

Impact of structural biologists and the Protein Data Bank on small-molecule drug discovery and development

Received for publication, November 2, 2020, and in revised form, February 2, 2021. Published, Papers in Press, March 18, 2021, <https://doi.org/10.1016/j.jbc.2021.100559>

Stephen K. Burley*

From the Research Collaboratory for Structural Bioinformatics Protein Data Bank, Institute for Quantitative Biomedicine, Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, Piscataway, New Jersey, USA; Rutgers Cancer Institute of New Jersey, Robert Wood Johnson Medical School, New Brunswick, New Jersey, USA; Research Collaboratory for Structural Bioinformatics Protein Data Bank, San Diego Supercomputer Center, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, California, USA

Edited by John Denu

The Protein Data Bank (PDB) is an international core data resource central to fundamental biology, biomedicine, bioenergy, and biotechnology/bioengineering. Now celebrating its 50th anniversary, the PDB houses >175,000 experimentally determined atomic structures of proteins, nucleic acids, and their complexes with one another and small molecules and drugs. The importance of three-dimensional (3D) biostructure information for research and education obtains from the intimate link between molecular form and function evident throughout biology. Among the most prolific consumers of PDB data are biomedical researchers, who rely on the open access resource as the authoritative source of well-validated, expertly curated biostructures. This review recounts how the PDB grew from just seven protein structures to contain more than 49,000 structures of human proteins that have proven critical for understanding their roles in human health and disease. It then describes how these structures are used in academe and industry to validate drug targets, assess target druggability, characterize how tool compounds and other small-molecules bind to drug targets, guide medicinal chemistry optimization of binding affinity and selectivity, and overcome challenges during preclinical drug development. Three case studies drawn from oncology exemplify how structural biologists and open access to PDB structures impacted recent regulatory approvals of antineoplastic drugs.

This invited contribution to the Protein Data Bank (PDB)'s 50th anniversary celebration issue of the *Journal of Biological Chemistry* summarizes my perspective on the impact of structural biologists and open access to data from the PDB (1, 2) on discovery and early-stage development of small-molecule therapeutic agents. The PDB is universally recognized as being comprehensive, authoritative, FAIR (Findable, Accessible, Interoperable, and Reusable (3)), and trustworthy (<https://coretrustseal.org>). It is central to research and education in fundamental biology, biomedicine, biotechnology/bioengineering, and energy sciences worldwide (Fig. 1; (4)).

Materials for this review were drawn from PDB archive holdings; peer-reviewed publications; anecdotal accounts presented publicly at scientific meetings; my professional experience in academe (Furlaud Professor, The Rockefeller University; Full Investigator, Howard Hughes Medical Institute; University Professor and Henry Rutgers Chair, Rutgers, The State University of New Jersey) and in industry (Chief Scientific Officer, SGX Pharmaceuticals, Incorporated; Distinguished Lilly Scholar, Eli Lilly and Company); and my ongoing roles as Director of the US-funded Research Collaboratory for Structural Bioinformatics Protein Data Bank and coleader of the Worldwide Protein Data Bank (wwPDB) partnership (see below).

I first describe how PDB coverage of human proteins broadened over the past 50 years to encompass more than 49,000 structures of human proteins and then describe PDB infrastructure that directly supports structural biologists and drug hunters (and indirectly all PDB data consumers). I go on to explain how PDB structures are being used in academe and industry to validate drug targets, assess target druggability, characterize how tool compounds and other small molecules bind to drug targets, guide medicinal chemistry optimization of binding affinity and selectivity, and overcome challenges during preclinical drug development. Finally, some quantitative information on the impact of PDB structures on drug discovery and development is presented together with illustrative case studies focused on recently approved antineoplastic agents.

For the avoidance of doubt, I freely acknowledge that structural sciences and 3D structure information are but two of the many bioscience tools/data that underpin drug discovery and early-stage drug development. Other essential elements of the drug hunting ecosystem include biochemistry, biophysical measurements, genome sequencing, molecular and cellular biology, bioimaging and mass spectrometry instrumentation, high-throughput biochemical and cell-based screening, medicinal chemistry, computational chemistry, molecular dynamics simulations, pharmacology, toxicology, statistics and machine learning, etc.

* For correspondence: Stephen K. Burley, Stephen.Burley@RCSB.org.



Stephen K. Burley, University Professor and Henry Rutgers Chair at Rutgers, The State University of New Jersey, and the Director of the RCSB Protein Data Bank. His research in both academia and industry has led to major contributions in structural biology, cancer research, and structure-guided drug discovery.

technological advances underpinning MX, including synchrotron beamlines (6), routine diffraction measurements at cryogenic temperatures (7), X-ray reflection phasing using anomalous scattering (8), facile *E. coli* expression of exogenous proteins (9), and simulated annealing structure refinement against measured structure factors (10). This period also coincided with a sea change in data sharing practices within the global structural biology community; one that was a decade or more ahead of other scientific disciplines. A grass-roots movement led by a cadre of prominent structural biologists influenced scientific journals and research funders to require PDB deposition of atomic coordinates as condition for publication or funding of structural biology research. Later in 2008, PDB deposition of MX experimental data became mandatory, which was essential for making validation of 3D structures more stringent (see below). Subsequently, deposition of experimental data for structures determined using NMR spectroscopy and 3DEM became mandatory.

Growth of the PDB archive

As of December 31, 2020, PDB holdings had grown nearly 25,000-fold to 173,000 structures from just seven macromolecular crystallography (MX) structures in 1971. The PDB is currently growing at the rate of ~10% per year. During 2020, 14,058 new structures were released by the PDB archive. The vast majority of these structures continue to come from MX (~80%), with 3D electron microscopy (3DEM; (5)) adding ~17%, and nuclear magnetic resonance (NMR) spectroscopy contributing 3%.

Growth of the PDB was far from linear over its first 50 years. The decade between 1991 and 2000 saw explosive growth (~27-fold) in the PDB archive from 507 structures (as of December 31, 1990) to 13,589 by the end of 2000. This period coincided with synergistic convergence of five critical

Human protein structure coverage in the PDB archive

Among drug hunters, there is an understandably strong preference for structures of the actual drug target proteins, be they human, bacterial, or viral. Even small differences in amino acid sequence (hence 3D structure and biochemical function) can be misleading during medicinal chemistry optimization of small-molecule drug candidates. The inaugural human protein structure deposited to the PDB in 1976 was that of a Bence-Jones immunoglobulin light-chain dimer (PDB ID: 1rei (11)). By the end of 1980, PDB holdings included only five human

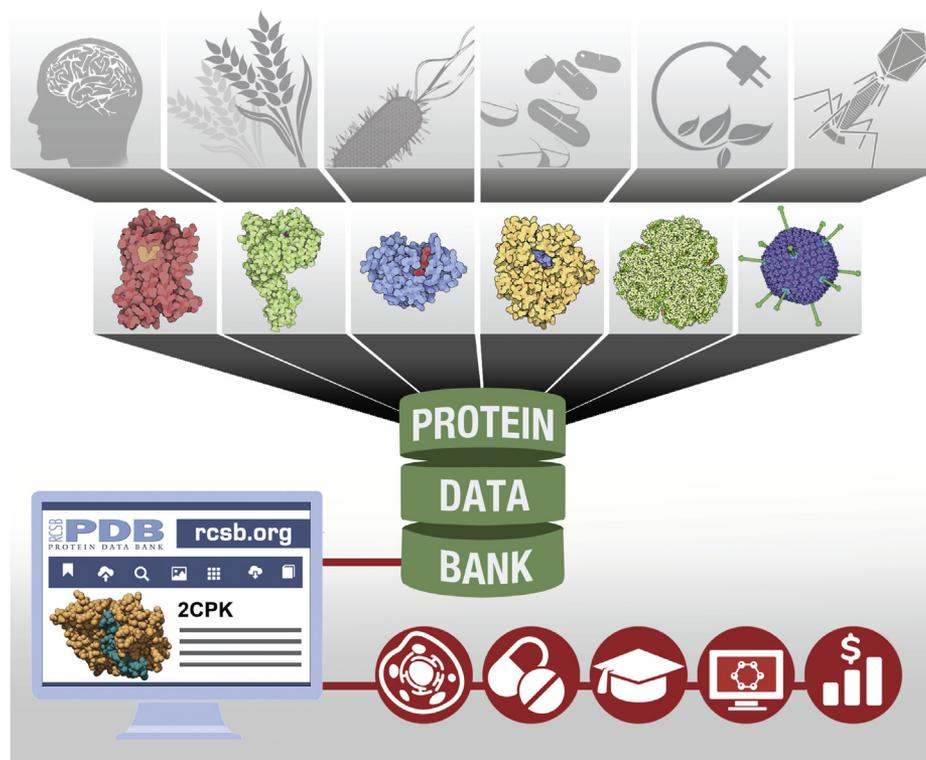


Figure 1. Availability of the Protein Data Bank archive helps transform 3D structure data to knowledge, benefiting basic and applied research and education.

proteins (~7% of archival contents), four of which were unique. One of these structurally characterized proteins, prealbumin (a.k.a. transthyretin, PDB ID: 2pab (12)), was subsequently validated as the target of tafamidis, a small-molecule drug (molecular weight or MW < 1000 Da) approved by the United States (US) Food and Drug Administration (FDA) in 2019 for treatment of adults diagnosed with cardiomyopathy due to transthyretin-mediated amyloidosis. Over the next decade, only 34 human protein structures were added to the archive. Four of them were subsequently validated as targets of approved small-molecule drugs, including hemoglobin S (PDB ID: 1hbs (13)), carbonic anhydrase 2 (PDB ID: 2ca2 (14)), elastase (PDB ID: 1hne (15)), and dihydrofolate reductase (PDB ID: 1dhf (16)).

