data standard is maintained by the wwPDB partnership together with the wwPDB PDBx/mmCIF Working Group (wwpdb.org/task/mmcif, accessed on 28 August 2022) [111]. wwPDB partners and the Working Group collaborate on developing terminologies for new and rapidly evolving methodologies and remediating (or enhancing) representations for existing data content.

In its role as wwPDB-designated PDB Archive Keeper, RCSB PDB is responsible for safeguarding >100 TB of digital information and a physical archive that includes correspondence and other archive-related artifacts dating back to the early 1970s. Snapshots of the digital information are preserved annually and following large-scale archive-wide data remediation campaigns, the most recent of which involved standardizing atom naming, etc. for >14,000 carbohydrate-containing structures in the PDB [56]. The size of the 2021 digital snapshot was ~1 TB, which does not include ~4.5 TB of 3DEM density map information archived in EMDB (also jointly managed by the wwPDB partnership).

In its role as wwPDB-designated Archive Keeper, RCSB PDB is responsible for weekly updates of the PDB archive using the following two-stage process:

Stage One releases sequence(s) for each distinct polymer (amino acid or nucleotide) in the structure; InChI string(s) for each distinct ligand; and crystallization *p*H value(s), where appropriate, on the wwPDB web portal (see www.wwpdb.org/ftp/pdb-ftp-sites, accessed on 28 August 2022) every Saturday by 03:00 Universal Time Coordinated (UTC). This first stage in the process supports weekly blind challenges for in silico prediction of protein structure (CAMEO, cameo3d.org, accessed on 28 August 2022 [113]) and small-molecule docking (CELPP, drugdesigndata.org/about/celpp, accessed on 28 August 2022 [114]).

Stage Two completes the weekly process every Wednesday at 00:00 UTC by releasing the updated PDB archive in full (currently adding ~300 new structures/week, updating previously released structures with literature citation information, etc., and on occasion removing obsolete structures).

PDB data are freely distributed online, providing universal open access to the archival information in two forms (latest archive, files.wwpdb.org/pub/pdb/data, accessed on 28 August 2022; and latest and prior versions of archive, files-versioned.wwpdb.org, accessed on 28 August 2022). Hypertext Transfer Protocol (HTTP) and remote sync (rsync) are recommended for access; File Transfer Protocol (FTP) access will be retired in late 2024. PDB data are also made available without storage fees or egress charges by Amazon Web Services (AWS) through its Open Data Sponsorship Program (registry.opendata.aws/pdb-3d-structural-biology-data/, accessed on 28 August 2022).

Global PDB archive data downloads in 2021 reached a record high of 2,364,150,827 structure data files, which represents an ~80% increase vs. the previous record of 1,323,213,832 set in 2020. Approximately 70% of global structure data file downloads in 2021 originated from the FTP archive. The remainder were accessed by users of wwPDB member web portals.

2.8. All Three Kingdoms of Life Are Represented in the PDB Archive

As of mid-2022, MX, 3DEM, NMR, EC, and microED had been used collectively to determine >190,000 3D biostructures housed in the PDB archive, which encompasses proteins from organisms representing all living kingdoms (Figure 7). Archaebacterial proteins were the least numerous (totaling 5664 structures), followed by bacteria (65,967 structures). PDB holdings of eukaryotic protein structures exceeded 105,000, with more than half being human in origin. There is limited PDB coverage across the so-called model organisms, with mouse proteins being most numerous at >8000 structures.

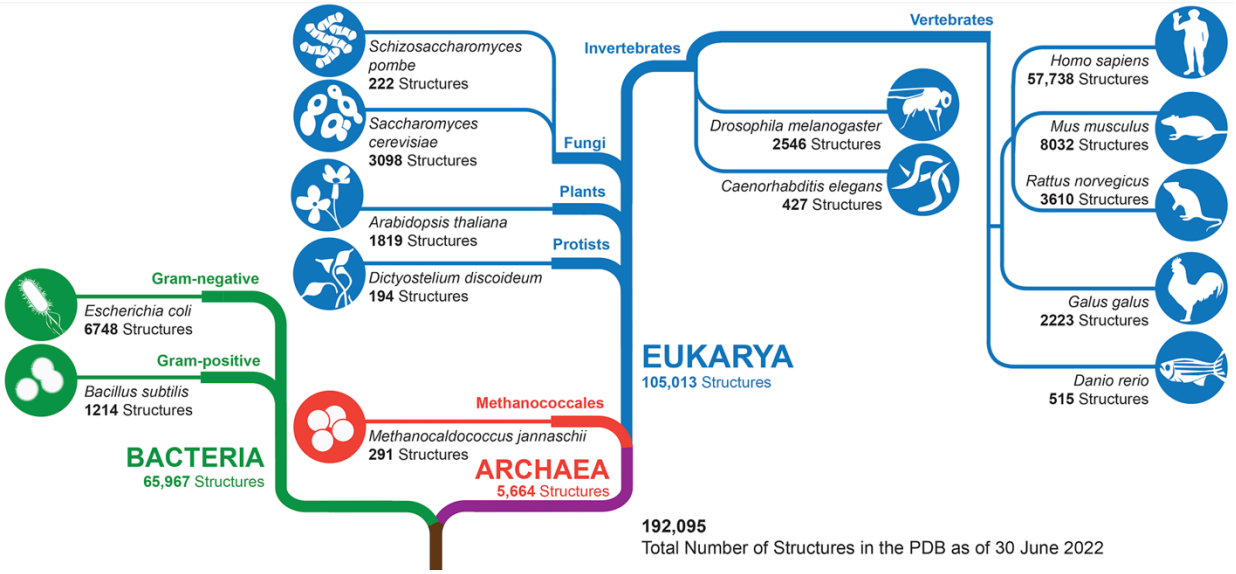

Figure 7. Phylogenetic Tree showing PDB holdings (as of mid-2022). Within each of the three branches, PDB structure totals are provided for selected organisms. N.B.: The PDB also houses 3D structures that solely contain nucleic acids (DNA, RNA, DNA-RNA hybrids, etc.) and/or viral proteins or human-designed proteins, which collectively accounted for ~8% of archival holdings as of mid-2022.

2.9. PDB Data Delivery/Usage Metrics

Most RCSB PDB users access the archive through our RCSB.org research-focused web portal, which makes PDB data available at no cost with no limitations on usage via the Creative Commons CC0 1.0 Universal license (creativecommons.org/publicdomain/zero/1.0/, accessed on 28 August 2022). In 2021, 6,845,233 unique internet protocol (IP) addresses from more than 240 countries and territories recognized by the United Nations (Figure 8A) were used to access RCSB.org (exceeding the 2020 pandemic lock-down record of 6,677,853). Figure 8B ranks RCSB.org utilization for the top ten user countries for 2019–2021. Not surprisingly, the US–RCSB PDB's host country–has the largest percentage of users, followed by the world's two most populous nations, India and the People's Republic of China.

We estimate that ~99% of PDB data consumers are not experts in structural biology. Their research interests are extremely broad, encompassing fundamental biology, biomedicine, energy sciences, bioengineering, and biotechnology [115,116]. Beyond the natural, physical, mathematical, and engineering sciences, there is also use of PDB data by social scientists (e.g., economists, [117,118]).

The RCSB.org web portal provides added value to PDB users that goes well beyond the content of the archive itself. On a weekly basis, RCSB PDB integrates PDB data with information from ~50 trusted external resources (Table 1). Integrating individual PDB structures with information from trusted external resources ensures that the RCSB.org web portal operates as a "living data resource." Scholarly journal articles describing PDB structures are static documents, reflecting what was known about the biomolecule(s) at the time of publication. Thereafter, it is not uncommon for new biological or biochemical functions of a macromolecule to come to light, or new disease-causing mutations to be identified. Such new findings are integrated with PDB data every week, thereby ensuring that RCSB.org users have access to the most current information pertaining to every 3D biostructure in the public domain.

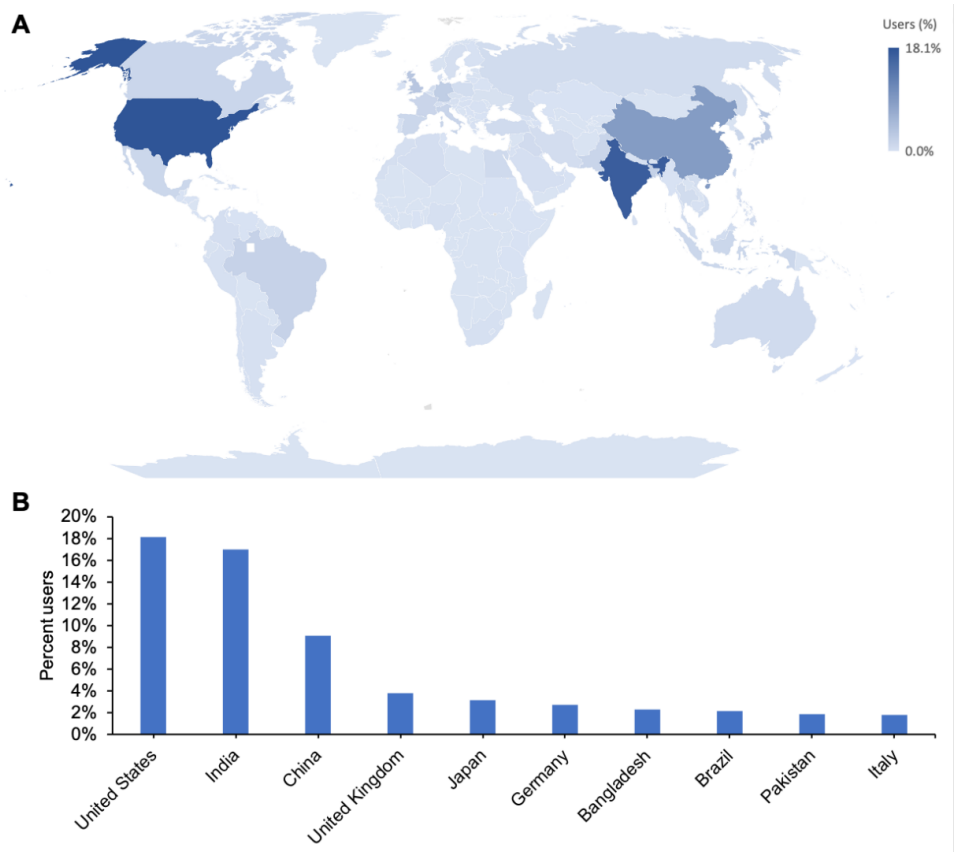

Figure 8. (**A**). Geographic distribution of RCSB.org users by country. (**B**). Top 10 countries with the highest percentage of users from 2019–2021. Data from Google Analytics.

