RCSB Protein Data Bank: Enabling biomedical research and drug discovery

David S. Goodsell[1,2,3], Christine Zardecki[1,2] , Luigi Di Costanzo[1,2], Jose M. Duarte[4], Brian P. Hudson[1,2], Irina Persikova[1,2], Joan Segura[4], Chenghua Shao[1,2], Maria Voigt[1,2], John D. Westbrook[1,2], Jasmine Y. Young[1,2], and Stephen K. Burley[1,2,4,5]

[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] The Scripps Research Institute, La Jolla, CA 92037, 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] Rutgers Cancer Institute of New Jersey, Rutgers, The State University of New Jersey, New Brunswick, NJ 08903, USA

Running Header:
RCSB Protein Data Bank: Enabling biomedical research and drug discovery

Keywords:
protein structure and function, structural biology, Protein Data Bank, PDB, structure-guided drug discovery, integral membrane proteins, GPCR, ion channel, transporter, ubiquitin ligase

Abstract

The Protein Data Bank (PDB) archive currently holds >154,000 atomic level three-dimensional (3D) structures of biomolecules experimentally determined using crystallography, NMR spectroscopy, and electron microscopy. The archive was established in 1971 as the first open-access, digital-data resource in biology, and is now managed by the Worldwide Protein Data Bank partnership (wwPDB; wwPDB.org). US PDB operations are the responsibility of the RCSB Protein Data Bank (RCSB PDB; RCSB.org; based at Rutgers University, UC San Diego, and UC San Francisco). The RCSB PDB serves millions of RCSB.org users worldwide by delivering PDB data integrated with ~40 external biodata resources, providing rich structural views of fundamental biology, biomedicine, and energy sciences. In addition, the RCSB PDB outreach/education portal serves hundreds of thousands of PDB101.RCSB.org users worldwide, who are primarily university educators and their undergraduate students. Not counted in these usage metrics are the many PDB users working in biopharmaceutical companies, wherein copies of the PDB archive are retained within firewalls for interoperation with proprietary structures. Recently published work has shown that the holdings of the PDB archive facilitated discovery and development of nearly 90% of the 210 new medical entities (or new drugs) approved by the US Food and Drug Administration (US FDA) between 2010 and 2016. Growth of the PDB archive in terms of size and complexity is reported herein, together with reviews of archival holdings of protein targets important for discovery and development of new drugs (e.g., G-protein coupled receptors, voltage-gated ion channels, ligand-gated ion channels, transporters, and E3 ubiquitin ligases).

Overview of the PDB Archive, the wwPDB, and the RCSB PDB

Protein Data Bank (PDB) archive was established in 1971 as the first open access digital-data resource in the biological sciences with seven protein structures.[1;2] Current PDB archival holdings encompass >154,000 atomic level structures of proteins, DNA, and RNA, experimentally determined by macromolecular X-ray crystallography (MX: ~90%), nuclear magnetic resonance spectroscopy (NMR: ~9%), and 3D electron microscopy (3DEM: ~1%). Nearly three quarters (~73%) of PDB structures also include one or more ligands (e.g., enzyme co-factors and inhibitors, US FDA approved drugs, metals), and ~10% of PDB structures include one or more carbohydrate components. Virtually all PDB structures were determined with the support of research funding from governments or private philanthropies, and the PDB archive is now widely regarded as an international public good. Replacement value of current PDB archival holdings is conservatively estimated at >15 billion US dollars.[3]

The PDB archive is jointly managed by the Worldwide Protein Data Bank partnership (wwPDB; wwPDB.org),[4] consisting of the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank,[5,6] Protein Data Bank Japan (PDBj),[7] Protein Data Bank in Europe (PDBe),[8] and BioMagResBank (BMRB).[9] The wwPDB operates under an international agreement (wwpdb.org/about/agreement). Adhering to the FAIR principles of Findability, Accessibility, Interoperability, and Reusability,[10] wwPDB partners standardize, collect, validate, biocurate, securely store, and remediate macromolecular structure data as a single global archive for *Data Depositors* and disseminate these data *via* FTP to *Data Consumers*, all at no charge with no restrictions on data usage.