Coincident with explosive growth in the PDB from 1991 to 2000, archival holdings of human protein structures grew nearly 68-fold to 2644. Advances in exogenous protein over-expression in insect and mammalian cell hosts further accelerated the pace of structure determination for human proteins in the 2000s (17). By the end of 2010, the archive included 16,640 human protein structures. Among the most exciting of these new structures were the first examples of human G-protein-coupled receptors (reviewed in (18)), which represent an important class of drug targets (19, 20).

As of December 31, 2020, PDB holdings of human protein structures had grown to 49,723 (~29% of the archive). In 2020, additions of first examples of human protein structures numbered 1430. This annual growth metric, reflecting ongoing increases in structural coverage of the human proteome, consistently exceeded 1000 structures/year from 2016 through 2020. Given the reductionist approach taken by many structural biologists, particularly those using either MX or NMR, these metrics should not be conflated with per residue coverage of the human proteome. SwissModel (21) provides useful tools for assessing structural coverage of human proteins and the impact of homology modeling on our knowledge of the human proteome in 3D (see <https://swissmodel.expasy.org/repository/species/9606>).

PDB infrastructure supporting structural biologists and drug discovery

Since 2003, the PDB has been managed jointly by the wwPDB partnership (<https://wwPDB.org>; (2, 22)), including the US-funded Research Collaboratory for Structural Bioinformatics Protein Data Bank or RCSB PDB (<https://RCSB.org>; (19, 23, 24)), Protein Data Bank in Europe or PDBe (<https://PDBe.org>; (25)), Protein Data Bank Japan or PDBj (<https://PDBj.org>; (26)), and Biological Magnetic Resonance Data Bank or BMRB (<https://BMRB.io>; (27)).

The wwPDB OneDep software system (<https://deposit.wwpdb.org>) for deposition, validation, biocuration of incoming structures supports PDB data depositors worldwide. OneDep helps to ensure that every structure coming into the archive is rigorously validated and expertly biocurated (28–30). The OneDep system is also used for periodic remediation of PDB data (31–34). Joint wwPDB management of the archive depends critically on active and productive engagement with community

thought leaders on matters pertaining to data standards and validation of both structures and supporting experimental data and metadata. wwPDB Validation Task Forces (35–37) and specialized workshops (38) have played important roles in developing community-endorsed validation tools that help ensure archive data quality, accuracy, and integrity. Every structure in the PDB archive is accompanied by an official, public wwPDB validation report that drug hunters and other PDB data consumers can rely on to assess the reliability of the atomic coordinates of the macromolecule and any bound ligands.

Data stored in the PDB archive conform to the PDBx/mmCIF data dictionary (39, 40). In 2014, PDBx/mmCIF became the internationally recognized, official data and metadata standard for the PDB (41). The RCSB PDB (with wwPDB partners and the wwPDB PDBx/mmCIF Working Group) coordinates PDBx/mmCIF development and hosts a public repository for data standards, metadata specifications, tutorials, and links for accessing relevant software tools (<http://mmcif.wwpdb.org>). The PDBx/mmCIF framework enables automated checking of data consistency. PDB chemical and molecular data (33, 42) are also managed with PDBx/mmCIF (<https://www.wwpdb.org/data/ccd>; <https://www.wwpdb.org/data/bird>). Use of this fully extensible, machine-readable data standard helps to ensure that PDB data are FAIR. Provenance and quality information for each data item are faithfully preserved throughout all wwPDB deposition, biocuration, validation, and remediation procedures. Since 2017, the PDB archive has supported versioning, allowing replacement of atomic coordinates by the Depositor of Record when errors, *etc.*, require correction (43).

Within the PDBx/mmCIF schema, the wwPDB Chemical Component Dictionary (CCD) (42) is of particular importance to drug hunters. Every ligand (*i.e.*, Chemical Component) represented in the PDB archive (including amino acids, nucleotides, organic compounds, and ions) is defined in the CCD. It currently houses >32,500 unique ligands. The CCD is used to standardize atom nomenclature and chemical naming using 2D graphic matching during OneDep biocuration (30). In addition, ligand geometry and chirality checking is performed using Mogul (44) (courtesy of the Cambridge Crystallographic Data Centre (CCDC) (45), as recommended by wwPDB X-ray VTF (35)). wwPDB validation software is periodically updated to incorporate recommendations from community experts. Most recently, new ligand validation tools of particular benefit to drug hunters were added by the wwPDB. They were introduced in response to recommendations from the wwPDB/CCDC/D3R Ligand Validation Workshop (38) aimed at improving the validation and ultimately the quality of co-crystal structures in the PDB (46).

Utility of PDB structures for small-molecule drug discovery and development

Over the past two decades, structural biologists and structure-guided drug discovery have become firmly established within the biopharmaceutical industry (47, 48). 3D structures can explain how small-molecule ligands bind to their target proteins (*e.g.*, imatinib targeting Abl in chronic

myeloid leukemia (PDB ID: 1iep; (49, 50))). Structural data have also proven useful in overcoming some of the myriad challenges inherent in turning biochemically active compounds into potent drug-like molecules suitable for safety and efficacy testing in animals and humans (51). Most, probably all, major biopharmaceutical companies around the world maintain a copy of the PDB archive within their firewall. Bringing PDB data inside the firewall enables interoperation of public-domain PDB structures with proprietary structures generated at each of the companies. Conservative estimates suggest that proprietary protein structures held as trade secrets inside company firewalls are comparable in aggregate number to the current size of the PDB. Willingness on the part of industry structural biologists (and company leadership) to contribute some of these data to the PDB is both much appreciated and essential for fostering ongoing technical innovations within the experimental and computational ecosystems that make structure-guided drug discovery possible (and successful).

Public-domain 3D structure data archived in the PDB are used in small-molecule drug discovery and early-stage drug development at five points within the process.

Target biology

Function follows form in biology. Atomic-level 3D structures made freely available from the PDB provide functional insights that are not always readily apparent from amino acid sequence (reviewed in (4, 20)). Simply put, there is no substitute for a direct look at the 3D structure of a potential drug discovery target. This information helps researchers understand how it works at the atomic level (*i.e.*, molecular mechanism) and how it contributes to human health and disease. Equally important, use of 3D structures to interpret the results of human genome sequencing studies (*e.g.*, driver mutations specific to tumors, genome-wide associations) influences target selection in many therapeutic areas. Industry colleagues have frequently remarked to me that “the first thing they do when starting a new drug-discovery project is to search the PDB and look hard at potential target structure(s).”

Developing the fullest possible understanding of the role that a given drug discovery target plays in human health and disease represents a critical determinant of success. An influential industry-wide analysis of the causes of attrition during drug discovery and development campaigns documented that efficacy failures account for an overall attrition rate of nearly 30% (52). A more recent analysis identified lack of efficacy as the cause of up to 66% of failures during Phase II clinical trials (53). Efficacy failures occur when the drug candidate engages the target, achieves the desired biochemical end point (*e.g.*, enzyme inhibition) without serious adverse events, yet fails to deliver the desired clinical benefit. Hence, the importance of gaining comprehensive knowledge of target biology well before embarking on expensive and lengthy human clinical trials.

Target druggability

3D structures enable visualization of surface features (*e.g.*, clefts, invaginations) likely to bind small organic compounds and

thereby inhibit enzyme action or some other biochemical or biological function (reviewed in (54)). Much of the free energy of binding of small molecules to proteins comes not from the enthalpic contribution (ΔH), but from the entropic contribution ($-\Delta S$) to the Gibbs free energy change ($\Delta G = \Delta H - T\Delta S$) of the system upon ligand binding. Higher-resolution MX and 3DEM structures of proteins frequently reveal low-entropy water molecules occupying concave surface features. When small-molecule ligands bind within such clefts, they displace most, if not all, of the bound water molecules expelling them into the bulk solvent and increasing the entropy of the system. It is, therefore, widely believed to be more difficult to discover small molecules that target flatter, relatively featureless protein surfaces (*e.g.*, protein–protein interaction interfaces) *versus* deeply invaginated clefts characteristic of enzyme active sites (55). Prequalifying a protein as a target amenable to a small-molecule drug(s) using 3D structure information can increase the probability of finding suitable lead compounds and thereafter reduce the likelihood of attrition during medicinal chemistry optimization.

Small-molecule binding

Public-domain cocrystal structures frequently provide useful precompetitive information concerning binding of tool compounds to potential drug discovery targets (56). Even more powerful are the many cocrystal structure studies carried out within biopharmaceutical companies that directly assess in 3D how small-molecule hits coming from biochemical or cell-based assays (57) or biophysical measurements (58) bind to would-be drug targets. *In silico* virtual screening exercises carried out computationally with millions of small molecules can also provide useful information regarding potential starting points for medicinal chemistry.