Table 1. Trusted external resources/data content integrated weekly with PDB archival data by RCSB PDB from rcsb.org/docs/general-help/data-from-external-resources-integrated-into-rcsb-pdb (accessed on 28 August 2022). (N.B.: In response to community input, RCSB PDB continues to integrate new external data resources.).

Resource	Description
AlphaFold DB [73,74]	Computed Structure Models by AlphaFold2
ATC	Anatomical Therapeutic Chemical (ATC) Classification System from World Health Organization
Binding MOAD [119]	Binding affinities
BindingDB [120]	Binding affinities
BMRB [13]	BMRB-to-PDB mappings
CATH [121]	Protein structure classification
CCDC [122]	Cambridge Structural Database (CSD)
ChEBI [123]	Chemical entities of biological interest
ChEMBL [124]	Manually curated database of bioactive molecules with drug-like properties
DrugBank [125]	Drug and drug target data
ECOD [126]	Evolutionary Classification of Protein Domains
EMDB [11]	3DEM density maps and associated metadata
ExplorEnz [127]	IUBMB Enzyme nomenclature and classification

Table 1. *Cont.*

Resource	Description
Gencode [128]	Gene structure data
Gene Ontology [129]	Gene structure data
Genotype-Tissue Expression (GTEx) [130]	Tissue-specific gene expression data
GlyCosmos [131]	Web portal integrating the glycosciences with the life sciences
GlyGen [132]	Data integration and dissemination resource for carbohydrates and glycoconjugates
GlyTouCan [133]	Glycan structure repository
Human Gene Nomenclature Committee (genenames.org, accessed on 28 August 2022)	Human gene name nomenclature and genomic information
IMGT [134]	International ImMunoGeneTics information system
Immune Epitope Database [135]	Antibody and T cell epitopes
International Mouse Phenotyping Consortium (mousephenotype.org, accessed on 28 August 2022)	Mouse gene phenotype data
InterPro [136]	Classification of Protein Families
MemProtMD [137]	Database of Membrane Proteins Embedded in Lipid Bilayers
ModelArchive (modelarchive.org accessed on 28 August 2022)	Computed Structure Models (e.g., by RoseTTAFold)
Mpstruc [138]	Classification of transmembrane protein structures
NCBI Gene [139]	Gene info, reference sequences, etc.
NCBI Taxonomy [139]	Organism classification
NDB [140]	Experimentally determined nucleic acids and complex assemblies
OPM [141]	Orientations of Proteins in Membranes database; Classification of transmembrane protein structures and membrane segments
PDBbind-CN [142]	Binding affinities
PDBflex [143]	Protein structure flexibility
PDBTM [144]	Protein Data Bank of Transmembrane Proteins
Pharos [145]	Drug targets and diseases
ProteinDiffraction.org (proteindiffraction.org, accessed on 28 August 2022)	Diffraction images
PubChem [146]	Chemical information
PubMed [139]	Citation information
PubMedCentral [139]	Open access literature
RECOORD [147]	NMR structure ensembles
RESID [148]	Protein modifications
SAbDab [149]	The Structural Antibody Database
Thera-SAbDab [150]	Therapeutic Structural Antibody Database
SBGrid [151]	Structural Biology Data Grid/diffraction images
SCOP [152]	Structural Classification of Proteins
SCOPe [153]	Structural Classification of Proteins—extended
SIFTS [154]	Structure, function, taxonomy, sequence
UniProt [155]	Protein sequences and annotations

PDB data utilization worldwide is also mediated by third parties that repackage and reuse the archival information. While the RCSB PDB is unable to assess utilization of the archive via third parties, review of the Nucleic Acids Research Online Molecular Biology Database Collection [156], which comprises databases from *Nucleic Acids Research* annual Database Issues, identified 460 external data resources that distribute repackaged PDB data (Supplementary Materials Table S1). Additional utilization of PDB data occurs within all major biopharmaceutical companies and many smaller biotechnology companies that maintain copies of the archive inside company firewalls. They frequently use PDB data alongside proprietary MX structures determined by company structural biologists or their contractors. Most, if not all, global biopharmaceutical companies (e.g., Pfizer, Novartis, Eli Lilly and Company) rely on structure-guided drug discovery of small-molecule, orally bioavailable therapeutic agents, which typically begins with scanning of PDB archival holdings for a public domain structure of the target protein to begin the discovery process [25,157,158]. They also make use of PDB structures when engineering new biologic agents (monoclonal antibodies, cytokines, etc.) for use as injectables [159].

Literature searching provides another means of assessing utilization and impact of PDB data. As of mid-2022, 162,262 (~84%) of PDB structures are described in 75,497 unique primary publications, the vast majority of which appeared in peer-reviewed journals. Citation analyses carried out using EuropePMC revealed that in 2021, the PDB was mentioned by name in 23,030 publications. It further documented that PDB IDs were mentioned in 585,903 publications in 2021. An RCSB PDB study published in 2018 [160] documented that citations of PDB data spanned the sciences, literally from Agriculture to Zoology. Not surprisingly, nearly 90% of published PDB structures analyzed in 2018 were cited by journals in the area of Biochemistry and Molecular Biology. High impact within other areas of biomedicine (Cell Biology, Pharmacology and Pharmacy, Microbiology, Genetics and Heredity) was, as expected, also documented. Further RCSB PDB analyses on this topic highlighted PDB structure publications that were frequently cited in scientific journals focused on Materials Science, Physics, Computer Science, Chemistry, Engineering, and Mathematics [116].

Searching of the patent literature in August 2022 also documented substantial impact of PDB data. Directed searches for PDB mentions using the US Patent and Trademark Office website (uspto.gov, accessed on 28 August 2022) identified nearly 19,000 in-process patent applications and ~10,000 issued US patents (vs. ~20,000 in process applications and ~6500 issued patents in June 2017 [160]). Analogous searches of global patent literature using PatSeer (patseer.com) documented ~90,000 issued patents and patent applications in process worldwide that include PDB mentions (vs. ~50,000 in June 2017 [160]).

Finally, RCSB PDB also operates a second web portal focused on outreach and education (PDB101.RCSB.org, with PDB-101 denoting an introductory course) [161]. PDB-101 was launched in 2011 to support PDB archive exploration and training by university faculty, postdoctoral researchers, undergraduate and graduate students, school teachers and their pupils, and the general public. It was established to help train the next generation of PDB users and promote structural biology and protein science to non-experts. Regularly published features include the highly popular *Molecule of the Month* series [162], 3D biostructure-related activities, molecular animations and videos, and educational curricula, many of which are organized around a public health topic [163]. The *Guide to Understanding PDB Data* covers key topics, including file format information and explanations of the types of data included with a PDB entry. Materials are organized into various categories (Health and Disease, Molecules of Life, Biotech and Nanotech, and Structures and Structure Determination) and searchable by keyword (e.g., cancer, checkpoint therapy, antibody). Although it is not as intensively accessed as our RCSB.org research-focused web portal, there is substantial utilization of PDB101.RCSB.org by users from around the world (Figure 9).

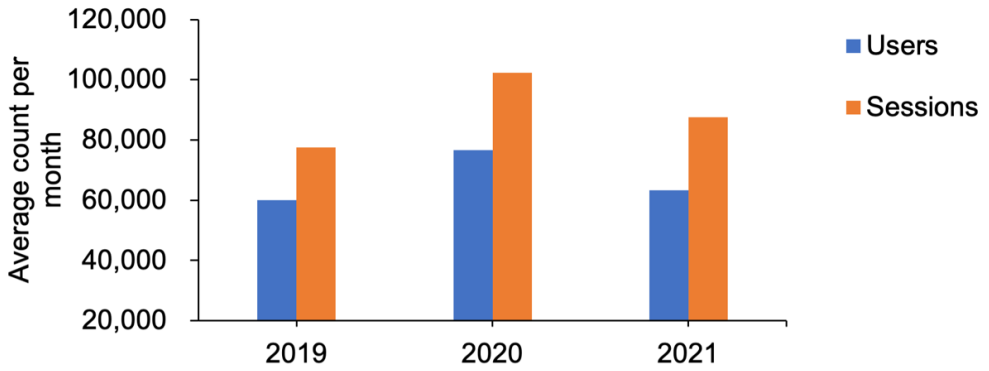

Figure 9. Average monthly usage of PDB-101 (PDB101.RCSB.org, accessed on 28 August 2022) from 2019–2021. Data from Google Analytics.

2.10. Impact of PDB Data on Computational Structure Modeling

Use of PDB data to compute 3D structure information for other proteins is well-established. For many years, publicly available computational services (e.g., Modeller/ModBase [164–166] and ProMod3/SWISS-MODEL, [167,168] and Rosetta [169]) used comparative or homology modeling to predict protein structures. This approach depends on finding an experimentally-determined protein structure in the PDB with an amino acid sequence similar to that of the target protein to use as a modeling template or scaffold. Homology modeling typically succeeds when a structural template with >40% sequence identity is available. Like MR, homology modeling is often useful because of the parsimony of macromolecular evolution.

As the PDB archive grew, template-free computational structure modeling became possible for very small globular proteins. Continuous advances in both homology modeling and template-free protein structure prediction were fostered by two community-led blind challenges (i.e., CASP [170], and the weekly Continuous Automated Model EvaluatiOn (or CAMEO) online challenge [113]). Both CASP and CAMEO rely on coordination with structural biologists and the wwPDB to ensure relevant structure data are not publicly released before each challenge concludes.

Google DeepMind emerged as the top performer in the 2020 CASP challenge [170]. Its AlphaFold2 software uses artificial intelligence/machine learning (AI/ML) to predict 3D structures of smaller globular proteins with accuracies comparable to that of low-resolution experimental methods [74]. It was rightly heralded as a major breakthrough in de novo protein structure prediction. Subsequently, the Rosetta team led by David A. Baker (University of Washington/Howard Hughes Medical Institute) released RoseTTAFold [76] and then RoseTTAFold2, which also use AI/ML methods to generate computed structure models (CSMs) of proteins with reported accuracies comparable to that of AlphaFold2. Figure 10 contrasts experimental structure determination with computed structure model calculation. At the time of writing, CSMs for nearly every protein sequence represented in UniProt [155] generated by DeepMind using AlphaFold2 were publicly available from AlphaFold DB [73–75]. Some of the CSMs generated by computational biologists operating independently of DeepMind (using RoseTTAFold, AlphaFold2, etc.) are available from the open access ModelArchive (modelarchive.org, accessed on 28 August 2022).