US PDB operations are the responsibility of the RCSB Protein Data Bank (RCSB PDB; RCSB.org) with financial support from the National Science Foundation, the National Institute of General Medical Sciences, the National Cancer Institute, the National Institute of Allergy and Infectious Disease, and Department of Energy. The RCSB PDB has performance sites at Rutgers, The State University of New Jersey, the San Diego Supercomputer Center at the University of California San Diego (UCSD), and the University of California San Francisco (UCSF). RCSB PDB also serves as the global Archive Keeper, responsible for ensuring disaster recovery of PDB data and coordinating weekly archival updates among wwPDB partners in Europe and Asia.

RCSB PDB serves millions of users worldwide, primarily through the web portal at RCSB.org. The website, described recently in *Nucleic Acids Research*,[6] provides tools and services to access and explore PDB content. Each week, PDB data are integrated with ~40 external data resources to provide rich, up-to-date structural views of fundamental biology, biomedicine, and energy sciences. Data can be searched and explored through individual Structure Summary Pages, or as groups of structures displayed in tabular reports.

Since RCSB PDB users (*Data Consumers*) extend well beyond experts in structural biology,[11,12] our website features are designed to enable finding a variety of structures related to a particular topic using search tools (e.g., by sequence, sequence similarity, small molecule name). The website also offers alternatives to searching, such as the Browse by Annotation tool that organizes PDB structures into hierarchical trees based upon several different classifications, including Anatomical Therapeutic Chemical (ATC) drug classification system developed by the Word Health Organization Collaborating Centre for Drug Statistics Methodology (http://www.whocc.no/atc_ddd_index/); protein residue modifications in the PDB archive using the protein modification ontology (PSI-MOD) from the Proteomics Standards Initiative;[13] and Biological Process**,** Cellular Component**,** and Molecular Function based upon descriptions from the Gene Ontology (GO) Consortium[14] mapped to corresponding PDB structures by the SIFTS initiative.[15]

Different visualization options are available. Protein Feature View offers graphical summaries of full-length UniProt[16] protein sequences and how they correspond to PDB entries, together with annotations from external databases (such as Pfam),[17] homology model information,[18,19] predicted regions of protein disorder, and hydrophobic regions. Rapid 3D visualization of structures large and small is possible using the NGL viewer,[20] which includes specialized options for viewing ligand interactions and electron density maps. Many RCSB PDB features available on RCSB.org are also provided as Web Services supporting programmatic access.

A separate website, PDB101.rcsb.org ("101", as in an entry level course), hosts educational materials to encourage learning about proteins and nucleic acids in 3D. A main focus is the Molecule of the Month series,[21] currently in its 20th year of telling molecular stories about structure and function. Other materials include molecular origami paper models, posters, animations, and curricular materials. A “Guide to Understanding PDB Data” is built around more PDB-specific information: PDB Data, Visualizing Structures, Reading Coordinate Files, and Potential Challenges (including biological assembly *versus* asymmetric unit).

All RCSB PDB activities are supported by robust infrastructure that ensures 24/7/365 support of millions of *Data Depositors* and *Data Consumers* worldwide. A full description of RCSB PDB services was recently published,[6] along various analyses of the impact of structural biologists, the PDB archive, and the RCSB PDB have also been published.[12,22,23]

Growth in the Size and Complexity of the PDB Archive

Figure 1 illustrates the growth in the PDB archive since 2000. Atomic coordinates for >11,200 new structures together with experimental data/metadata (~7.6% year-on-year growth). Most of these new structures in 2018 were determined using MX (~88.8%), with the remainder determined by 3DEM (~7.6%), and NMR (~3.5%). The number of 3DEM structures populating the archive has been growing rapidly since structural biologists ushered in the "resolution revolution"[24] (Figure 1B). Since 2016, annual 3DEM structure depositions have exceeded NMR structure depositions. Global data deposition statistics, maintained from 2000 onwards, are updated on a weekly basis (http://www.wwpdb.org/stats/deposition).