When it comes to understanding the chemotype binding properties of a would-be drug target, the preferred method is fragment screening (59). This approach was pioneered in the 2000s by a number of structure-guided drug discovery companies, including Astex Pharmaceuticals, Plexxikon, and SGX Pharmaceuticals, Inc. It typically involves soaking of preformed protein crystals with one or more low-molecular-weight chemical fragments (MW~150–300; frequently referred to as scaffolds) and then using high-throughput X-ray crystallography to visualize any bound fragments. Hit rates for fragment screening campaigns are typically 1 to 5%, permitting efficiency gains by using compound mixtures (60). Following fragment screening, a subset of the bound scaffolds occupying the desired site on the surface of the target (*e.g.*, enzyme active site) are selected for chemical elaboration on the bases of patent novelty, synthetic tractability, *etc.* Today, fragment screening is widely used throughout the biopharmaceutical industry (59). In some organizations, fragment libraries are prescreened against target proteins using biophysical tools such as solution NMR spectroscopy (61) and surface plasmon resonance (62). Successful prosecution of fragment/scaffold hits detected by whatever means, however, almost invariably depends on repeated access to cocrystal structures of bound compounds during medicinal chemistry lead optimization.

Typical fragment or scaffold libraries consist of 500 to 5000 compounds, each with multiple reactive sites capable of supporting automated or semiautomated chemistry with large numbers (*i.e.*, sometimes tens of thousands) of commercially available modifying substituents. Well-designed fragment libraries have the potential to be elaborated into 10^{15} or more unique chemical structure variations (assuming 1000 fragments each with three sites of chemical diversity with 10,000 possible substituents at each reactive site). Thus, the potential chemical diversity of fragment-based approaches to drug discovery far outstrips even the largest compound screening libraries assembled in either academe or industry (typically no more than a few million compounds). Notwithstanding this impressive diversity metric, even the best designed fragment library will never provide access to the enormous number of compounds possible. For reference, the number of distinct molecular structures of MW <500 Da containing only carbon, oxygen, nitrogen, hydrogen, and fluorine atoms that obey the valence rules of chemistry is estimated to be $\sim 10^{60}$ (63).

Finally, structural characterization of compound hits detected by any screening method provides valuable insights into how various chemotypes bind (or are predicted to bind) to target proteins. 3D structural information regarding screening hits that are not selected as fragments/scaffolds for medicinal chemistry optimization is frequently used to support decision-making by medicinal chemists (64). Knowledge of the chemotype binding properties of the target site can motivate selection of chemical substituents with which to modify fragment/scaffold hits. This information can also be used later in the optimization process to further optimize the fit of the elaborated lead compound to its binding site.

Structure-guided lead optimization

PDB structures and ancillary data stored in the archive regarding sample production, crystallization, *etc.*, constitute important precompetitive information routinely used by drug hunters. Open access to these data (without limitations on usage) facilitates early-stage drug discovery writ large. Whenever practicable, nearly every the major biopharmaceutical company makes intensive use of cocrystal structures to guide optimization of small-molecule ligand potency from screening hits to lead compounds to drug candidates (reviewed in (47)). In the most favorable cases, knowledge of cocrystal structures of potential off-target proteins (*e.g.*, GSK-3 β : inhibition of this protein kinase causes hyperglycemia) can be utilized to help ensure the desired selectivity profile and reduce the likelihood of off-target toxicity. In the absence of experimental cocrystal structures of the target protein, *in silico* docking tools are commonly used to guide lead optimization (reviewed in (65)). Where an experimental 3D structure of the target protein is not available, homology models are routinely combined with these same *in silico* docking tools. Machine learning approaches are also being used with increasing frequency to drive medicinal chemistry campaigns (reviewed in (66)).

Structural guidance of medicinal chemistry decision-making is particularly important when optimizing the physicochemical

properties of would-be drug candidates. Lipinski's "Rule of 5" (67) has often been touted as a basis for determining whether or not a small molecule is "drug like." Close reading of Lipinski's landmark paper, however, reveals that the Rule of 5 pertains to oral bioavailability, not drug likeness *per se*. Clinical trial experiences have repeatedly shown that more stringent limits on molecular weight (MW < 400 Da instead of 500 Da) are correlated with increased likelihood of successful outcomes (68). Lower drug candidate lipophilicity (as judged by cLogP, the calculated \log_{10} of the partition coefficient between octanol and water) is also correlated with improved clinical trial outcomes (69). Lipophilicity appears to be a critical determinant of nonspecific binding to proteins unrelated to the drug target and consequently unwanted side effects and clinical adverse events. Precise knowledge of how lead compounds bind to target proteins informs decision-making regarding the chemical modifications necessary to maintain cLogP <3 (not <5 as specified in the Rule of 5), while avoiding addition of atoms that increase MW beyond ~ 400 Da. Other molecular design considerations influenced by knowledge of 3D structure focus on avoiding synthesis of overly flat compounds, because small molecules lacking Sp³ carbons and chiral centers tend can be poorly soluble in aqueous solution (70).

Optimization of ADME properties

3D structures of proteins are used to overcome ADME (Absorption–Distribution–Metabolism–Excretion) challenges (reviewed in (51)), typically identified during early-stage (*i.e.*, preclinical) development. More than 500 PDB structures of cytochrome P450 enzymes (Cyp; earliest PDB ID released 2003: 1og2 (71)) help inform additional medicinal chemistry efforts aimed at eliminating Cyp inhibition by drug candidates (or Cyp-mediated metabolism of same) while preserving potent and selective target binding. More than 30 PDB structures of P-glycoprotein multidrug transporters (earliest PDB ID released 2009: 3g5u (72)) are available from the PDB to help medicinal chemists understand and overcome drug resistance due to cellular efflux pumps, which can interfere with the efficacy of cancer chemotherapeutic agents. Finally, three PDB structures of the human ether-a-go-go (commonly referred to as hERG) related potassium channel (earliest PDB ID released 2017: 5va1 (73)) can provide structural insights into potentially life-threatening small-molecule binding that can induce long QT syndrome, bringing with it the risk of sudden death due to the cardiac arrhythmia known as *Torsades des Pointes*.

Quantitative measures of the impact of structural biologists and the PDB on small-molecule drug discovery

Published case studies (*e.g.*, (74)) and anecdotal accounts presented at scientific meetings leave no doubt as to the importance of contributions to drug discovery made by structural biologists working in the biopharmaceutical industry. Their use of the PDB is corroborated by explicit references to the archive detected in more than 50,000 issued

patents and patent applications in process worldwide during 2016 (4).

Given the highly competitive nature of any industry and appropriate concerns of trade secrecy, publication and mandatory PDB deposition of proprietary structures by biopharmaceutical companies often lag the actual research by years. Much of the time, unfortunately, publication never happens because of resource limitations or organizational impediments to making data public. My own biopharmaceutical industry publications describing proprietary research projects attest to the fact that what you see in print and in the PDB is just the tip of the iceberg in terms of the accomplishments of industry structural biologists and their impact on drug discovery (75–78).

Two quantitative analyses of the impact of structural biologists and PDB structures on drug approvals have been published recently. The first such analysis (79) reported examinations of PDB archival holdings to identify 3D structures relevant to the 171 new small-molecule drugs approved by the US FDA from the beginning of 2010 to the end of 2016. Approximately 92% of these small-molecule drugs (158/171) have known protein targets. In 2018, the PDB archive held 5364 drug/target-related structures containing a known protein target of one of these 171 small-molecule drugs and/or the small-molecule drug itself. In approximately 20% of cases (30/171), the number of PDB structures of the drug target exceeded 100, reflecting the importance of structure-guided approaches for prosecution of certain medicinal chemistry campaigns (e.g., thrombin, earliest PDB ID released 1992: 1fph (80); HIV-protease inhibitors, earliest PDB ID released 1990: 4hvp (81)). This quantitative study also identified 277 PDB structures containing one of the US FDA-approved small-molecule drugs approved between 2010 and 2016.

Freely available structures downloaded from the PDB undoubtedly facilitated a number of these successful drug discovery and development efforts. In most cases, the earliest example of a related PDB structure of the target protein for a given small-molecule drug came from an academic research group. The remainder came from industry or, rarely, from combined academic–industry research teams. More than half of the 5364 drug/target-related structures were deposited to the PDB well before the new drug was approved by the US FDA for clinical use. The median time between earliest deposition of a related PDB structure and US FDA drug approval exceeded 10 years. This metric is significant because development of a newly discovered drug candidate can extend over many years, with the bulk of this time devoted to clinical trials. On average, a new medicine will be 10 to 15 years in the making before regulatory approval (82).

Bosutinib case study

Bosutinib, a potent inhibitor of the human nonreceptor tyrosine kinases Src and Abl (Fig. 2, A), was approved by US FDA in 2012 for treatment of adults with newly diagnosed chronic phase Philadelphia chromosome positive (Ph+) chronic myelogenous leukemia (CML) and chronic,

accelerated, or blast phase Ph+ CML with resistance or intolerance to prior therapy. In 2018, the PDB archive held a total of 179 protein structures related to Src or Abl (e.g., wild-type form of bosutinib's original target Src, mutant forms of Src, phosphorylated forms of Src, bosutinib bound to Src, bosutinib bound to Abl, other small-molecule inhibitors bound to Src, and bosutinib bound to protein kinases other than Src and Abl). Beginning in 1994 with publication and PDB deposition of the earliest structure of human Src (Src homology domain 3 or SH3) by structural biologists at the Glaxo Research Institute (PDB ID: 1shd (83)), PDB structures have provided detailed insights into the biology of Src as a nonreceptor tyrosine kinase, its role as a driver of tumor growth, and inhibition by various small molecules. This work culminated in successful clinical trials of bosutinib targeting Abl in Ph+ CML and regulatory approval. (N.B.: The first PDB structure of the Abl protein kinase catalytic domain determined by an academic research team was released in 2000 (84)). Figure 2, B and C illustrate the cocrystal structures of bosutinib bound to the catalytic domains of Src and Abl, respectively.