Of particular importance when evaluating CSMs for use in research are pLDDT (predicted local distance difference test) scores or confidence estimates generated by AlphaFold2 [74,171]. pLDDT scores (scaled between 0 and 100) denote polypeptide chain segments as very high confidence (pLDDT $\geq$ 90), confident (90 > pLDDT $\geq$ 70), low confidence (70 > pLDDT $\geq$ 50), and very low confidence (pLDDT < 50). We do not yet know how much enhanced AI/ML methods will improve prediction accuracy and expand the scope thereof to larger, multidomain proteins, but history shows us that continued growth of the PDB should only help in this regard.

It is no exaggeration to say that neither AlphaFold2 nor RoseTTAFold2 would exist today without open access to complete, rigorously validated, expertly biocurated 3D

biostructure data from the PDB [172]. Looking ahead, use of AI/ML methods for accurate prediction of structures of macromolecular assemblies and, perhaps even more challenging, transient intermolecular interactions that underpin complex regulatory processes in biology will depend critically on continued growth in the number of 3DEM structures of large molecular machines deposited to the PDB. Successful application of AI/ML methods for predicting small-molecule ligand binding to protein targets may not be possible in the near term given current PDB data deposition trends. The number of co-crystal structures of small molecules binding to proteins in the PDB is dwarfed by 3D structure data collectively held as trade secrets across the biopharmaceutical industry. Contributions of significantly more co-crystal structure data from industry would almost certainly fuel advances in prediction of small-molecule binding to proteins. With sufficient data placed in the public domain, we can reasonably expect that AI/ML methods would accelerate drug discovery and development efforts in both academe and industry for the greater good [172].

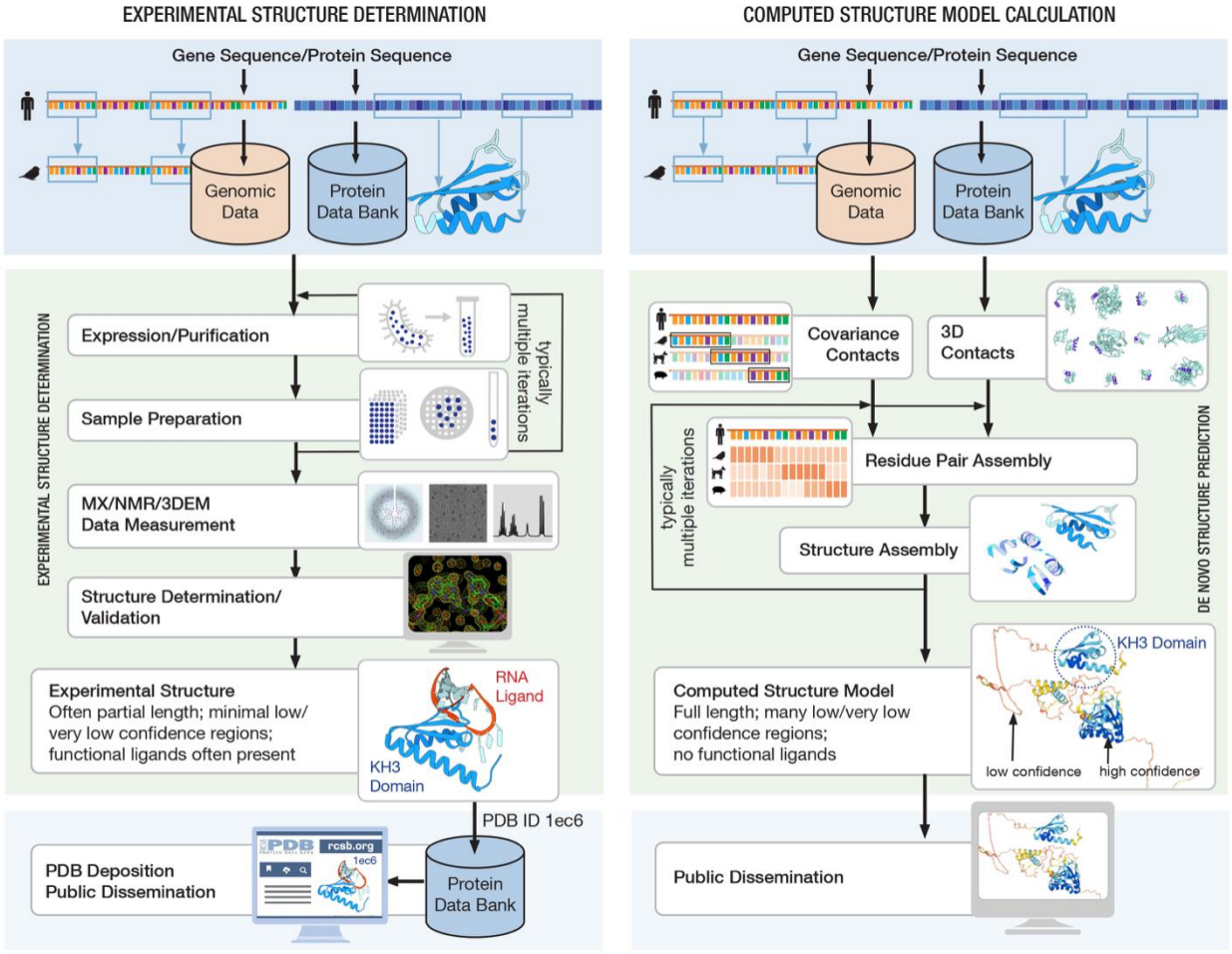

Figure 10. Experimental approaches for determination of protein structures and computational methods for predicting structures both rely on open access to genomic and 3D structure data. Here, methods for determining the structure of the RNA-binding protein Nova-2 are shown. The MX structure (**left**) was determined for an isolated domain of the protein bound to its RNA target. The computed structure (**right**) includes the entire polypeptide chain, which is predicted to include three well-folded domains (blue/cyan) connected by apparently unstructured linkers (yellow/orange). Image adapted from *New England Journal of Medicine*, Stephen K. Burley, Wadih Arap, Renata Pasqualini, Predicting Proteome-Scale Protein Structure with Artificial Intelligence, 385, 2191–2194 [173].Copyright © 2022 Massachusetts Medical Society. Reprinted with permission.

2.11. Future Directions

The futures of structural biologists and the PDB appear even brighter, contrary to post-AlphaFold2 rumors to the effect that experimental structural biology is on the verge of precipitous decline. Depositions of structures to the PDB in 2022 are on track to exceed those in all previous years. Experimentally determined 3D biostructures are highly prized accomplishments. Medium-to-high resolution experimental structures (e.g., MX structures better than 3.5 Å resolution) are more accurate than CSMs [174]. Moreover, they frequently contain bound small-molecule ligands of biological or biomedical importance. They may also include more than one macromolecule, providing information regarding homo- and hetero-meric assemblies that underpin the workings of complex molecular machines.

CSMs generated with AI/ML methods are of considerable interest to experimental structural biologists. Many are taking a "glass half full" approach to this information. They often rely on CSMs of large multi-domain eukaryotic proteins for designing protein expression constructs by excluding low confidence and very low confidence regions when generating truncations suitable for MX, NMR, or 3DEM studies. (N.B.: CSMs are not eligible for archiving in the PDB, because they do not involve measurements from a sample of the biological macromolecule for which the structure is determined.)

The future of experimental structural biology is also looking bright. Researchers are tackling ever larger and more complex macromolecular machines using so-called integrative or hybrid methods that combine experimental measurements from more than one biophysical technique. Anticipating this trend, a wwPDB Integrative/Hybrid Methods (IHM) Task Force was assembled to make recommendations regarding data archiving and structure validation [175,176]. As an interim measure, the wwPDB established PDB-Dev as a standalone prototype system [177–179] for archiving and publicly disseminating integrative structures and associated data. Integrative structure determination entails making measurements using complementary experimental methods (e.g., 3DEM and chemical cross-linking) and converting the results into spatial restraints that are applied to with known starting structures of molecular components to determine the structures of complex macromolecular assemblies.

The PDB-Dev software system supports data collection, processing, curation, validation, archiving, and distribution of integrative biostructures. It is underpinned by ModelCIF (github.com/ihmwg/ModelCIF, accessed on September 29 2022), an expanded set of data standards based on the PDBx/mmCIF data standard (above) for representing integrative structures and associated experimental restraints; a software library that supports the new data standards; a data harvesting system for collecting heterogeneous data from diverse experimental techniques, methods for curating, validating and visualizing integrative structures; and web services for distributing archived data. The PDB-Dev prototype system has allowed structural biologists to make their integrative structures publicly available, including but by no means limited to those involved in transport of proteins and nucleic acids across the nuclear envelope (nuclear pore complex [180]), regulation of gene expression (expressome complex [181]), cellular vesicle trafficking (exocyst complex [182]), and regulation of genomic architecture (BAF complex [183]). Importantly, the PDB-Dev data standard was designed to interoperate with PDBx/mmCIF and the PDB, so that integrative structures and related metadata can eventually be archived in the PDB.

In parallel with building PDB-Dev, wwPDB partners are working to establish a federated network of interoperating structural biology data resources, as recommended by the IHM Task Force [176]. This effort involves collaboration with other experimental data repositories (e.g., SASBDB [184] and PRIDE [185]). Tools are being created to support automated data exchange between PDB-Dev and these and other biodata repositories (e.g., BioImage Archive, www.ebi.ac.uk/bioimage-archive, accessed on 28 August 2022 [186]). The overarching goal of the wwPDB partnership is to foster federation of structural biology data resources across length scales ranging from atoms to individual proteins to macromolecular machines to organelles to cells and eventually tissues to maximize the impact that

atomic level 3D biostructures will have on research and education across basic and applied biological, biomedical and energy sciences.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/biom12101425/s1, Table S1 Cumulative list of external data resources identified as repackaging and redistributing PDB data.

Author Contributions: Conceptualization, S.K.B. and H.M.B.; software, J.M.D., Z.F., R.L., E.P., D.W.P., Y.R., C.S., M.V. and J.D.W.; writing—original draft preparation, S.K.B.; writing—review and editing, S.K.B., H.M.B., J.M.D., Z.F., J.W.F., B.P.H., R.L., E.P., D.W.P., Y.R., A.S., M.S., C.S., B.V., M.V., J.Y.Y. and C.Z.; visualization, C.S., J.W.F., B.P.H., D.W.P., C.S. and M.V.; supervision, J.M.D., R.L., Y.R., J.Y.Y. and C.Z.; funding acquisition, S.K.B., A.S. and B.V. All authors have read and agreed to the published version of the manuscript.