The global structural biology community has also been depositing increasingly more complex structures into the PDB archive. Figures 2A and 2B reflect the complexity of structures deposited to the PDB archive *versus* time. Growth in the number of distinct small-molecule ligands represented in the PDB Chemical Component Dictionary (CCD)[25] is illustrated in Figure 2A (2498 new ligands were added in 2018, corresponding to year-on-year growth of 7.7%). Entries in the PDB CCD include amino acids; nucleosides and nucleotides; carbohydrates; metals and other ions; crystallization and buffer solutes; enzyme co-factors, substrates, and products; prosthetic groups (e.g., heme); oligopeptides; small organic molecules; and pharmacologic agents. In parallel with the growth of the CCD, the average size of each PDB entry, as gauged by mean aggregate molecular weight of the biological assembly, is also growing (Figure 2B). Not surprisingly, 3DEM has contributed disproportionately to the growth in the number of larger PDB structures since early 2014 (Figure 2C).

Drug Target Structures in the PDB

Structure-guided drug discovery is a well-established tool for large and small biopharmaceutical companies alike.[26] 3D structural studies frequently aid in optimizing small-molecule ligand affinity and selectivity for target proteins [e.g., vemurafenib approved for treatment of the 50% of late stage metastatic

melanoma patients with the Val600→Glu mutation that activates the BRAF protein kinase, PDB structure 3og7[27]]. A recent RCSB PDB analysis[23] documented that United States Food and Drug Administration (US FDA) approval of 88% of 210 new molecular entities (NMEs or new drugs from 2010 to 2016) was facilitated by open access to ~6,000 PDB structures containing the protein targeted by the NME and/or the new drug itself. More than half of these structures were described in the scientific literature and publicly released >10 years before final drug approval. Moreover, these structures were cited in a significant fraction of the more than 2 million papers reporting publicly-funded, pre-competitive research on the drug targets that influenced drug company investment decisions, leading ultimately to the US FDA approvals and patient access to new life altering drugs. Finally, the impact of structural biologists and the PDB archive on US FDA new drug approvals was similar across all therapeutic areas.

Integral Membrane Protein Structures in the PDB

Recent successes enjoyed by structural biologists studying integral membrane proteins document that the PDB archive will represent an increasingly important source of pre-competitive information supporting ongoing and future drug discovery campaigns directed at these challenging targets. More than 50% of the targets of current US FDA approved drugs are integral membrane proteins.[28] The vast majority of these drug targets fall within four well-studied protein families (G-protein coupled receptors or GPCRs: ~30%; voltage-gated ion channels or VGICs: 8%; ligand-gated ion channels or LGICs: 7%; and transporters: 7%; examples of each are displayed Figure 3). The following sections briefly review current PDB holdings and highlight opportunities for structure-guided drug discovery for each of these major classes of target proteins.

G-protein Coupled Receptors (GPCRs): PDB archival holdings of GPCRs at the time of writing are summarized in Table 1. Since publication of the landmark structure of bovine rhodopsin (PDB 1f88) in 2000[29], more than 300 GPCR structures from four of the five GPCR sub-families have been determined and deposited in the archive, including A-Rhodopsin, B1-Secretin, C-Glutamate, and F-Frizzled/Taste 2 (but not B2-Adhesion).[30] The vast majority of these structures were determined using MX (~91%), with a small number coming from NMR (~2%), and a growing number coming from 3DEM (~7%). Initially, GPCR structure depositions to the PDB were restricted to the Rhodopsin sub-family, many of them crystallized using lipidic mesophases[31] and visualized as chimeras with entire proteins or smaller protein domains inserted into extra-membranous