The second quantitative analysis of the impact of the PDB on drug approvals (85) reported examinations of archival holdings to identify 3D structures relevant to the 54 new antineoplastic small-molecule drugs approved by the US FDA from the beginning of 2010 to the end of 2018. The protein targets of every one of these small molecule drugs are known. In 2019, the PDB archive contained 2115 drug/target-related structures containing a known protein target of one of these small molecule drugs and/or the small-molecule drug itself. More than three-quarters of the recently approved small-molecule drugs (41/54, ~76%) can be found in PDB structures that reveal in 3D at atomic detail how the drug binds to its target protein and, in some cases, to off targets. Virtually all of the 2115 related PDB structures were deposited well before corresponding drug approvals by the US FDA (median time between earliest PDB deposition and regulatory approval exceeded 14 years).

Evidence for the impact of structure-guided approaches on small-molecule antineoplastic drug approvals

The second quantitative analysis also examined the impact of structure-guided discovery on US FDA approvals of antineoplastic drugs between 2010 and 2018 (85). For 39/54 (~72%) small-molecule drugs, there is either direct or indirect evidence that structure-guided lead optimization (typically starting with public domain PDB structures) was employed by one or more biopharmaceutical companies pursuing the drug target. Not surprising, a majority of these 39 cases correspond to 28 small-molecule drugs targeting one or more protein kinases. The bosutinib case study presented above reflects the somewhat serendipitous nature of protein kinase drug discovery. In this case, the target for which drug approval was obtained, Abl, was not Src, the protein for which the compound was originally intended. It is not unusual for protein kinase inhibitors to interdict the activity of more than one

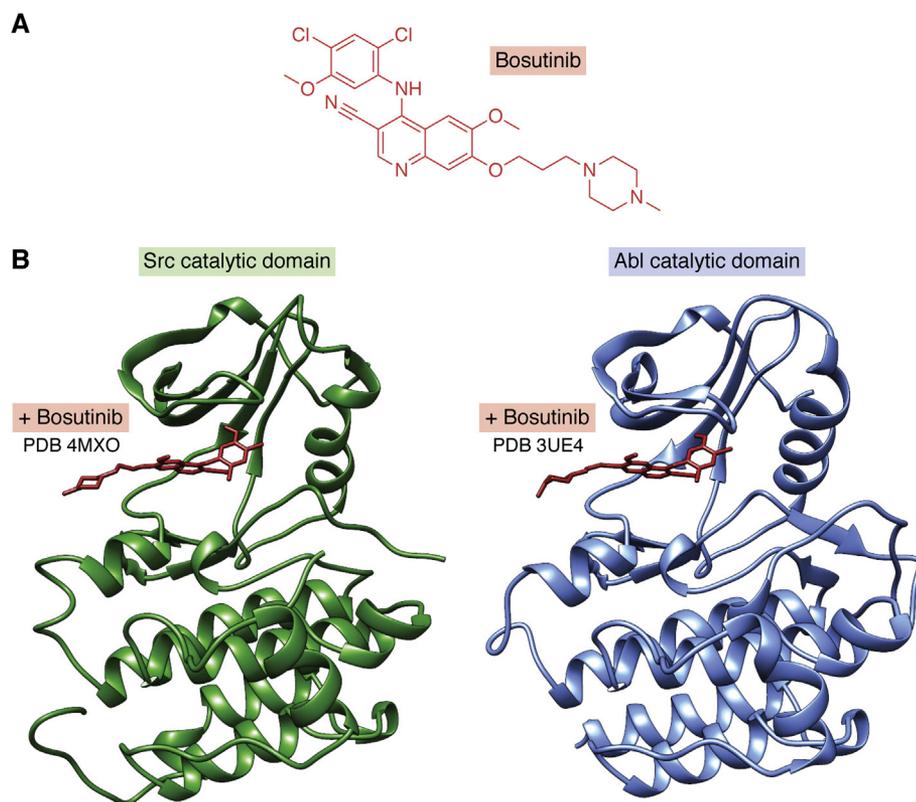


Figure 2. Bosutinib binding to Src and Abl. *A*, chemical structure of bosutinib. *B*, cocrystal structures of bosutinib (red) bound to the catalytic domains of Src (Left, green; PDB ID: 4mxo (106) and Abl) (Right, blue; PDB ID: 3ue4 (107)) non-receptor tyrosine kinases.

closely related protein kinase with beneficial effects in the clinic. Imatinib, for example, was originally approved for treatment of Ph⁺ CML targeting Abl. Following extraordinarily successful initial clinical outcomes (86), imatinib was shown to inhibit another protein kinase known as Kit, mutants of which cause gastrointestinal stromal tumors (GIST; (87)). Subsequent clinical trials with individuals diagnosed with metastatic GIST allowed the sponsor company Novartis to obtain a US FDA line-extension approval for imatinib in 2012, substantially increasing the number of patients benefiting from the drug with a commensurate impact on sales.

Two additional case studies follow illustrating the impact of academic and industrial structural biologists and PDB data on discovery of three dual cyclin-dependent kinase 4/cyclin-dependent kinase 6 (CDK4/CDK6) inhibitors and an isocitrate dehydrogenase 2 (IDH2) inhibitor.

CDK4/CDK6 case study

Among the 28 kinase inhibitors that were identified as products of structure-guided drug discovery are three compounds (palbociclib, ribociclib, and abemaciclib) approved for treatment of breast cancer that target CDK4 and CDK6. These two closely related cyclin-dependent kinases control progression through the G1 phase of the cell cycle, playing central roles in cell proliferation and tumorigenesis. Discovery and near simultaneous approval of the dual inhibitors resulted from parallel efforts by three large biopharmaceutical companies (Pfizer, Novartis, and Eli Lilly and Co). They competed head to head targeting the hinge regions of the two closely related enzymes. Each discovery team

would have relied on open access to tens of previously determined CDK structures in the PDB. They would also have been aided by open access to thousands of other protein kinase structures archived in the PDB. Most of these structures were contributed by academic researchers (earliest protein kinase PDB ID released 1993: 2cpk (88); earliest CDK4 PDB ID released 2009: 2w9f (89); earliest CDK6 PDB ID released 1999: 1bi8 (90)). Thanks to Pfizer, cocrystal structures of each dual inhibitor bound to CDK6 are available from the PDB (palbociclib PDB ID: 5L2i (91); ribociclib PDB ID: 5L2t (91); and abemaciclib PDB ID: 5L2s (91)). Close inspection of the modes of inhibitor binding revealed both common (*i.e.*, hydrogen bonding engagement of the hinge region) and disparate features of CDK6-ligand interactions for the three inhibitors (Fig. 3).

This case study also exemplifies how structure-guided lead optimization can be used to discover chemically distinct drug candidates that block the same enzyme active site, allowing each sponsor company to secure its own composition of matter patent protection. For oncology patients, availability of multiple drugs targeting the same protein can be beneficial. A given patient may tolerate side effects (due to off-target binding) of one drug better than those of another. For example, side effect profiles are known to impact patient adherence to treatment of Ph⁺ CML using imatinib and other small-molecule inhibitors of Abl (92).

IDH2 case study

IDH2 is a homodimeric, NADP(+)-dependent, mitochondrial enzyme responsible for catalyzing oxidative

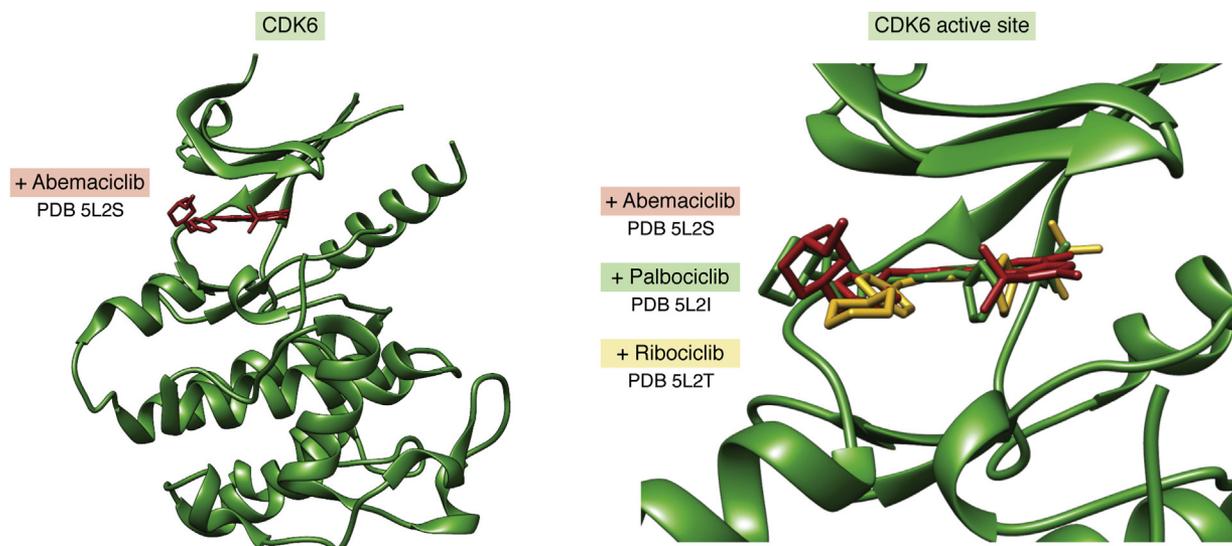


Figure 3. Hinge-binding inhibitors targeting two cyclin-dependent kinases. Left, CDK6 (green) bound to abemaciclib (red) (PDB ID: 5L2s (91)). Right, active site of CDK6 (PDB ID: 5L2s (91)) showing bound abemaciclib (red; PDB ID: 5L2s (91)), overlaid with palbociclib (green; PDB ID: 5L2i (91)) and ribociclib (yellow; PDB ID: 5L2t. (91)).