Funding: RCSB PDB core operations are jointly funded by the National Science Foundation (NSF; DBI-1832184, PI: S.K.B.), the US Department of Energy (DE-SC0019749, PI: S.K.B.), and the National Cancer Institute, the National Institute of Allergy and Infectious Diseases, and the National Institute of General Medical Sciences of the National Institutes of Health (R01GM133198, PI: S.K.B.). Other funding awards to RCSB PDB by the NSF and to PDBe by the UK Biotechnology and Biological Research Council are jointly supporting development of a Next Generation PDB archive (DBI-2019297, PI: S.K.B.; BB/V004247/1, PI: Sameer Velankar) and new Mol* features (DBI-2129634, PI: S.K.B.; BB/W017970/1, PI: Sameer Velankar). PDB-dev development supported NSF awards DBI-1756248, DBI-2112966 (PI: B.V.) and DBI-1756250, DBI-2112967 (A.S.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: PDB data are made freely available by the wwPDB (wwPDB.org, accessed on 28 August 2022).

Acknowledgments: The authors thank the tens of thousands of structural biologists worldwide who deposited structures to the PDB since 1971 and the many millions of researchers, educators, and students around the world who consume PDB data. We also gratefully acknowledge contributions to the success of the PDB archive made by past members of RCSB PDB and our Worldwide Protein Data Bank partners (PDBe, PDBj, EMDB, and BMRB).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Protein Data Bank. Crystallography: Protein Data Bank. *Nat. New Biol.* **1971**, *233*, 223. [CrossRef]
2. Berman, H.M.; Henrick, K.; Nakamura, H. Announcing the worldwide Protein Data Bank. *Nat. Struct. Biol.* **2003**, *10*, 980. [CrossRef]
3. wwPDB consortium. Protein Data Bank: The single global archive for 3D macromolecular structure data. *Nucleic Acids Res.* **2019**, *47*, D520–D528. [CrossRef]
4. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. [CrossRef]
5. Burley, S.K.; Bhikadiya, C.; Bi, C.; Bittrich, S.; Chen, L.; Crichlow, G.; Christie, C.H.; Dalenberg, K.; Costanzo, L.D.; Duarte, J.M.; et al. RCSB Protein Data Bank: Powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering, and energy sciences. *Nucleic Acid Res.* **2021**, *49*, D437–D451. [CrossRef] [PubMed]
6. Burley, S.K.; Bhikadiya, C.; Bi, C.; Bittrich, S.; Chen, L.; Crichlow, G.V.; Duarte, J.M.; Dutta, S.; Fayazi, M.; Feng, Z.; et al. RCSB Protein Data Bank: Celebrating 50 years of the PDB with new tools for understanding and visualizing biological macromolecules in 3D. *Protein Sci.* **2022**, *31*, 187–208. [CrossRef]
7. Burley, S.K.; Bhikadiya, C.; Bi, C.; Bittrich, S.; Chao, H.; Chen, L.; Craig, P.A.; Crichlow, G.V.; Dalenberg, K.; Duarte, J.M.; et al. RCSB Protein Data Bank: Tools for visualizing and understanding biological macromolecules in 3D. *Protein Sci.* 2022; *submitted*.
8. Armstrong, D.R.; Berrisford, J.M.; Conroy, M.J.; Gutmanas, A.; Anyango, S.; Choudhary, P.; Clark, A.R.; Dana, J.M.; Deshpande, M.; Dunlop, R.; et al. PDBe: Improved findability of macromolecular structure data in the PDB. *Nucleic Acids Res.* **2020**, *48*, D335–D343. [CrossRef]

9. Bekker, G.J.; Yokochi, M.; Suzuki, H.; Ikegawa, Y.; Iwata, T.; Kudou, T.; Yura, K.; Fujiwara, T.; Kawabata, T.; Kurisu, G. Protein Data Bank Japan: Celebrating our 20th anniversary during a global pandemic as the Asian hub of three dimensional macromolecular structural data. *Protein Sci.* **2022**, *31*, 173–186. [CrossRef] [PubMed]
10. Tagari, M.; Newman, R.; Chagoyen, M.; Carazo, J.M.; Henrick, K. New electron microscopy database and deposition system. *Trends Biochem. Sci.* **2002**, *27*, 589. [CrossRef]
11. Lawson, C.L.; Patwardhan, A.; Baker, M.L.; Hryc, C.; Garcia, E.S.; Hudson, B.P.; Lagerstedt, I.; Ludtke, S.J.; Pintilie, G.; Sala, R.; et al. EMDataBank unified data resource for 3DEM. *Nucleic Acids Res.* **2016**, *44*, D396–D403. [CrossRef] [PubMed]
12. Ulrich, E.L.; Akutsu, H.; Doreleijers, J.F.; Harano, Y.; Ioannidis, Y.E.; Lin, J.; Livny, M.; Mading, S.; Maziuk, D.; Miller, Z.; et al. BioMagResBank. *Nucleic Acids Res.* **2008**, *36*, D402–D408. [CrossRef]
13. Romero, P.R.; Kobayashi, N.; Wedell, J.R.; Baskaran, K.; Iwata, T.; Yokochi, M.; Maziuk, D.; Yao, H.; Fujiwara, T.; Kurusu, G.; et al. BioMagResBank (BMRB) as a Resource for Structural Biology. *Methods Mol. Biol.* **2020**, *2112*, 187–218. [CrossRef] [PubMed]
14. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **2016**, *3*, 1–9. [CrossRef] [PubMed]
15. van der Aalst, W.M.P.; Bichler, M.; Heinzl, A. Responsible Data Science. *Bus. Inf. Syst. Eng.* **2017**, *59*, 311–313. [CrossRef]
16. Moore, P.B. The PDB and the ribosome. *J. Biol. Chem.* **2021**, *296*, 100561. [CrossRef] [PubMed]
17. Johnson, J.E.; Olson, A.J. Icosahedral virus structures and the protein data bank. *J. Biol. Chem.* **2021**, *296*, 100554. [CrossRef]
18. Neidle, S. Beyond the double helix: DNA structural diversity and the PDB. *J. Biol. Chem.* **2021**, *296*, 100553. [CrossRef] [PubMed]
19. Westhof, E.; Leontis, N.B. An RNA-centric historical narrative around the Protein Data Bank. *J. Biol. Chem.* **2021**, *296*, 100555. [CrossRef] [PubMed]
20. Prestegard, J.H. A perspective on the PDB's impact on the field of glycobiology. *J. Biol. Chem.* **2021**, *296*, 100556. [CrossRef] [PubMed]
21. Li, F.; Egea, P.F.; Vecchio, A.J.; Asial, I.; Gupta, M.; Paulino, J.; Bajaj, R.; Dickinson, M.S.; Ferguson-Miller, S.; Monk, B.C.; et al. Highlighting membrane protein structure and function: A celebration of the Protein Data Bank. *J. Biol. Chem.* **2021**, *296*, 100557. [CrossRef]
22. Chiu, W.; Schmid, M.F.; Pintilie, G.D.; Lawson, C.L. Evolution of standardization and dissemination of cryo-EM structures and data jointly by the community, PDB, and EMDB. *J. Biol. Chem.* **2021**, *296*, 100560. [CrossRef] [PubMed]
23. Pan, X.; Kortemme, T. Recent advances in de novo protein design: Principles, methods, and applications. *J. Biol. Chem.* **2021**, *296*, 100558. [CrossRef]
24. Murray, D.; Petrey, D.; Honig, B. Integrating 3D structural information into systems biology. *J. Biol. Chem.* **2021**, *296*, 100562. [CrossRef]
25. Burley, S.K. Impact of structural biologists and the Protein Data Bank on small-molecule drug discovery and development. *J. Biol. Chem.* **2021**, *296*, 100559. [CrossRef]
26. Taylor, S.S.; Wu, J.; Bruystens, J.G.H.; Del Rio, J.C.; Lu, T.W.; Kornev, A.P.; Ten Eyck, L.F. From structure to the dynamic regulation of a molecular switch: A journey over 3 decades. *J. Biol. Chem.* **2021**, *296*, 100746. [CrossRef]
27. Wolberger, C. How structural biology transformed studies of transcription regulation. *J. Biol. Chem.* **2021**, *296*, 100741. [CrossRef]
28. Wilson, I.A.; Stanfield, R.L. 50 Years of structural immunology. *J. Biol. Chem.* **2021**, *296*, 100745. [CrossRef]
29. Saibil, H.R. The PDB and protein homeostasis: From chaperones to degradation and disaggregase machines. *J. Biol. Chem.* **2021**, *296*, 100744. [CrossRef]
30. Michalska, K.; Joachimiak, A. Structural genomics and the Protein Data Bank. *J. Biol. Chem.* **2021**, *296*, 100747. [CrossRef]
31. Sali, A. From integrative structural biology to cell biology. *J. Biol. Chem.* **2021**, *296*, 100743. [CrossRef]
32. Miller, M.D.; Phillips, G.N., Jr. Moving beyond static snapshots: Protein dynamics and the Protein Data Bank. *J. Biol. Chem.* **2021**, *296*, 100749. [CrossRef]
33. Richardson, J.S.; Richardson, D.C.; Goodsell, D.S. Seeing the PDB. *J. Biol. Chem.* **2021**, *296*, 100742. [CrossRef] [PubMed]
34. Cohen, A.E. A new era of synchrotron-enabled macromolecular crystallography. *Nat. Methods* **2021**, *18*, 433–434. [CrossRef]
35. Kern, D. From structure to mechanism: Skiing the energy landscape. *Nat. Methods* **2021**, *18*, 435–436. [CrossRef] [PubMed]
36. Vinothkumar, K.R. Expanding capabilities and infrastructure for cryo-EM. *Nat. Methods* **2021**, *18*, 437–438. [CrossRef] [PubMed]
37. Das, R. RNA structure: A renaissance begins? *Nat. Methods* **2021**, *18*, 439. [CrossRef]
38. Li, X. Cryo-electron tomography: Observing the cell at the atomic level. *Nat. Methods* **2021**, *18*, 440–441. [CrossRef] [PubMed]
39. Wozny, M.R.; Kukulski, W. Molecular visualization of cellular complexity. *Nat. Methods* **2021**, *18*, 442–443. [CrossRef]
40. Narykov, O.; Srinivasan, S.; Korkin, D. Computational protein modeling and the next viral pandemic. *Nat. Methods* **2021**, *18*, 444–445. [CrossRef] [PubMed]
41. Luthey-Schulten, Z. Integrating experiments, theory and simulations into whole-cell models. *Nat. Methods* **2021**, *18*, 446–447. [CrossRef]
42. Bonvin, A. 50 years of PDB: A catalyst in structural biology. *Nat. Methods* **2021**, *18*, 448–449. [CrossRef]
43. Bourne, P.E.; Addess, K.J.; Bluhm, W.F.; Chen, L.; Deshpande, N.; Feng, Z.; Fleri, W.; Green, R.; Merino-Ott, J.C.; Townsend-Merino, W.; et al. The distribution and query systems of the RCSB Protein Data Bank. *Nucleic Acids Res.* **2004**, *32*, D223–D225. [CrossRef] [PubMed]