loops (e.g., T4 phage lysozyme[32]) that facilitate crystal lattice formation without perturbing the structure of the 7-transmembrane helix (7-TM) domain. At present, the PDB archive contains structures for more than 60 unique GPCRs (representing examples or orthologs of ~15% of the entire complement of more than 800 GPCRs encoded by the human genome). GPCR structures have been elucidated in both active and inactive conformational states, some including bound small-molecule ligands or drugs, bound peptide/protein ligands, bound heterotrimeric G proteins, and in some cases stabilizing Fab fragments and/or camelid-nanobodies.[33] Structure-guided drug discovery for GPCRs (particularly Class A members) using MX is currently being pursued within many of the large biopharmaceutical companies, targeting both receptors represented within the PDB and novel receptors, exclusive to one or more companies.

Table 1. **G-protein Coupled Receptors** in the PDB archive. Table generated in June 2019 using sequence searching with representative members of each class.

	All	Class A (Rhodopsin)	Class B1 (Secretin)	Class B2 (Adhesion)	Class C (Glutamate)	Class F (Frizzled/Taste 2)
Structures	339	295	23	0	8	13
X-ray	311	278	15	0	6	12
Resolution (Å)		7.7-1.7	3.3-1.9		3.1-2.2	3.9-2.4
NMR*	6	6		0		0
3DEM	22	11	8	0	2	1
Resolution (Å)		4.5-3.0	4.1-3.0		4.0	3.8
Unique Receptors	62	52	6	0	2	2

* One Solid State NMR entry (PDB 2lnl)[34]

The first 3DEM structure of a GPCR to become publicly available (PDB 5vai)[35] revealed the structure of glucagon-like peptide 1 (GLP1) being recognized by the glucagon-like peptide-1 receptor (GLP1-R: active conformation) that was embedded in a detergent micelle and bound to a G-protein hetero-trimer (Figure 3A). Underscoring the power of cryo-electron microscopy to enable structural studies of large/complex and very challenging samples, this Class B1 (Secretin) GPCR was visualized at the atomic level in the act of recognizing its 31-residue peptide hormone ligand, while engaging with a G-protein hetero-trimer. Comparison of this structure with the structures of other activated GPCRs and G-protein hetero-trimers previously deposited to the PDB documents that inclusion of a stabilizing nanobody (NB-35)[36] in the sample deposited onto EM

grids does not appear to have perturbed the 3D structures of either the GPCR or the G-proteins.

GLP1-R[37] is the target of seven oligopeptide agonists (exenatide, liraglutide, lixisenatide, albiglutide, dulaglutide, and semaglutide) approved by the US FDA for treatment of type II diabetes mellitus. These biologic agents, the newest of which was approved in 2017, mimic endogenous GLP1 and promote secretion of insulin by the patient's own pancreas in response to elevated glucose levels. The principal advantage of this treatment strategy *versus* older/cheaper small-molecule insulin secretagogues is it carries a lower risk of hypoglycemia. At present, there are no publicly-available structures of any of the seven approved GLP1-R agonists bound to full-length GLP1-R. With open access to PDB structure 5vai, detailed knowledge of how 5vai and related structures were determined, and recent acquisitions of 3DEM instrumentation by biopharmaceutical companies, the stage is now set for structure-guided discovery of the next generation of GLP1-R agonists with improved pharmacologic properties (i.e., longer half-lives that will permit less frequent dosing and improve the likelihood of compliance). It also appears highly likely that 3DEM will shortly reveal one or more structures of Class B2 (Adhesion) GPCRs, some of which are drug discovery targets[38,39], and all of which have thus far eluded 3D structure determination by any experimental method.