decarboxylation of isocitrate to 2-oxoglutarate in mitochondria. The first PDB structure of a mammalian IDH2 (porcine, 96% identical in amino acid sequence to human) was deposited to the PDB in 2002 by academic researchers (PDB ID: 1lwd (93)). No similar structures were publicly disclosed until 2013, when a small biotechnology company (Agiros, Inc) published the structure of the R140Q mutant form of human IDH2 bound to a small-molecule inhibitor (PDB ID: 4ja8 (94)). IDH2 was pursued as a drug discovery target because certain point mutations in IDH2 confer a gain of function on malignant cells, resulting in accumulation and secretion of the oncometabolite (*R*)-2-hydroxyglutarate (reviewed in (95)). In 2018, the PDB archive housed six MX structures of human IDH2, all of which were contributed by biopharmaceutical companies (*i.e.*, Novartis or Agios). A structure-guided drug discovery campaign at Agios yielded enasidenib (PDB ID: 5i96 (96)), which was approved by US FDA in 2017 for treatment of relapsed or refractory acute myeloid leukemia in individuals with specific mutations of the IDH2 gene confirmed by an FDA-approved diagnostic test.

Figure 4, A and B illustrate the allosteric mechanism by which enasidenib inhibits the R140Q mutant form of IDH2. The dimeric enzyme consists of two central domains that participate in dimerization and two pairs of flanking domains. A single enasidenib molecule binds to a site deep within the dimer interface well away from either of enzyme active site (Fig. 4A). The inhibited form of the enzyme remains homodimeric, and neither the central nor the two flanking domains undergo significant conformational changes due to drug binding ($C\alpha$ root-mean-square deviations for each domain range from 0.8 to 1.0 Å). Instead, a Rube Goldberg Machine-like action at a distance (see <https://www.rubegoldberg.com>) initiated by drug binding causes the entrances to the two active sites to be splayed open (ASN141- $C\alpha$ to HIS175- $C\alpha$ separation increases from 5.7 to 9.0 to 10.0 Å), thereby compromising substrate binding.

Conclusion

Over the next 50 years, we can expect the PDB archive to grow in size and importance as it provides ever larger and more diverse communities of data consumers with open access to well-validated, expertly biocurated 3D structures of biological macromolecules. wwPDB biocurators are already seeing significant flows of new structures coming into the PDB from X-ray free-electron laser studies of enzymes in action (97) and cryo-electron tomography structures of macromolecular machines at work inside living cells that have been caught in the act by flash freezing (98). Both of these rapidly evolving experimental techniques promise to reveal more about the inner workings of the natural world and bring new drug targets into focus. Looking further ahead, integrative (or hybrid) methods (99, 100) will bring 3D structures of even greater import into the PDB over the coming decades. Integrative methods do not rely on a single established method (be it MX, 3DEM, or NMR spectroscopy). Instead, they combine complementary techniques of measurement and computational analysis to resolve very large, complex macromolecular systems, some of which will be the drug targets of tomorrow.

In closing, I would be remiss if I did not highlight the myriad ways in which structural biologists and the wwPDB partners are working together to combat the COVID-19 pandemic. Since late January 2020, more than 1000 structures of SARS-CoV-2 proteins have been deposited into the PDB (<http://rcsb.org/covid19>). The vast majority of these structures were publicly released well before their appearance in the scientific literature. They are contributing to the fight against the virus in real time (101). More than 100 PDB structures of the viral spike surface glycoprotein are informing successful vaccine design efforts and enabling atomic-level 3D characterization of neutralizing antibodies in approved by the US FDA for passive immunization. Equally important are literally 100s of cocrystal structures of viral proteins bound to small-molecule ligands (102). Strikingly similar among *coronaviridae*, various essential

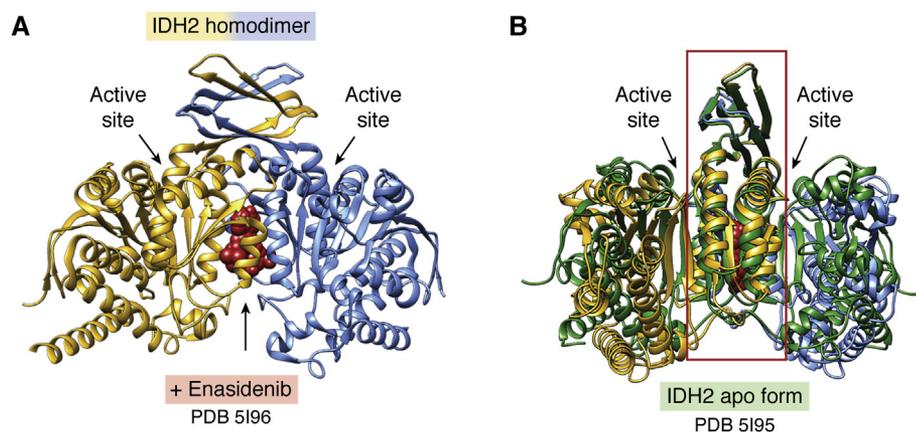


Figure 4. *A*, allosteric inhibitor enasidenib (red) targeting homodimeric IDH2 (yellow and blue monomers; PDB ID: 5i96 (96)). *B*, superposition of apo (green PDB ID: 5i95 (96)) and inhibited (yellow and blue monomers; PDB ID: 5i96 (96)) forms of IDH1 computed by overlaying the central domains (boxed in red). Arrows indicate the locations of the two active sites. The view in (*B*) is rotated $\sim 45^\circ$ about the vertical relative to (*A*).

enzymes (main protease, papain-like proteinase, RNA-dependent RNA polymerase, etc.) represent attractive targets for structure-guided discovery of small-molecule antiviral drugs (103, 104). Finally, PDB data are enabling insight into the evolution of SARS-CoV-2 proteins (105) as they change in amino acid sequence and 3D structure, sometimes rendering the virus more transmissible and possibly even deadlier.

Acknowledgments—First and foremost, the author thanks the many thousands of structural biologists who deposited structures to the PDB since 1971 and the many millions of individuals from around the world who consume PDB data. We also thank Dr Kirk L. Clark and Ms Christine Zardecki for insightful comments, Dr John Westbrook for PDB data analyses, and Dr Brian Hudson, Ms Maria Voigt, and Ms Luciana Giono for figure preparation. Finally, the author gratefully acknowledges contributions to the success of the PDB archive made by all members of RCSB PDB (past and present) and our Worldwide Protein Data Bank partners (Protein Data Bank in Europe, Protein Data Bank Japan, Biological Magnetic Resonance Bank, and Electron Microscopy Data Bank).

Author contributions—S. K. B. is responsible for all aspects of this article.

Funding and additional information—RCSB PDB is jointly funded by the National Science Foundation (DBI-1832184), the U.S. Department of Energy (DE-SC0019749), and the National Cancer Institute, National Institute of Allergy and Infectious Diseases, and National Institute of General Medical Sciences of the National Institutes of Health under grant R01GM133198. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institutes of Health.

Conflict of interest—The author declares no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are: 3D, Three-dimensional/Three dimensions; 3DEM, Three-dimensional Electron Microscopy; Abl, Abelson Proto-oncogene 1, Nonreceptor Tyrosine Kinase; ADME, Absorption-Distribution-Metabolism-Excretion; BMRB, Biological Magnetic Resonance Data Bank; CDK4, Cyclin-dependent Kinase 4; CDK6, Cyclin-dependent Kinase 6; CML, Chronic Myelogenous Leukemia; D3R, Drug Design Data Resource; DNA, Deoxyribonucleic Acid; EMDB, Electron Microscopy Data

Bank; FAIR, Findability, Accessibility, Interoperability, and Reusability; FDA, Food and Drug Administration; GIST, Gastrointestinal Stromal Tumor; HIV, Human Immunodeficiency Virus; ID, Identifier; IDH2, Isocitrate Dehydrogenase 2; mmCIF, Macromolecular Crystallographic Information File; MX, Macromolecular Crystallography; NMR, Nuclear Magnetic Resonance; OneDep, Unified wwPDB System for Deposition, Biocuration, and Validation; PDB, Protein Data Bank; PDBe, Protein Data Bank in Europe; PDBj, Protein Data Bank Japan; PDBx, Protein Data Bank Exchange Data Dictionary; PH+, Philadelphia Chromosome Positive; QT Interval, Electrocardiogram Q-wave to T-wave Interval; RCSB, Research Collaboratory for Structural Bioinformatics; RNA, Ribonucleic Acid; Src, Proto-oncogene Sarcoma Tyrosine-protein Kinase; US, United States; wwPDB, Worldwide Protein Data Bank.