44. Young, J.Y.; Westbrook, J.D.; Feng, Z.; Sala, R.; Peisach, E.; Oldfield, T.J.; Sen, S.; Gutmanas, A.; Armstrong, D.R.; Berrisford, J.M.; et al. OneDep: Unified wwPDB System for Deposition, Biocuration, and Validation of Macromolecular Structures in the PDB Archive. *Structure* **2017**, *25*, 536–545. [CrossRef] [PubMed]
45. Gore, S.; Sanz Garcia, E.; Hendrickx, P.M.S.; Gutmanas, A.; Westbrook, J.D.; Yang, H.; Feng, Z.; Baskaran, K.; Berrisford, J.M.; Hudson, B.P.; et al. Validation of Structures in the Protein Data Bank. *Structure* **2017**, *25*, 1916–1927. [CrossRef]
46. Feng, Z.; Westbrook, J.D.; Sala, R.; Smart, O.S.; Bricogne, G.; Matsubara, M.; Yamada, I.; Tsuchiya, S.; Aoki-Kinoshita, K.F.; Hoch, J.C.; et al. Enhanced validation of small-molecule ligands and carbohydrates in the protein databank. *Structure* **2021**, *29*, 393–400.e391. [CrossRef]
47. Young, J.Y.; Westbrook, J.D.; Feng, Z.; Peisach, E.; Persikova, I.; Sala, R.; Sen, S.; Berrisford, J.M.; Swaminathan, G.J.; Oldfield, T.J.; et al. Worldwide Protein Data Bank biocuration supporting open access to high-quality 3D structural biology data. *Database* **2018**, *2018*, bay002. [CrossRef]
48. Kendrew, J.C.; Dickerson, R.E.; Strandberg, B.E.; Hart, R.G.; Davies, D.R.; Phillips, D.C.; Shore, V.C. Structure of myoglobin: A three-dimensional Fourier synthesis at 2 A. resolution. *Nature* **1960**, *185*, 422–427. [CrossRef]
49. Yip, K.M.; Fischer, N.; Paknia, E.; Chari, A.; Stark, H. Atomic-resolution protein structure determination by cryo-EM. *Nature* **2020**, *587*, 157–161. [CrossRef] [PubMed]
50. Shao, C.; Westbrook, J.D.; Lu, C.; Bhikadiya, C.; Peisach, E.; Young, J.Y.; Duarte, J.M.; Lowe, R.; Wang, S.; Rose, Y.; et al. Simplified Quality Assessment for Small-molecule Ligands in the PDB Archive. *Structure* **2022**, *30*, 252–262. [CrossRef] [PubMed]
51. Blundell, T.L.; Johnson, L.N. *Protein Crystallography*; Academic Press: New York, NY, USA, 1976.
52. Rossmann, M.G. The molecular replacement method. *Acta Cryst. A* **1990**, *46 Pt 2*, 73–82. [CrossRef]
53. Read, R.J.; Adams, P.D.; Arendall, W.B., 3rd; Brunger, A.T.; Emsley, P.; Joosten, R.P.; Kleywegt, G.J.; Krissinel, E.B.; Lutteke, T.; Otwinowski, Z.; et al. A new generation of crystallographic validation tools for the protein data bank. *Structure* **2011**, *19*, 1395–1412. [CrossRef]
54. Shao, C.; Yang, H.; Westbrook, J.D.; Young, J.Y.; Zardecki, C.; Burley, S.K. Multivariate Analyses of Quality Metrics for Crystal Structures in the PDB Archive. *Structure* **2017**, *25*, 458–468. [CrossRef]
55. Adams, P.D.; Aertgeerts, K.; Bauer, C.; Bell, J.A.; Berman, H.M.; Bhat, T.N.; Blaney, J.M.; Bolton, E.; Bricogne, G.; Brown, D.; et al. Outcome of the First wwPDB/CCDC/D3R Ligand Validation Workshop. *Structure* **2016**, *24*, 502–508. [CrossRef] [PubMed]
56. Shao, C.; Feng, Z.; Westbrook, J.D.; Peisach, E.; Berrisford, J.; Ikegawa, Y.; Kurisu, G.; Velankar, S.; Burley, S.K.; Young, J.Y. Modernized Uniform Representation of Carbohydrate Molecules in the Protein Data Bank. *Glycobiology* **2021**, *31*, 1204–1218. [CrossRef]
57. Barends, T.R.M.; Stauch, B.; Cherezov, V.; Schlichting, I. Serial femtosecond crystallography. *Nat. Rev. Methods Prim.* **2022**, *2*, 59. [CrossRef]
58. Pearson, A.R.; Mehrabi, P. Serial synchrotron crystallography for time-resolved structural biology. *Curr. Opin. Struct. Biol.* **2020**, *65*, 168–174. [CrossRef]
59. Schmidt, M. Macromolecular movies, storybooks written by nature. *Biophys. Rev.* **2021**, *13*, 1191–1197. [CrossRef] [PubMed]
60. Olmos, J.L., Jr.; Pandey, S.; Martin-Garcia, J.M.; Calvey, G.; Katz, A.; Knoska, J.; Kupitz, C.; Hunter, M.S.; Liang, M.; Oberthuer, D.; et al. Enzyme intermediates captured "on the fly" by mix-and-inject serial crystallography. *BMC Biol.* **2018**, *16*, 59. [CrossRef] [PubMed]
61. Chapman, H.N.; Fromme, P.; Barty, A.; White, T.A.; Kirian, R.A.; Aquila, A.; Hunter, M.S.; Schulz, J.; DePonte, D.P.; Weierstall, U.; et al. Femtosecond X-ray protein nanocrystallography. *Nature* **2011**, *470*, 73–77. [CrossRef] [PubMed]
62. Kuhlbrandt, W. Biochemistry. The resolution revolution. *Science* **2014**, *343*, 1443–1444. [CrossRef] [PubMed]
63. Herzik, M.A., Jr. Cryo-electron microscopy reaches atomic resolution. *Nature* **2020**, *587*, 39–40. [CrossRef]
64. Passmore, L.A.; Russo, C.J. Specimen Preparation for High-Resolution Cryo-EM. *Methods Enzym.* **2016**, *579*, 51–86. [CrossRef]
65. Brilot, A.F.; Chen, J.Z.; Cheng, A.; Pan, J.; Harrison, S.C.; Potter, C.S.; Carragher, B.; Henderson, R.; Grigorieff, N. Beam-induced motion of vitrified specimen on holey carbon film. *J. Struct. Biol.* **2012**, *177*, 630–637. [CrossRef]
66. Li, X.; Mooney, P.; Zheng, S.; Booth, C.R.; Braunfeld, M.B.; Gubbens, S.; Agard, D.A.; Cheng, Y. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* **2013**, *10*, 584–590. [CrossRef]
67. Bai, X.C.; Fernandez, I.S.; McMullan, G.; Scheres, S.H. Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. *eLife* **2013**, *2*, e00461. [CrossRef]
68. Scheres, S.H. A Bayesian view on cryo-EM structure determination. *J. Mol. Biol.* **2012**, *415*, 406–418. [CrossRef]
69. Scheres, S.H.W. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **2012**, *180*, 519–530. [CrossRef]
70. Zhang, P. Advances in cryo-electron tomography and subtomogram averaging and classification. *Curr. Opin. Struct. Biol.* **2019**, *58*, 249–258. [CrossRef]
71. Zanetti, G.; Prinz, S.; Daum, S.; Meister, A.; Schekman, R.; Bacia, K.; Briggs, J.A. The structure of the COPII transport-vesicle coat assembled on membranes. *eLife* **2013**, *2*, e00951. [CrossRef]
72. Ni, T.; Sun, Y.; Seaton-Burn, W.; Al-Hazeem, M.M.J.; Zhu, Y.; Yu, X.; Liu, L.-N.; Zhang, P. Tales of Two α-Carboxysomes: The Structure and Assembly of Cargo Rubisco. *bioRxiv* **2022**. [CrossRef]