Voltage-Gated Ion channels (VGICs): Voltage-gated ion channels open and close ion-selective pores in response to small changes in membrane potential, playing central roles in nerve signal transmission. They form a large superfamily that includes voltage-gated sodium (Na_v), calcium (Ca_v), potassium (K_v), and other ion channels, encoded by at least 143 human genes,[40,41] making them the third largest family of signaling proteins after GPCRs and protein/lipid kinases. Pioneering work on the homo-tetrameric potassium channel KcsA from *S. lividans*[42] and *A. pernix* K_vAP[43] revealed the mechanistic bases for ion selectivity and gating at the atomic level, but structural information for Na_v and Ca_v channels, which lack the structural 4-fold symmetry seen in KcsA, has only recently become available.

Voltage-gated sodium channels give rise to the rapid action potentials that mediate nerve transmission, making them targets for natural and designed toxins, inhibitors, and drugs.[44,45] Many examples are found in nature, including the potent and exquisitely selective tetrodotoxin, a neurotoxin found in puffer fish (and other organisms) with a lethal dose being less than a milligram. Many anticonvulsants, antiarrhythmics, and local anesthetics, such as lamotrigine, flecainide, and lidocaine, also act by blocking these channels.[46] The PDB currently contains >750 VGIC structures (Table 2).

Table 2. **Voltage-gated Ion Channels** available from the PDB Archive. Table generated in July 2019 based on Gene Ontology or GO # 0005244.

	Voltage-gated ion channels	Voltage-gated potassium channel activity	High voltage-gated calcium channel activity	Voltage-gated proton channel activity	NMDA glutamate receptor activity	Voltage-gated anion channel activity	Voltage-gated ion channel activity involved in regulation of postsynaptic membrane potential	Voltage-gated ion channel activity involved in regulation of presynaptic membrane potential
Structures	756	235	237	8	135	31	21	17
X-ray	494	156	132	7	94	11	16	8
Resolution (Å)	1.2-4.8	1.2-4.8	1.4-4.4	1.4-3.5	1.3-4.0	1.6-4.1	1.4-2.6	1.9-3.0
NMR	64	30	25	1	2	2	5	7
3DEM	197	48	80	0	39	18	0	2
Resolution (Å)	2.9-35	2.9-10	3.0-35	n/a	4.5-16.5	3.2-6.6	n/a	3.0-3.8
Hybrid	1	1	0	0	0	0	0	0
Unique Channels	105	41	30	1	6	7	4	4

The human genome encodes nine voltage-gated sodium channels (Na_v, designated $Na_v1.1$ to $Na_v1.9$), which support a range of cellular and biological functions. $Na_v1.1$, $Na_v1.2$, $Na_v1.3$, and $Na_v1.6$ are expressed primarily in the central nervous system; $Na_v1.4$ is found in the skeletal muscle; $Na_v1.5$ is found in cardiac muscle; and $Na_v1.7$, $Na_v1.8$, and $Na_v1.9$ are typically found in the peripheral tissues. Much of the early structural work on Na_v channels was performed using bacterial homologs, which have a simpler homo-tetrameric structure and proved relatively easy to express, purify, and crystallize. These first MX structures were reported in 2011 for *A. butzleri* Na_vAb (PDB 3rvy, 3rvz, 3rw0).[47] In contrast, the mammalian channel is composed of one long alpha chain, which forms four membrane-spanning domains similar in arrangement to the four identical bacterial subunits. In addition, the alpha subunit associates with one or more copies of five beta-subunits (beta1, beta1B, beta2, beta3, and beta4). Recently, 3DEM structures have been determined for human $Na_v1.4$/beta2 (PDB 6agf),[48] $Na_v1.2$/beta2 with a conotoxin (PDB 6j8e),[49] and $Na_v1.7$/beta1/beta2 with tetrodotoxin and saxotoxin (PDB 6j8i and others, Figure 3B).[49]

Drug discovery efforts have focused considerable resources on $Na_v1.7$.[50] This work began in earnest following the 2004 discovery that a $Na_v1.7$ gain-of-function

mutation causes persistent pain[51]. In 2006, a loss-of-function mutation was identified in several Pakistani street performers, who show no sensitivity to pain while walking on hot coals.[52] Selectivity remains an elusive challenge in this arena. Pair-wise sequence identities among the nine human Na_v VGICs exceed 70%. To complicate matters further, multiple functional binding sites for both large and small molecules are present on the solvent-accessible surfaces of these integral membrane proteins. Prior to the availability of 3DEM structures of Na_v VGICs, much of the early drug design work was performed using homology models based on distantly related bacterial homologs.