References

- Protein Data Bank (1971) Crystallography: Protein Data Bank. *Nat. New Biol.* **233**, 223
- wwPDB Consortium (2019) Protein Data Bank: The single global archive for 3D macromolecular structure data. *Nucleic Acids Res.* **47**, D520–D528
- 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., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., et al. (2016) The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* **3**, 160018
- Burley, S. K., Berman, H. M., Christie, C., Duarte, J. M., Feng, Z., Westbrook, J., Young, J., and Zardecki, C. (2018) RCSB Protein Data Bank: Sustaining a living digital data resource that enables breakthroughs in scientific research and biomedical education. *Protein Sci.* **27**, 316–330
- Mitra, A. K. (2019) Visualization of biological macromolecules at near-atomic resolution: Cryo-electron microscopy comes of age. *Acta Crystallogr. F Struct. Biol. Commun.* **75**, 3–11
- Helliwell, J. R. (1998) Synchrotron radiation facilities. *Nat. Struct. Biol.* **5 Suppl**, 614–617
- Hope, H. (1988) Cryocrystallography of biological macromolecules: A generally applicable method. *Acta Crystallogr.* **B44**, 22–26
- Hendrickson, W. A. (1991) Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254**, 51–58
- Studier, F. W., and Moffatt, B. A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189**, 113–130
- Brünger, A. T., Kuriyan, J., and Karplus, M. (1987) Crystallographic R factor refinement by molecular dynamics. *Science* **235**, 458–460

11. Epp, O., Lattman, E. E., Schiffer, M., Huber, R., and Palm, W. (1975) The molecular structure of a dimer composed of the variable portions of the Bence-Jones protein REI refined at 2.0-Å resolution. *Biochemistry* **14**, 4943–4952
12. Blake, C. C., Geisow, M. J., Oatley, S. J., Rerat, B., and Rerat, C. (1978) Structure of prealbumin: Secondary, tertiary and quaternary interactions determined by Fourier refinement at 1.8 Å. *J. Mol. Biol.* **121**, 339–356
13. Padlan, E. A., and Love, W. E. (1985) Refined crystal structure of deoxyhemoglobin S. II. Molecular interactions in the crystal. *J. Biol. Chem.* **260**, 8280–8291
14. Eriksson, A. E., Kylsten, P. M., Jones, T. A., and Liljas, A. (1988) Crystallographic studies of inhibitor binding sites in human carbonic anhydrase II: A pentacoordinated binding of the SCN⁻ ion to the zinc at high pH. *Proteins* **4**, 283–293
15. Engh, R., Lobermann, H., Schneider, M., Wiegand, G., Huber, R., and Laurell, C. B. (1989) The S variant of human alpha 1-antitrypsin, structure and implications for function and metabolism. *Protein Eng.* **2**, 407–415
16. Davies, J. F., 2nd, Delcamp, T. J., Prendergast, N. J., Ashford, V. A., Freisheim, J. H., and Kraut, J. (1990) Crystal structures of recombinant human dihydrofolate reductase complexed with folate and 5-deazafofolate. *Biochemistry* **29**, 9467–9479
17. Assenberg, R., Wan, P. T., Geisse, S., and Mayr, L. M. (2013) Advances in recombinant protein expression for use in pharmaceutical research. *Curr. Opin. Struct. Biol.* **23**, 393–402
18. Granier, S., and Kobilka, B. (2012) A new era of GPCR structural and chemical biology. *Nat. Chem. Biol.* **8**, 670–673
19. Burley, S. K., Bhikadiya, C., Bi, C., Bittrich, S., Chen, L., Crichlow, G., Christie, C. H., Dalenberg, K., Costanzo, L. D., Duarte, J. M., Dutta, S., Feng, Z., Ganesan, S., Goodsell, D. S., Ghosh, S., *et al.* (2021) 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 Acids Res.* **49**, D437–D451
20. Goodsell, D. S., Zardecki, C., Di Costanzo, L., Duarte, J. M., Hudson, B. P., Persikova, I., Segura, J., Shao, C., Voigt, M., Westbrook, J. D., Young, J. Y., and Burley, S. K. (2020) RCSB Protein Data Bank: Enabling biomedical research and drug discovery. *Protein Sci.* **29**, 52–65
21. Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., and Schwede, T. (2018) SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303
22. Berman, H. M., Henrick, K., and Nakamura, H. (2003) Announcing the worldwide Protein Data Bank. *Nat. Struct. Biol.* **10**, 980
23. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242
24. Rose, Y., Duarte, J. M., Lowe, R., Segura, J., Bi, C., Bhikadiya, C., Chen, L., Rose, A. S., Bittrich, S., Burley, S. K., and Westbrook, J. D. (2020) RCSB Protein Data Bank: Architectural advances towards integrated searching and efficient access to macromolecular structure data from the PDB archive. *J. Mol. Biol.*, 166704
25. 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., Gane, P., Gaborova, R., Gupta, D., Haslam, P., Koca, J., *et al.* (2020) PDB: Improved findability of macromolecular structure data in the PDB. *Nucleic Acids Res.* **48**, D335–D343
26. Kinjo, A. R., Bekker, G. J., Suzuki, H., Tsuchiya, Y., Kawabata, T., Ikegawa, Y., and Nakamura, H. (2017) Protein Data Bank Japan (PDBj): Updated user interfaces, resource description framework, analysis tools for large structures. *Nucleic Acids Res.* **45**, D282–D288
27. Ulrich, E. L., Akutsu, H., Doreleijers, J. F., Harano, Y., Ioannidis, Y. E., Lin, J., Livny, M., Mading, S., Maziuk, D., Miller, Z., Nakatani, E., Schulte, C. F., Tolmie, D. E., Kent Wenger, R., Yao, H., *et al.* (2008) YoungMagResBank. *Nucleic Acids Res.* **36**, D402–D408
28. 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., Chen, L., Chen, M., Di Costanzo, L., Dimitropoulos, D., Gao, G., *et al.* (2017) OneDep: Unified wwPDB system for deposition, biocuration, and validation of macromolecular structures in the PDB archive. *Structure* **25**, 536–545
29. 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., Ikegawa, Y., Kobayashi, N., Lawson, C. L., Mading, S., Mak, L., *et al.* (2017) Validation of the structures in the Protein Data Bank. *Structure* **25**, 1916–1927
30. 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., Gutmanas, A., Igarashi, R., Armstrong, D. R., Baskaran, K., Chen, L., *et al.* (2018) Worldwide Protein Data Bank biocuration supporting open access to high-quality 3D structural biology data. *Database (Oxford)* **2018**, bay002
31. Henrick, K., Feng, Z., Bluhm, W. F., Dimitropoulos, D., Doreleijers, J. F., Dutta, S., Flippen-Anderson, J. L., Ionides, J., Kamada, C., Krissinel, E., Lawson, C. L., Markley, J. L., Nakamura, H., Newman, R., Shimizu, Y., *et al.* (2008) Remediation of the Protein Data Bank archive. *Nucleic Acids Res.* **36**, D426–D433
32. Lawson, C. L., Dutta, S., Westbrook, J. D., Henrick, K., and Berman, H. M. (2008) Representation of viruses in the remediated PDB archive. *Acta Crystallogr. D Biol. Crystallogr.* **D64**, 874–882
33. Dutta, S., Dimitropoulos, D., Feng, Z., Persikova, I., Sen, S., Shao, C., Westbrook, J., Young, J., Zhuravleva, M. A., Kleywegt, G. J., and Berman, H. M. (2014) Improving the representation of peptide-like inhibitor and antibiotic molecules in the Protein Data Bank. *Biopolymers* **101**, 659–668
34. Sen, S., Young, J., Berrisford, J. M., Chen, M., Conroy, M. J., Dutta, S., Di Costanzo, L., Gao, G., Ghosh, S., Hudson, B. P., Igarashi, R., Kengaku, Y., Liang, Y., Peisach, E., Persikova, I., *et al.* (2014) Small molecule annotation for the Protein Data Bank. *Database (Oxford)* **2014**, bau116
35. 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., Perrakis, A., Richardson, J. S., Sheffler, W. H., Smith, J. L., Tickle, I. J., *et al.* (2011) A new generation of crystallographic validation tools for the protein data bank. *Structure* **19**, 1395–1412
36. 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., Berman, H. M., Kleywegt, G. J., and Markley, J. L. (2013) Recommendations of the wwPDB NMR validation task force. *Structure* **21**, 1563–1570
37. Henderson, R., Sali, A., Baker, M. L., Carragher, B., Devkota, B., Downing, K. H., Egelman, E. H., Feng, Z., Frank, J., Grigorieff, N., Jiang, W., Ludtke, S. J., Medalia, O., Penczek, P. A., Rosenthal, P. B., *et al.* (2012) Outcome of the first electron microscopy validation task force meeting. *Structure* **20**, 205–214
38. 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., Burley, S. K., Case, D. A., Clark, K. L., Darden, T., Emsley, P., *et al.* (2016) Outcome of the first wwPDB/CCDC/D3R ligand validation workshop. *Structure* **24**, 502–508
39. Westbrook, J., Henrick, K., Ulrich, E. L., and Berman, H. M. (2005) 3.6.2 the Protein Data Bank exchange data dictionary. In: Hall, S. R., McMahon, B., eds. *International Tables for Crystallography*, Springer, Dordrecht, The Netherlands: 195–198
40. Fitzgerald, P. M. D., Westbrook, J. D., Bourne, P. E., McMahon, B., Watenpaugh, K. D., and Berman, H. M. (2005) 4.5 Macromolecular dictionary (mmCIF). In: Hall, S. R., McMahon, B., eds. *International Tables for Crystallography G. Definition and Exchange of Crystallographic Data*, Springer, Dordrecht, The Netherlands: 295–443
41. Adams, P. D., Afonine, P. V., Baskaran, K., Berman, H. M., Berrisford, J., Bricogne, G., Brown, D. G., Burley, S. K., Chen, M., Feng, Z., Flensburg, C., Gutmanas, A., Hoch, J. C., Ikegawa, Y., Kengaku, Y., *et al.* (2019) Announcing mandatory submission of PDBx/mmCIF format files for crystallographic depositions to the Protein Data Bank (PDB). *Acta Crystallogr. D Struct. Biol.* **75**, 451–454
42. Westbrook, J. D., Shao, C., Feng, Z., Zhuravleva, M., Velankar, S., and Young, J. (2015) The chemical component dictionary: Complete