73. Varadi, M.; Anyango, S.; Deshpande, M.; Nair, S.; Natassia, C.; Yordanova, G.; Yuan, D.; Stroe, O.; Wood, G.; Laydon, A.; et al. AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **2022**, *50*, D439–D444. [CrossRef]
74. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Zidek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [CrossRef]
75. Senior, A.W.; Evans, R.; Jumper, J.; Kirkpatrick, J.; Sifre, L.; Green, T.; Qin, C.; Zidek, A.; Nelson, A.W.R.; Bridgland, A.; et al. Improved protein structure prediction using potentials from deep learning. *Nature* **2020**, *577*, 706–710. [CrossRef]
76. Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G.R.; Wang, J.; Cong, Q.; Kinch, L.N.; Schaeffer, R.D.; et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **2021**, *373*, 871–876. [CrossRef]
77. Mosalaganti, S.; Obarska-Kosinska, A.; Siggel, M.; Taniguchi, R.; Turonova, B.; Zimmerli, C.E.; Buczak, K.; Schmidt, F.H.; Margiotta, E.; Mackmull, M.T.; et al. AI-based structure prediction empowers integrative structural analysis of human nuclear pores. *Science* **2022**, *376*, eabm9506. [CrossRef]
78. Turk, M.; Baumeister, W. The promise and the challenges of cryo-electron tomography. *FEBS Lett.* **2020**, *594*, 3243–3261. [CrossRef]
79. Pintilie, G.; Zhang, K.; Su, Z.; Li, S.; Schmid, M.F.; Chiu, W. Measurement of atom resolvability in cryo-EM maps with Q-scores. *Nat. Methods* **2020**, *17*, 328–334. [CrossRef] [PubMed]
80. Wang, Z.; Patwardhan, A.; Kleywegt, G.J. Validation analysis of EMDB entries. *Acta Crystallogr. Sect. D Struct. Biol.* **2022**, *78*, 542–552. [CrossRef]
81. Burley, S.K.; Berman, H.M.; Chiu, W.; Dai, W.; Flatt, J.W.; Hudson, B.P.; Kaelber, J.; Khare, S.; Kulczyk, A.; Lawson, C.L.; et al. Electron Microscopy Holdings of the Protein Data Bank: Impact of the Resolution Revolution and Implications for the Future. *Biophys Rev.* 2022; *submitted*.
82. Williamson, M.P.; Havel, T.F.; Wuthrich, K. Solution conformation of proteinase inhibitor IIA from bull seminal plasma by 1H nuclear magnetic resonance and distance geometry. *J. Mol. Biol.* **1985**, *182*, 295–315. [CrossRef]
83. Kaptein, R.; Zuiderweg, E.R.; Scheek, R.M.; Boelens, R.; van Gunsteren, W.F. A protein structure from nuclear magnetic resonance data. lac repressor headpiece. *J. Mol. Biol.* **1985**, *182*, 179–182. [CrossRef]
84. Driscoll, P.C.; Gronenborn, A.M.; Beress, L.; Clore, G.M. Determination of the three-dimensional solution structure of the antihypertensive and antiviral protein BDS-I from the sea anemone Anemonia sulcata: A study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing. *Biochemistry* **1989**, *28*, 2188–2198. [CrossRef] [PubMed]
85. Kaptein, R.; Boelens, R.; Scheek, R.M.; van Gunsteren, W.F. Protein structures from NMR. *Biochemistry* **1988**, *27*, 5389–5395. [CrossRef] [PubMed]
86. Gronenborn, A.M.; Bax, A.; Wingfield, P.T.; Clore, G.M. A powerful method of sequential proton resonance assignment in proteins using relayed 15N-1H multiple quantum coherence spectroscopy. *FEBS Lett.* **1989**, *243*, 93–98. [CrossRef]
87. Clore, G.M.; Appella, E.; Yamada, M.; Matsushima, K.; Gronenborn, A.M. Three-dimensional structure of interleukin 8 in solution. *Biochemistry* **1990**, *29*, 1689–1696. [CrossRef] [PubMed]
88. Pfander, R.; Neumann, L.; Zweckstetter, M.; Seger, C.; Holak, T.A.; Tampe, R. Structure of the active domain of the herpes simplex virus protein ICP47 in water/sodium dodecyl sulfate solution determined by nuclear magnetic resonance spectroscopy. *Biochemistry* **1999**, *38*, 13692–13698. [CrossRef]
89. Montelione, G.T.; Nilges, M.; Bax, A.; Guntert, P.; Herrmann, T.; Richardson, J.S.; Schwieters, C.D.; Vranken, W.F.; Vuister, G.W.; Wishart, D.S.; et al. Recommendations of the wwPDB NMR Validation Task Force. *Structure* **2013**, *21*, 1563–1570. [CrossRef]
90. Vostrikov, V.V.; Mote, K.R.; Verardi, R.; Veglia, G. Structural dynamics and topology of phosphorylated phospholamban homopentamer reveal its role in the regulation of calcium transport. *Structure* **2013**, *21*, 2119–2130. [CrossRef] [PubMed]
91. Lapinaite, A.; Simon, B.; Skjaerven, L.; Rakwalska-Bange, M.; Gabel, F.; Carlomagno, T. The structure of the box C/D enzyme reveals regulation of RNA methylation. *Nature* **2013**, *502*, 519–523. [CrossRef]
92. Lu, M.; Russell, R.W.; Bryer, A.J.; Quinn, C.M.; Hou, G.; Zhang, H.; Schwieters, C.D.; Perilla, J.R.; Gronenborn, A.M.; Polenova, T. Atomic-resolution structure of HIV-1 capsid tubes by magic-angle spinning NMR. *Nat. Struct. Mol. Biol.* **2020**, *27*, 863–869. [CrossRef]
93. Jehle, S.; Vollmar, B.S.; Bardiaux, B.; Dove, K.K.; Rajagopal, P.; Gonen, T.; Oschkinat, H.; Klevit, R.E. N-terminal domain of alphaB-crystallin provides a conformational switch for multimerization and structural heterogeneity. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 6409–6414. [CrossRef]
94. Gauto, D.F.; Estrozi, L.F.; Schwieters, C.D.; Effantin, G.; Macek, P.; Sounier, R.; Sivertsen, A.C.; Schmidt, E.; Kerfah, R.; Mas, G.; et al. Integrated NMR and cryo-EM atomic-resolution structure determination of a half-megadalton enzyme complex. *Nat. Commun.* **2019**, *10*, 2697. [CrossRef]
95. Puthenveetil, R.; Vinogradova, O. Solution NMR: A powerful tool for structural and functional studies of membrane proteins in reconstituted environments. *J. Biol. Chem.* **2019**, *294*, 15914–15931. [CrossRef]
96. Pervushin, K.V.; Orekhov, V.; Popov, A.I.; Musina, L.; Arseniev, A.S. Three-dimensional structure of (1-71)bacterioopsin solubilized in methanol/chloroform and SDS micelles determined by 15N-1H heteronuclear NMR spectroscopy. *Eur. J. Biochem.* **1994**, *219*, 571–583. [CrossRef]
97. Bondarenko, V.; Wells, M.M.; Chen, Q.; Tillman, T.S.; Singewald, K.; Lawless, M.J.; Caporoso, J.; Brandon, N.; Coleman, J.A.; Saxena, S.; et al. Structures of highly flexible intracellular domain of human alpha7 nicotinic acetylcholine receptor. *Nat. Commun.* **2022**, *13*, 793. [CrossRef]

98. Morag, O.; Sgourakis, N.G.; Baker, D.; Goldbourt, A. The NMR-Rosetta capsid model of M13 bacteriophage reveals a quadrupled hydrophobic packing epitope. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 971–976. [CrossRef]
99. Lange, O.F.; Lakomek, N.A.; Fares, C.; Schroder, G.F.; Walter, K.F.; Becker, S.; Meiler, J.; Grubmuller, H.; Griesinger, C.; de Groot, B.L. Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science* **2008**, *320*, 1471–1475. [CrossRef] [PubMed]
100. Religa, T.L.; Sprangers, R.; Kay, L.E. Dynamic regulation of archaeal proteasome gate opening as studied by TROSY NMR. *Science* **2010**, *328*, 98–102. [CrossRef] [PubMed]
101. Gutmanas, A.; Adams, P.D.; Bardiaux, B.; Berman, H.M.; Case, D.A.; Fogh, R.H.; Guntert, P.; Hendrickx, P.M.; Herrmann, T.; Kleywegt, G.J.; et al. NMR Exchange Format: A unified and open standard for representation of NMR restraint data. *Nat. Struct. Mol. Biol.* **2015**, *22*, 433–434. [CrossRef] [PubMed]
102. DeLisle, C.F.; Malooley, A.L.; Banerjee, I.; Lorieau, J.L. Pro-islet amyloid polypeptide in micelles contains a helical prohormone segment. *FEBS J.* **2020**, *287*, 4440–4457. [CrossRef] [PubMed]
103. Henderson, R.; Baldwin, J.M.; Ceska, T.A.; Zemlin, F.; Beckmann, E.; Downing, K.H. Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *J. Mol. Biol.* **1990**, *213*, 899–929. [CrossRef]
104. Nannenga, B.L.; Gonen, T. The cryo-EM method microcrystal electron diffraction (MicroED). *Nat. Methods* **2019**, *16*, 369–379. [CrossRef]
105. Shi, D.; Nannenga, B.L.; Iadanza, M.G.; Gonen, T. Three-dimensional electron crystallography of protein microcrystals. *eLife* **2013**, *2*, e01345. [CrossRef] [PubMed]
106. Martynowycz, M.W.; Shiriaeva, A.; Ge, X.; Hattne, J.; Nannenga, B.L.; Cherezov, V.; Gonen, T. MicroED structure of the human adenosine receptor determined from a single nanocrystal in LCP. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2106041118. [CrossRef]
107. Nannenga, B.L.; Shi, D.; Hattne, J.; Reyes, F.E.; Gonen, T. Structure of catalase determined by MicroED. *eLife* **2014**, *3*, e03600. [CrossRef]
108. Martynowycz, M.W.; Clabbers, M.T.B.; Hattne, J.; Gonen, T. Ab initio phasing macromolecular structures using electron-counted MicroED data. *Nat. Methods* **2022**, *19*, 724–729. [CrossRef]
109. Westbrook, J.; Bourne, P.E. STAR/mmCIF: An extensive ontology for macromolecular structure and beyond. *Bioinformatics* **2000**, *16*, 159–168. [CrossRef]
110. Fitzgerald, P.M.D.; Westbrook, J.D.; Bourne, P.E.; McMahon, B.; Watenpaugh, K.D.; Berman, H.M. 4.5 Macromolecular dictionary (mmCIF). In *International Tables for Crystallography G. Definition and Exchange of Crystallographic Data*; Hall, S.R., McMahon, B., Eds.; Springer: Dordrecht, The Netherlands, 2005; pp. 295–443.
111. Westbrook, J.D.; Young, J.Y.; Shao, C.; Feng, Z.; Guranovic, V.; Lawson, C.; Vallat, B.; Adams, P.D.; Berrisford, J.M.; Bricogne, G.; et al. PDBx/mmCIF Ecosystem: Foundational semantic tools for structural biology. *J. Mol. Biol.* **2022**, *434*, 167599. [CrossRef]
112. Hall, S.R.; Allen, F.H.; Brown, I.D. The crystallographic information file (CIF): A new standard archive file for crystallography. *Acta Crystallogr. Sect. A Found. Crystallogr.* **1991**, *47*, 655–685. [CrossRef]
113. Haas, J.; Barbato, A.; Behringer, D.; Studer, G.; Roth, S.; Bertoni, M.; Mostaguir, K.; Gumienny, R.; Schwede, T. Continuous Automated Model EvaluatiOn (CAMEO) complementing the critical assessment of structure prediction in CASP12. *Proteins Struct. Funct. Genet.* **2018**, *86* (Suppl. S1), 387–398. [CrossRef]
114. Wagner, J.R.; Churas, C.P.; Liu, S.; Swift, R.V.; Chiu, M.; Shao, C.; Feher, V.A.; Burley, S.K.; Gilson, M.K.; Amaro, R.E. Continuous Evaluation of Ligand Protein Predictions: A Weekly Community Challenge for Drug Docking. *Structure* **2019**, *27*, 1326–1335. [CrossRef]
115. Markosian, C.; Di Costanzo, L.; Sekharan, M.; Shao, C.; Burley, S.K.; Zardecki, C. Analysis of impact metrics for the Protein Data Bank. *Sci. Data* **2018**, *5*, 180212. [CrossRef] [PubMed]
116. Feng, Z.; Verdiguel, N.; Di Costanzo, L.; Goodsell, D.S.; Westbrook, J.D.; Burley, S.K.; Zardecki, C. Impact of the Protein Data Bank Across Scientific Disciplines. *Data Sci. J.* **2020**, *19*, 1–14. [CrossRef]
117. Sullivan, K.P.; Brennan-Tonetta, P.; Marxen, L.J. Economic Impacts of the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank. 2017. Available online: https://doi.org/10.2210/rcsb_pdb/pdb-econ-imp-2017 (accessed on 28 August 2022).
118. Hill, R.; Stein, C. *Scooped! Estimating Rewards for Priority in Science*; Working Paper; Massachusetts Institute of Technology: Cambridge, MA, USA, 2019.
119. Ahmed, A.; Smith, R.D.; Clark, J.J.; Dunbar, J.B., Jr.; Carlson, H.A. Recent improvements to Binding MOAD: A resource for protein-ligand binding affinities and structures. *Nucleic Acids Res.* **2015**, *43*, D465–D469. [CrossRef]
120. Gilson, M.K.; Liu, T.; Baitaluk, M.; Nicola, G.; Hwang, L.; Chong, J. BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Res.* **2016**, *44*, D1045–D1053. [CrossRef] [PubMed]
121. Sillitoe, I.; Bordin, N.; Dawson, N.; Waman, V.P.; Ashford, P.; Scholes, H.M.; Pang, C.S.M.; Woodridge, L.; Rauer, C.; Sen, N.; et al. CATH: Increased structural coverage of functional space. *Nucleic Acids Res.* **2021**, *49*, D266–D273. [CrossRef] [PubMed]
122. Groom, C.R.; Bruno, I.J.; Lightfoot, M.P.; Ward, S.C. The Cambridge Structural Database. *Acta Cryst. B Struct. Sci. Cryst. Eng. Mater.* **2016**, *72*, 171–179. [CrossRef] [PubMed]
123. Hastings, J.; Owen, G.; Dekker, A.; Ennis, M.; Kale, N.; Muthukrishnan, V.; Turner, S.; Swainston, N.; Mendes, P.; Steinbeck, C. ChEBI in 2016: Improved services and an expanding collection of metabolites. *Nucleic Acids Res.* **2016**, *44*, D1214–D1219. [CrossRef]