Today, medicinal chemists are sifting through multiple sites of action of natural toxins and poisons with the aim of finding druggable sites with potential for specificity, and then targeting them with small molecules, peptides, or antibodies. Notwithstanding insights from these new structures, serious challenges remain for structure-guided drug discovery. Na_v VGICs are conformationally dynamic, existing in multiple functional states (e.g., closed/resting, open, and closed/inactivated) each of which will need to be structurally characterized. Single-particle 3DEM, however, offers a critical advantage *versus* MX in that multiple conformations of a macromolecular assembly can be accommodated *via* focused classification procedures[53] to reveal multiple structural states on the EM grid.[54]

Ligand-Gated Ion Channels (LGICs): Ligand-gated ion channels mediate transmission of signals across nerve synapses in response to binding of neurotransmitters. There are three major structural classes of these channels (Table 3): pentameric "Cys-loop" receptors, ionotropic glutamate receptors, and P2X receptors.[55] The pentameric Cys-loop receptors include excitatory cation-selective channels, such as the nicotinic acetylcholine receptors and inhibitory anion-selective channels (e.g., the $GABA_A$ receptor). In 2005, Nigel Unwin's ground-breaking EM structure of the nicotinic acetylcholine receptor from the marbled electric ray (PDB 2bg9) revealed at the atomic-level both ligand-binding subunits and channel geometry.[56] A large collection of toxins, poisons, and drugs act through these pentameric receptors, including two well-known poisons curare and strychnine;[57] anesthetics and alcohol;[58] benzodiazopine antidepressants;[59] and the antiparasitic agent ivermectin.[60]

Table 3. **Ligand-Gated Ion Channels** in the PDB Archive. Table generated in July 2019 based on Gene Ontology or GO # 0015276.

	All*	Cyclic nucleotide-gated ion channel activity	Extracellular ligand-gated ion channel activity	Intracellular ligand-gated ion channel activity	Ligand-gated anion channel activity	Ligand-gated cation channel activity	Ligand-gated ion channel activity involved in regulation of presynaptic membrane potential
Structures	968	38	647	241	88	865	159
X-ray	685	26	506	122	59	612	115
Resolution	1.15-7.4	1.65-3.28	1.15-4.79	1.21-7.4	1.55-3.8	1.15-7.4	1.24-3.96
NMR	57	5	29	18	9	47	2
3DEM	223	4	112	98	20	203	45
Resolution (Å)	2.94-50	3.4-3.51	2.95-50	2.94-8.5	3.04-6.6	2.94-50	3.8-16.5
EC	3	3		3		3	
Resolution (Å)	3.54-3.8	3.54-3.8		3.54-3.8		3.54-3.8	
Unique genes	84	8	45	24	11	69	5

* N.B.: PDB structures may appear in multiple LGIC classification categories.

Ionotropic glutamate receptors (iGluR) fall into four main classes, based on their small-molecule binding properties: AMPA receptors (GluA1-4), kainate receptors (GluK1-5), NMDA receptors (GluN1, GluN2A-D, GluN3A-B), and delta receptors (GluD1-2).[61,62] These polypeptide chains can form both homo- and hetero-tetramers, and associate with a variety of modulatory auxiliary subunits. They are modular in structure. An amino-terminal domain (homologous to bacterial periplasmic binding proteins) mediates dimerization between subunits of the same iGluR subfamily. The C-terminal portion contains the extracellular agonist-binding domain, which consists of two segments separated by the portion that forms the membrane-spanning ion channel pore. MX analysis of extracellular fragments of iGluR proved instrumental in characterizing some of the functionally important properties of these channels.[63] Beginning in 2009 with publication of the MX structure of GluA2 AMPA receptor (PDB 3kg2),[64] work in this area has moved rapidly. Today, multiple 3DEM structures of iGluR and their complexes with ligands, toxins, and accessory proteins are also available.[65,66]