- descriptions of constituent molecules in experimentally determined 3D macromolecules in the Protein Data Bank. *Bioinformatics* **31**, 1274–1278
43. Zardecki, C., Young, J., Westbrook, J., Feng, Z., Peisach, E., Persikova, I., Liang, Y., Gutmanas, A., Berrisford, J., Ikegawa, Y., Chen, M., Burley, S. K., Valenkar, S., and Kurisu, G. (2019) Enabling depositor-initiated PDB coordinate replacement through file versioning. *Acta Cryst.* **A75**, a25
 44. Bruno, I. J., Cole, J. C., Kessler, M., Luo, J., Motherwell, W. D., Purkis, L. H., Smith, B. R., Taylor, R., Cooper, R. I., Harris, S. E., and Orpen, A. G. (2004) Retrieval of crystallographically-derived molecular geometry information. *J. Chem. Inf. Comput. Sci.* **44**, 2133–2144
 45. Groom, C. R., Bruno, I. J., Lightfoot, M. P., and Ward, S. C. (2016) The Cambridge structural database. *Acta Crystallogr. B* **72**, 171–179
 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., Kurisu, G., Velankar, S., Burley, S. K., and Young, J. Y. (2021) Enhanced validation of small-molecule ligands and carbohydrates in the Protein Data Bank. *Structure*. <https://doi.org/10.1016/j.str.2021.02.004>
 47. Blundell, T. L. (2017) Protein crystallography and drug discovery: Recollections of knowledge exchange between academia and industry. *IUCr* **4**, 308–321
 48. Klebe, G. (2013) *Drug Design: Methodology, Concepts, and Mode-of-Action*, Springer, Berlin
 49. Nagar, B., Bornmann, W. G., Pellicena, P., Schindler, T., Veach, D. R., Miller, W. T., Clarkson, B., and Kuriyan, J. (2002) Crystal structures of the kinase domain of c-Abl in complex with the small molecule inhibitors PD173955 and imatinib (STI-571). *Cancer Res.* **62**, 4236–4243
 50. Capdeville, R., Buchdunger, E., Zimmermann, J., and Matter, A. (2002) Glivec (STI571, imatinib), a rationally developed, targeted anticancer drug. *Nat. Rev. Drug Discov.* **1**, 493–502
 51. Stoll, F., Goller, A. H., and Hillisch, A. (2011) Utility of protein structures in overcoming ADMET-related issues of drug-like compounds. *Drug Discov. Today* **16**, 530–538
 52. Kola, I., and Landis, J. (2004) Can the pharmaceutical industry reduce attrition rates? *Nat. Rev. Drug Discov.* **3**, 711–715
 53. Boyer, S., Brealey, C., and Davis, A. M. (2015) Attrition in drug discovery and development. In: Alex, A., Harris, C. J., Smith, D. A., eds. *Attrition in the Pharmaceutical Industry: Reasons, Implications, and Pathways Forward*, Wiley, Hoboken, NJ: 5–45
 54. Brown, K. K., Hann, M. M., Lakdawala, A. S., Santos, R., Thomas, P. J., and Todd, K. (2018) Approaches to target tractability assessment - a practical perspective. *Medchemcomm* **9**, 606–613
 55. Jones, S., and Thornton, J. M. (2004) Searching for functional sites in protein structures. *Curr. Opin. Chem. Biol.* **8**, 3–7
 56. Arrowsmith, C. H., Audia, J. E., Austin, C., Baell, J., Bennett, J., Blagg, J., Bountra, C., Brennan, P. E., Brown, P. J., Bunnage, M. E., Buser-Doepner, C., Campbell, R. M., Carter, A. J., Cohen, P., Copeland, R. A., et al. (2015) The promise and peril of chemical probes. *Nat. Chem. Biol.* **11**, 536–541
 57. Blay, V., Tolani, B., Ho, S. P., and Arkin, M. R. (2020) High-throughput screening: Today's biochemical and cell-based approaches. *Drug Discov. Today*. <https://doi.org/10.1016/j.drudis.2020.07.024>
 58. Renaud, J. P., Chung, C. W., Danielson, U. H., Egnér, U., Hennig, M., Hubbard, R. E., and Nar, H. (2016) Biophysics in drug discovery: Impact, challenges and opportunities. *Nat. Rev. Drug Discov.* **15**, 679–698
 59. Erlanson, D. A., Fesik, S. W., Hubbard, R. E., Jahnke, W., and Jhoti, H. (2016) Twenty years on: The impact of fragments on drug discovery. *Nat. Rev. Drug Discov.* **15**, 605–619
 60. Burley, S. K., Hirst, G., Sprengeler, P., and Reich, S. (2010) Fragment-based structure-guided drug discovery: Strategy, process, and lessons from human protein kinases. In: Merz, K., Jr., Ringe, D., Reynolds, C., eds. *Drug Design: Structure- and Ligand-Based Approaches*, Cambridge University Press, Cambridge, UK: 30–40
 61. Lepre, C. A. (2011) Practical aspects of NMR-based fragment screening. *Methods Enzymol.* **493**, 219–239
 62. Navratilova, I., and Hopkins, A. L. (2010) Fragment screening by surface plasmon resonance. *ACS Med. Chem. Lett.* **1**, 44–48
 63. Bohacek, R. S., McMartin, C., and Guida, W. C. (1996) The art and practice of structure-based drug design: A molecular modeling perspective. *Med. Res. Rev.* **16**, 3–50
 64. Burns, M. C., Howes, J. E., Sun, Q., Little, A. J., Camper, D. V., Abbott, J. R., Phan, J., Lee, T., Waterson, A. G., Rossanese, O. W., and Fesik, S. W. (2018) High-throughput screening identifies small molecules that bind to the RAS:SOS:RAS complex and perturb RAS signaling. *Anal. Biochem.* **548**, 44–52
 65. Lin, J. H. (2016) Review structure- and dynamics-based computational design of anticancer drugs. *Biopolymers* **105**, 2–9
 66. Lo, Y. C., Rensi, S. E., Torng, W., and Altman, R. B. (2018) Machine learning in chemoinformatics and drug discovery. *Drug Discov. Today* **23**, 1538–1546
 67. Lipinski, C., Lombardo, F., Dominy, B. W., and Feeney, P. J. (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **23**, 3–29
 68. Wenlock, M. C., Austin, R. P., Barton, P., Davis, A. M., and Leeson, P. D. (2003) A comparison of physicochemical property profiles of development and marketed oral drugs. *J. Med. Chem.* **46**, 1250–1256
 69. Waring, M. J., Arrowsmith, J., Leach, A. R., Leeson, P. D., Mandrell, S., Owen, R. M., Piraudeau, G., Pennie, W. D., Pickett, S. D., Wang, J., Wallace, O., and Weir, A. (2015) An analysis of the attrition of drug candidates from four major pharmaceutical companies. *Nat. Rev. Drug Discov.* **14**, 475–486
 70. Lovering, F., Bikker, J., and Humblet, C. (2009) Escape from flatland: Increasing saturation as an approach to improving clinical success. *J. Med. Chem.* **52**, 6752–6756
 71. Williams, P. A., Cosme, J., Ward, A., Angove, H. C., Matak Vinkovic, D., and Jhoti, H. (2003) Crystal structure of human cytochrome P450 2C9 with bound warfarin. *Nature* **424**, 464–468
 72. Martinez, L., Arnaud, O., Henin, E., Tao, H., Chaptal, V., Doshi, R., Andrieu, T., Dussurgey, S., Tod, M., Di Pietro, A., Zhang, Q., Chang, G., and Falson, P. (2014) Understanding polyspecificity within the substrate-binding cavity of the human multidrug resistance P-glycoprotein. *FEBS J.* **281**, 673–682
 73. Wang, W., and MacKinnon, R. (2017) Cryo-EM structure of the open human ether-a-go-go-related K(+) channel hERG. *Cell* **169**, 422–430. e410
 74. Hu, T., Sprague, E. R., Fodor, M., Stams, T., Clark, K. L., and Cowan-Jacob, S. W. (2018) The impact of structural biology in medicine illustrated with four case studies. *J. Mol. Med. (Berl.)* **96**, 9–19
 75. Buchanan, S. G., Hendle, J., Lee, P. S., Smith, C. R., Bounaud, P. Y., Jessen, K. A., Tang, C. M., Huser, N. H., Felce, J. D., Froning, K. J., Peterman, M. C., Aubol, B. E., Gessert, S. F., Sauder, J. M., Schwinn, K. D., et al. (2009) SGX523 is an exquisitely selective, ATP-competitive inhibitor of the MET receptor tyrosine kinase with antitumor activity *in vivo*. *Mol. Cancer Ther.* **8**, 3181–3190
 76. O'Hare, T., Eide, C. A., Tyner, J. W., Corbin, A. S., Wong, M. J., Buchanan, S., Holme, K., Jessen, K. A., Tang, C., Lewis, H. A., Romero, R. D., Burley, S. K., and Deininger, M. W. (2008) SGX393 inhibits the CML mutant Bcr-AblT315I and preempts *in vitro* resistance when combined with nilotinib or dasatinib. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 5507–5512
 77. Sabini, E., Hazra, S., Konrad, M., Burley, S. K., and Lavie, A. (2007) Structural basis for activation of the therapeutic L-nucleoside analogs 3TC and troxacitabine by human deoxycytidine kinase. *Nucleic Acids Res.* **35**, 186–192
 78. Atwell, S., Adams, J. M., Badger, J., Buchanan, M. D., Feil, I. K., Froning, K. J., Gao, X., Hendle, J., Keegan, K., Leon, B. C., Muller-Dieckmann, H. J., Nienaber, V. L., Noland, B. W., Post, K., Rajashankar, K. R., et al. (2004) A novel mode of Gleevec binding is revealed by the structure of spleen tyrosine kinase. *J. Biol. Chem.* **279**, 55827–55832
 79. Westbrook, J. D., and Burley, S. K. (2019) How structural biologists and the Protein Data Bank contributed to recent FDA new drug approvals. *Structure* **27**, 211–217
 80. Nar, H. (2012) The role of structural information in the discovery of direct thrombin and factor Xa inhibitors. *Trends Pharmacol. Sci.* **33**, 279–288