124. Gaulton, A.; Hersey, A.; Nowotka, M.; Bento, A.P.; Chambers, J.; Mendez, D.; Mutowo, P.; Atkinson, F.; Bellis, L.J.; Cibrian-Uhalte, E.; et al. The ChEMBL database in 2017. *Nucleic Acids Res.* **2017**, *45*, D945–D954. [CrossRef]
125. Wishart, D.S.; Feunang, Y.D.; Guo, A.C.; Lo, E.J.; Marcu, A.; Grant, J.R.; Sajed, T.; Johnson, D.; Li, C.; Sayeeda, Z.; et al. DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Res.* **2018**, *46*, D1074–D1082. [CrossRef]
126. Cheng, H.; Liao, Y.; Schaeffer, R.D.; Grishin, N.V. Manual classification strategies in the ECOD database. *Proteins Struct. Funct. Genet.* **2015**, *83*, 1238–1251. [CrossRef]
127. McDonald, A.G.; Boyce, S.; Tipton, K.F. ExplorEnz: The primary source of the IUBMB enzyme list. *Nucleic Acids Res.* **2009**, *37*, D593–D597. [CrossRef] [PubMed]
128. Harrow, J.; Frankish, A.; Gonzalez, J.M.; Tapanari, E.; Diekhans, M.; Kokocinski, F.; Aken, B.L.; Barrell, D.; Zadissa, A.; Searle, S.; et al. GENCODE: The reference human genome annotation for The ENCODE Project. *Genome Res.* **2012**, *22*, 1760–1774. [CrossRef]
129. Gene Ontology Consortium. The Gene Ontology resource: Enriching a GOld mine. *Nucleic Acids Res.* **2021**, *49*, D325–D334. [CrossRef]
130. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **2020**, *369*, 1318–1330. [CrossRef]
131. Yamada, I.; Shiota, M.; Shinmachi, D.; Ono, T.; Tsuchiya, S.; Hosoda, M.; Fujita, A.; Aoki, N.P.; Watanabe, Y.; Fujita, N.; et al. The GlyCosmos Portal: A unified and comprehensive web resource for the glycosciences. *Nat. Methods* **2020**, *17*, 649–650. [CrossRef]
132. York, W.S.; Mazumder, R.; Ranzinger, R.; Edwards, N.; Kahsay, R.; Aoki-Kinoshita, K.F.; Campbell, M.P.; Cummings, R.D.; Feizi, T.; Martin, M.; et al. GlyGen: Computational and Informatics Resources for Glycoscience. *Glycobiology* **2020**, *30*, 72–73. [CrossRef] [PubMed]
133. Tiemeyer, M.; Aoki, K.; Paulson, J.; Cummings, R.D.; York, W.S.; Karlsson, N.G.; Lisacek, F.; Packer, N.H.; Campbell, M.P.; Aoki, N.P.; et al. GlyTouCan: An accessible glycan structure repository. *Glycobiology* **2017**, *27*, 915–919. [CrossRef]
134. Lefranc, M.P.; Giudicelli, V.; Duroux, P.; Jabado-Michaloud, J.; Folch, G.; Aouinti, S.; Carillon, E.; Duvergey, H.; Houles, A.; Paysan-Lafosse, T.; et al. IMGT(R), the international ImMunoGeneTics information system(R) 25 years on. *Nucleic Acids Res.* **2015**, *43*, D413–D422. [CrossRef]
135. Vita, R.; Overton, J.A.; Greenbaum, J.A.; Ponomarenko, J.; Clark, J.D.; Cantrell, J.R.; Wheeler, D.K.; Gabbard, J.L.; Hix, D.; Sette, A.; et al. The immune epitope database (IEDB) 3.0. *Nucleic Acids Res.* **2015**, *43*, D405–D412. [CrossRef]
136. Blum, M.; Chang, H.Y.; Chuguransky, S.; Grego, T.; Kandasaamy, S.; Mitchell, A.; Nuka, G.; Paysan-Lafosse, T.; Qureshi, M.; Raj, S.; et al. The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.* **2021**, *49*, D344–D354. [CrossRef]
137. Newport, T.D.; Sansom, M.S.P.; Stansfeld, P.J. The MemProtMD database: A resource for membrane-embedded protein structures and their lipid interactions. *Nucleic Acids Res.* **2019**, *47*, D390–D397. [CrossRef] [PubMed]
138. White, S.H.; Snider, C. Membrane Proteins of Known 3D Structure (MPStruc). Available online: http://blanco.biomol.uci.edu/mpstruc/ (accessed on 28 August 2022).
139. Sayers, E.W.; Bolton, E.E.; Brister, J.R.; Canese, K.; Chan, J.; Comeau, D.C.; Connor, R.; Funk, K.; Kelly, C.; Kim, S.; et al. Database resources of the national center for biotechnology information. *Nucleic Acids Res.* **2022**, *50*, D20–D26. [CrossRef]
140. Berman, H.M.; Olson, W.K.; Beveridge, D.L.; Westbrook, J.; Gelbin, A.; Demeny, T.; Hsieh, S.H.; Srinivasan, A.R.; Schneider, B. The nucleic acid database. A comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.* **1992**, *63*, 751–759. [CrossRef]
141. Lomize, M.A.; Lomize, A.L.; Pogozheva, I.D.; Mosberg, H.I. OPM: Orientations of proteins in membranes database. *Bioinformatics* **2006**, *22*, 623–625. [CrossRef] [PubMed]
142. Su, M.; Yang, Q.; Du, Y.; Feng, G.; Liu, Z.; Li, Y.; Wang, R. Comparative Assessment of Scoring Functions: The CASF-2016 Update. *J. Chem. Inf. Model.* **2019**, *59*, 895–913. [CrossRef]
143. Hrabe, T.; Li, Z.; Sedova, M.; Rotkiewicz, P.; Jaroszewski, L.; Godzik, A. PDBFlex: Exploring flexibility in protein structures. *Nucleic Acids Res.* **2016**, *44*, D423–D428. [CrossRef]
144. Tusnady, G.E.; Dosztanyi, Z.; Simon, I. Transmembrane proteins in the Protein Data Bank: Identification and classification. *Bioinformatics* **2004**, *20*, 2964–2972. [CrossRef]
145. Nguyen, D.T.; Mathias, S.; Bologa, C.; Brunak, S.; Fernandez, N.; Gaulton, A.; Hersey, A.; Holmes, J.; Jensen, L.J.; Karlsson, A.; et al. Pharos: Collating protein information to shed light on the druggable genome. *Nucleic Acids Res.* **2017**, *45*, D995–D1002. [CrossRef]
146. Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B.A.; Thiessen, P.A.; Yu, B.; et al. PubChem in 2021: New data content and improved web interfaces. *Nucleic Acids Res.* **2021**, *49*, D1388–D1395. [CrossRef]
147. Nederveen, A.J.; Doreleijers, J.F.; Vranken, W.; Miller, Z.; Spronk, C.A.; Nabuurs, S.B.; Guntert, P.; Livny, M.; Markley, J.L.; Nilges, M.; et al. RECOORD: A recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank. *Proteins Struct. Funct. Genet.* **2005**, *59*, 662–672. [CrossRef]
148. Garavelli, J.S. The RESID Database of Protein Modifications as a resource and annotation tool. *Proteomics* **2004**, *4*, 1527–1533. [CrossRef]
149. Dunbar, J.; Krawczyk, K.; Leem, J.; Baker, T.; Fuchs, A.; Georges, G.; Shi, J.; Deane, C.M. SAbDab: The structural antibody database. *Nucleic Acids Res.* **2014**, *42*, D1140–D1146. [CrossRef]