81. Wlodawer, A., and Vondrasek, J. (1998) Inhibitors of HIV-1 protease: A major success of structure-assisted drug design. *Annu. Rev. Biophys. Biomol. Struct.* **27**, 249–284
82. Berdigaliyev, N., and Aljofan, M. (2020) An overview of drug discovery and development. *Future Med. Chem.* **12**, 939–947
83. Gilmer, T., Rodriguez, M., Jordan, S., Crosby, R., Alligood, K., Green, M., Kimery, M., Wagner, C., Kinder, D., Charifson, P., Hassell, A. M., Willard, D., Luther, M., Rusnak, D., Sternbach, D. D., *et al.* (1994) Peptide inhibitors of src SH3-SH2-phosphoprotein interactions. *J. Biol. Chem.* **269**, 31711–31719
84. Schindler, T., Bornmann, W., Pellicena, P., Miller, W. T., Clarkson, B., and Kuriyan, J. (2000) Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. *Science* **289**, 1938–1942
85. Westbrook, J. D., Soskind, R., Hudson, B. P., and Burley, S. K. (2020) Impact of Protein Data Bank on anti-neoplastic approvals. *Drug Discov. Today* **25**, 837–850
86. Hochhaus, A., Larson, R. A., Guilhot, F., Radich, J. P., Branford, S., Hughes, T. P., Baccarani, M., Deininger, M. W., Cervantes, F., Fujihara, S., Ortmann, C. E., Menssen, H. D., Kantarjian, H., O'Brien, S. G., Druker, B. J., *et al.* (2017) Long-term outcomes of imatinib treatment for chronic myeloid leukemia. *N. Engl. J. Med.* **376**, 917–927
87. Heinrich, M. C., Corless, C. L., Demetri, G. D., Blanke, C. D., von Mehren, M., Joensuu, H., McGreevey, L. S., Chen, C. J., Van den Abbeele, A. D., Druker, B. J., Kiese, B., Eisenberg, B., Roberts, P. J., Singer, S., Fletcher, C. D., *et al.* (2003) Kinase mutations and imatinib response in patients with metastatic gastrointestinal stromal tumor. *J. Clin. Oncol.* **21**, 4342–4349
88. Knighton, D., Zheng, J., Ten Eyck, L., Ashford, V., Xuong, N., Taylor, S., and Sowadski, J. (1991) Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science* **253**, 407–414
89. Day, P. J., Cleasby, A., Tickle, I. J., O'Reilly, M., Coyle, J. E., Holding, F. P., McMenamin, R. L., Yon, J., Chopra, R., Lengauer, C., and Jhoti, H. (2009) Crystal structure of human CDK4 in complex with a D-type cyclin. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 4166–4170
90. Russo, A. A., Tong, L., Lee, J. O., Jeffrey, P. D., and Pavletich, N. P. (1998) Structural basis for inhibition of the cyclin-dependent kinase Cdk6 by the tumour suppressor p16INK4a. *Nature* **395**, 237–243
91. Chen, P., Lee, N. V., Hu, W., Xu, M., Ferre, R. A., Lam, H., Bergqvist, S., Solowiej, J., Diehl, W., He, Y. A., Yu, X., Nagata, A., VanArsdale, T., and Murray, B. W. (2016) Spectrum and degree of CDK drug interactions predicts clinical performance. *Mol. Cancer Ther.* **15**, 2273–2281
92. Rychter, A., Jerzmanowski, P., Holub, A., Specht-Szwoch, Z., Kalinowska, V., Tegowska, U., Seferynska, I., Kolkowska-Lesniak, A., Lech-Maranda, E., and Gora-Tybor, J. (2017) Treatment adherence in chronic myeloid leukaemia patients receiving tyrosine kinase inhibitors. *Med. Oncol.* **34**, 104
93. Ceccarelli, C., Grodsky, N. B., Ariyaratne, N., Colman, R. F., and Bahnsen, B. J. (2002) Crystal structure of porcine mitochondrial NADP+-dependent isocitrate dehydrogenase complexed with Mn2+ and isocitrate. Insights into the enzyme mechanism. *J. Biol. Chem.* **277**, 43454–43462
94. Wang, F., Travins, J., DeLaBarre, B., Penard-Lacronique, V., Schalm, S., Hansen, E., Straley, K., Kernysky, A., Liu, W., Gliser, C., Yang, H., Gross, S., Artin, E., Saada, V., Mylonas, E., *et al.* (2013) Targeted inhibition of mutant IDH2 in leukemia cells induces cellular differentiation. *Science* **340**, 622–626
95. Molenaar, R. J., Maciejewski, J. P., Wilmink, J. W., and van Noorden, C. J. F. (2018) Wild-type and mutated IDH1/2 enzymes and therapy responses. *Oncogene* **37**, 1949–1960
96. Yen, K., Travins, J., Wang, F., David, M. D., Artin, E., Straley, K., Padyana, A., Gross, S., DeLaBarre, B., Tobin, E., Chen, Y., Nagaraja, R., Choe, S., Jin, L., Konteatis, Z., *et al.* (2017) AG-221, a first-in-class therapy targeting acute myeloid leukemia harboring oncogenic IDH2 mutations. *Cancer Discov.* **7**, 478–493
97. Orville, A. M. (2018) Entering an era of dynamic structural biology. *BMC Biol.* **16**, 55
98. Turk, M., and Baumeister, W. (2020) The promise and the challenges of cryo-electron tomography. *FEBS Lett.* **594**, 3243–3261
99. Sali, A., Berman, H. M., Schwede, T., Trewhella, J., Kleywegt, G., Burley, S. K., Markley, J., Nakamura, H., Adams, P., Bonvin, A. M., Chiu, W., Peraro, M. D., Di Maio, F., Ferrin, T. E., Grunewald, K., *et al.* (2015) Outcome of the first wwPDB hybrid/integrative methods task force workshop. *Structure* **23**, 1156–1167
100. Vallat, B., Webb, B., Westbrook, J. D., Sali, A., and Berman, H. M. (2018) Development of a prototype system for archiving integrative/hybrid structure models of biological macromolecules. *Structure* **26**, 894–904
101. Chodera, J., Lee, A. A., London, N., and von Delft, F. (2020) Crowd-sourcing drug discovery for pandemics. *Nat. Chem.* **12**, 581
102. Douangamath, A., Fearon, D., Gehertz, P., Krojer, T., Lukacik, P., Owen, C. D., Resnick, E., Strain-Damerell, C., Aimon, A., Abranyi-Balogh, P., Brandao-Neto, J., Carbery, A., Davison, G., Dias, A., Downes, T. D., *et al.* (2020) Crystallographic and electrophilic fragment screening of the SARS-CoV-2 main protease. *Nat. Commun.* **11**, 5047
103. Brzezinski, D., Kowiel, M., Cooper, D. R., Cymborowski, M., Grabowski, M., Wlodawer, A., Dauter, Z., Shabalina, I. G., Gilski, M., Rupp, B., Jaskolski, M., and Minor, W. (2021) Covid-19.bioreproducibility.org: A web resource for SARS-CoV-2-related structural models. *Protein Sci.* **30**, 115–124
104. Burley, S. K. (2020) How to help the free market fight coronavirus. *Nature* **580**, 167
105. Lubin, J. H., Zardecki, C., Dolan, E. M., Lu, C., Shen, Z., Dutta, S., Westbrook, J. D., Hudson, B. P., Goodsell, D. S., Williams, J. K., Voigt, M., Sarma, V., Xie, L., Venkatachalam, T., Arnold, S., *et al.* (2020) Evolution of the SARS-CoV-2 proteome in three dimensions (3D) during the first six months of the COVID-19 pandemic. *bioRxiv*. <https://doi.org/10.1101/2020.12.01.406637>
106. Levinson, N. M., and Boxer, S. G. (2014) A conserved water-mediated hydrogen bond network defines bosutinib's kinase selectivity. *Nat. Chem. Biol.* **10**, 127–132
107. Levinson, N. M., and Boxer, S. G. (2012) Structural and spectroscopic analysis of the kinase inhibitor bosutinib and an isomer of bosutinib binding to the Abl tyrosine kinase domain. *PLoS One* **7**, e29828