150. Raybould, M.I.J.; Marks, C.; Lewis, A.P.; Shi, J.; Bujotzek, A.; Taddese, B.; Deane, C.M. Thera-SAbDab: The Therapeutic Structural Antibody Database. *Nucleic Acids Res.* **2020**, *48*, D383–D388. [CrossRef] [PubMed]
151. Morin, A.; Eisenbraun, B.; Key, J.; Sanschagrin, P.C.; Timony, M.A.; Ottaviano, M.; Sliz, P. Collaboration gets the most out of software. *eLife* **2013**, *2*, e01456. [CrossRef]
152. Andreeva, A.; Kulesha, E.; Gough, J.; Murzin, A.G. The SCOP database in 2020: Expanded classification of representative family and superfamily domains of known protein structures. *Nucleic Acids Res.* **2020**, *48*, D376–D382. [CrossRef]
153. Chandonia, J.M.; Fox, N.K.; Brenner, S.E. SCOPe: Classification of large macromolecular structures in the structural classification of proteins-extended database. *Nucleic Acids Res.* **2019**, *47*, D475–D481. [CrossRef]
154. Dana, J.M.; Gutmanas, A.; Tyagi, N.; Qi, G.; O'Donovan, C.; Martin, M.; Velankar, S. SIFTS: Updated Structure Integration with Function, Taxonomy and Sequences resource allows 40-fold increase in coverage of structure-based annotations for proteins. *Nucleic Acids Res.* **2019**, *47*, D482–D489. [CrossRef]
155. UniProt Consortium. UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* **2021**, *49*, D480–D489. [CrossRef]
156. Rigden, D.J.; Fernandez, X.M. The 2022 Nucleic Acids Research database issue and the online molecular biology database collection. *Nucleic Acids Res.* **2022**, *50*, D1–D10. [CrossRef]
157. Westbrook, J.D.; Burley, S.K. How Structural Biologists and the Protein Data Bank Contributed to Recent FDA New Drug Approvals. *Structure* **2019**, *27*, 211–217. [CrossRef]
158. Westbrook, J.D.; Soskind, R.; Hudson, B.P.; Burley, S.K. Impact of Protein Data Bank on Anti-neoplastic Approvals. *Drug Discov. Today* **2020**, *25*, 837–850. [CrossRef]
159. Chiu, M.L.; Gilliland, G.L. Engineering antibody therapeutics. *Curr. Opin. Struct. Biol.* **2016**, *38*, 163–173. [CrossRef]
160. Burley, S.K.; Berman, H.M.; Christie, C.; Duarte, J.M.; Feng, Z.; Westbrook, J.; Young, J.; Zardecki, C. RCSB Protein Data Bank: Sustaining a living digital data resource that enables breakthroughs in scientific research and biomedical education. *Protein Sci.* **2018**, *27*, 316–330. [CrossRef]
161. Zardecki, C.; Dutta, S.; Goodsell, D.S.; Lowe, R.; Voigt, M.; Burley, S.K. PDB-101: Educational resources supporting molecular explorations through biology and medicine. *Protein Sci.* **2022**, *31*, 129–140. [CrossRef]
162. Goodsell, D.S.; Zardecki, C.; Berman, H.M.; Burley, S.K. Insights from 20 Years of the Molecule of the Month. *Biochem. Mol. Biol. Educ.* **2020**, *48*, 350–355. [CrossRef]
163. Goodsell, D.S.; Dutta, S.; Voigt, M.; Zardecki, C.; Burley, S.K. Molecular explorations of cancer biology and therapeutics at PDB-101. *Oncogene* **2022**, *41*, 4333–4335. [CrossRef]
164. Webb, B.; Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Bioinform.* **2016**, *54*, 5.6.1–5.6.37. [CrossRef] [PubMed]
165. Webb, B.; Sali, A. Protein structure modeling with MODELLER. *Methods Mol. Biol.* **2014**, *1137*, 1–15. [CrossRef] [PubMed]
166. Pieper, U.; Webb, B.M.; Dong, G.Q.; Schneidman-Duhovny, D.; Fan, H.; Kim, S.J.; Khuri, N.; Spill, Y.G.; Weinkam, P.; Hammel, M.; et al. ModBase, a database of annotated comparative protein structure models and associated resources. *Nucleic Acids Res.* **2014**, *42*, D336–D346. [CrossRef] [PubMed]
167. Biasini, M.; Schmidt, T.; Bienert, S.; Mariani, V.; Studer, G.; Haas, J.; Johner, N.; Schenk, A.D.; Philippsen, A.; Schwede, T. OpenStructure: An integrated software framework for computational structural biology. *Acta Crystallogr. Ser. D* **2013**, *69*, 701–709. [CrossRef]
168. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F.T.; de Beer, T.A.P.; Rempfer, C.; Bordoli, L.; et al. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, *46*, W296–W303. [CrossRef]
169. Alford, R.F.; Leaver-Fay, A.; Jeliazkov, J.R.; O'Meara, M.J.; DiMaio, F.P.; Park, H.; Shapovalov, M.V.; Renfrew, P.D.; Mulligan, V.K.; Kappel, K.; et al. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* **2017**, *13*, 3031–3048. [CrossRef] [PubMed]
170. Alexander, L.T.; Lepore, R.; Kryshtafovych, A.; Adamopoulos, A.; Alahuhta, M.; Arvin, A.M.; Bomble, Y.J.; Bottcher, B.; Breyton, C.; Chiarini, V.; et al. Target highlights in CASP14: Analysis of models by structure providers. *Proteins Struct. Funct. Genet.* **2021**, *89*, 1647–1672. [CrossRef]
171. Tunyasuvunakool, K.; Adler, J.; Wu, Z.; Green, T.; Zielinski, M.; Zidek, A.; Bridgland, A.; Cowie, A.; Meyer, C.; Laydon, A.; et al. Highly accurate protein structure prediction for the human proteome. *Nature* **2021**, *596*, 590–596. [CrossRef] [PubMed]
172. Burley, S.K.; Berman, H.M. Open-access data: A cornerstone for artificial intelligence approaches to protein structure prediction. *Structure* **2021**, *29*, 515–520. [CrossRef]
173. Burley, S.K.; Arap, W.; Pasqualini, R. Predicting Proteome-Scale Protein Structure with Artificial Intelligence. *N. Engl. J. Med.* **2021**, *385*, 2191–2194. [CrossRef]
174. Shao, C.; Bittrich, S.; Wang, W.; Burley, S.K. Assessing PDB Macromolecular Crystal Structure Confidence at the Individual Amino Acid Residue Level. *Structure*, 2022; *in press*.
175. Berman, H.M.; Adams, P.D.; Bonvin, A.A.; Burley, S.K.; Carragher, B.; Chiu, W.; DiMaio, F.; Ferrin, T.E.; Gabanyi, M.J.; Goddard, T.D.; et al. Federating Structural Models and Data: Outcomes from A Workshop on Archiving Integrative Structures. *Structure* **2019**, *27*, 1745–1759. [CrossRef] [PubMed]

176. Sali, A.; Berman, H.M.; Schwede, T.; Trewhella, J.; Kleywegt, G.; Burley, S.K.; Markley, J.; Nakamura, H.; Adams, P.; Bonvin, A.M.; et al. Outcome of the First wwPDB Hybrid/Integrative Methods Task Force Workshop. *Structure* **2015**, *23*, 1156–1167. [CrossRef] [PubMed]
177. Vallat, B.; Webb, B.; Westbrook, J.D.; Sali, A.; Berman, H.M. Development of a Prototype System for Archiving Integrative/Hybrid Structure Models of Biological Macromolecules. *Structure* **2018**, *26*, 894–904.e892. [CrossRef] [PubMed]
178. Vallat, B.; Webb, B.; Fayazi, M.; Voinea, S.; Tangmunarunkit, H.; Ganesan, S.J.; Lawson, C.L.; Westbrook, J.D.; Kesselman, C.; Sali, A.; et al. New system for archiving integrative structures. *Acta Crystallogr. Sect. D Struct. Biol.* **2021**, *77*, 1486–1496. [CrossRef]
179. Burley, S.K.; Kurisu, G.; Markley, J.L.; Nakamura, H.; Velankar, S.; Berman, H.M.; Sali, A.; Schwede, T.; Trewhella, J. PDB-Dev: A Prototype System for Depositing Integrative/Hybrid Structural Models. *Structure* **2017**, *25*, 1317–1318. [CrossRef]
180. Kim, S.J.; Fernandez-Martinez, J.; Nudelman, I.; Shi, Y.; Zhang, W.; Raveh, B.; Herricks, T.; Slaughter, B.D.; Hogan, J.A.; Upla, P.; et al. Integrative structure and functional anatomy of a nuclear pore complex. *Nature* **2018**, *555*, 475–482. [CrossRef]
181. O'Reilly, F.J.; Xue, L.; Graziadei, A.; Sinn, L.; Lenz, S.; Tegunov, D.; Blotz, C.; Singh, N.; Hagen, W.J.H.; Cramer, P.; et al. In-cell architecture of an actively transcribing-translating expressome. *Science* **2020**, *369*, 554–557. [CrossRef]
182. Ganesan, S.J.; Feyder, M.J.; Chemmama, I.E.; Fang, F.; Rout, M.P.; Chait, B.T.; Shi, Y.; Munson, M.; Sali, A. Integrative structure and function of the yeast exocyst complex. *Protein Sci.* **2020**, *29*, 1486–1501. [CrossRef]
183. Mashtalir, N.; Suzuki, H.; Farrell, D.P.; Sankar, A.; Luo, J.; Filipovski, M.; D'Avino, A.R.; St Pierre, R.; Valencia, A.M.; Onikubo, T.; et al. A Structural Model of the Endogenous Human BAF Complex Informs Disease Mechanisms. *Cell* **2020**, *183*, 802–817.e824. [CrossRef]
184. Kikhney, A.G.; Borges, C.R.; Molodenskiy, D.S.; Jeffries, C.M.; Svergun, D.I. SASBDB: Towards an automatically curated and validated repository for biological scattering data. *Protein Sci.* **2020**, *29*, 66–75. [CrossRef]
185. Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D.J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; et al. The PRIDE database and related tools and resources in 2019: Improving support for quantification data. *Nucleic Acids Res.* **2019**, *47*, D442–D450. [CrossRef] [PubMed]
186. Ellenberg, J.; Swedlow, J.R.; Barlow, M.; Cook, C.E.; Sarkans, U.; Patwardhan, A.; Brazma, A.; Birney, E. A call for public archives for biological image data. *Nat. Methods* **2018**, *15*, 849–854. [CrossRef]