As seen for the VGICs, the iGluRs display multiple sites for binding of toxins and poisons, and many of these LGICs are currently the focus of structure-guided drug discovery efforts (see the 2019 special issue of *ACS Med. Chem. Lett.* on Allosteric Modulation of Ionotropic Glutamate Receptors).[67] For example, memantine, an NMDA receptor channel blocker, has been approved for treatment of moderate-to-severe Alzheimer's patients.[68] A 3DEM structure of the hetero-trimeric GluN1/GluN2A/GluN2B NMDA receptor with a similar agent (MK-801, PDB 5uow,[69] Figure 3C) revealed the ligand binding site within a vestibule of the ion channel. Pre-clinical characterization of MK-801 underscores both the promise and the challenges posed by targeting these receptors. Neuroprotection was observed in animal models of stroke, traumatic brain injury, and Parkinsonism, accompanied by side effects of induced psychotic behavior and neuronal degeneration. A subsequently determined 3.6Å resolution MX structure of an N-terminal truncated form of the receptor enabled molecular dynamics simulations of MK-801 and memantine binding (PDB 5un1),[70] further advancing structure-guided design efforts aimed at improving side effect profiles.

Transporters: The transporters constitute a large, heterogenous class of membrane-spanning proteins involved in trafficking of small cargo molecules across lipid bilayers. Sequence mapping of PDB structures to the Transporter Classification Database (TCDB)[71] yielded 9,834 matches in the PDB archive (as of July 2019). Membrane transporters play central roles in ADME (absorption, distribution, metabolism, elimination) and pharmacodynamic properties of drugs.[72] The human genome encodes >400 membrane transporters that fall into two superfamilies: ATP-Binding Cassette or ABC Superfamily and SoLute Carrier (SLC) family. All ABC transporters and many SLC transporters function as active transporters, using either ATP or electrochemical gradients to drive transport.

Table 4. **Transporter Proteins** in the PDB Archive. Table generated in July 2019 using sequences from the Transporter Classification Database.[71]

	All	Class 1: Channels /Pores	Class 2: Electrochemical Potential-driven Transporters	Class 3: Primary Active Transporters	Class 4: Group Translocators	Class 5: Transmembrane Electron Carriers	Class 8: Accessory Factors Involved in Transport	Class 9: Incompletely Characterized Transport Systems
Structures	9834	4131	721	2203	80	96	1651	952
X-Ray	8207	3364	669	1911	52	79	1272	860
Resolution (Å)		7.6-0.82	5.97-1.0	7.78-0.88	3.91-1.45	3.7-0.99	7.81-0.73	7.0-0.85
NMR	716	313	15	80	21	13	216	58
3DEM	911	454	37	212	7	4	163	34
Resolution (Å)		50.0-1.9	14.0-3.0	37.0-2.0	4.3-2.6	3.8-3.1	35.0-2.6	35.0-3.0
Unique	3429	1318	214	810	46	43	655	343

ABC transporters were first identified in bacterial nutrient import systems, bearing a characteristic ATP-binding domain with a phosphate-binding loop (commonly known as the P-loop or Walker A motif) and a short "Leu-Ser-Gly-Gly-Gln" consensus sequence.[73] Similar motifs were later found in the bacterial multidrug-resistance export pump P-glycoprotein (P-gp). Subsequent studies revealed that 1-3% of bacterial genomes encode ABC transporters, which act as variously as importers or exporters. In all, the human genome encodes 48 ABC exporters, which fall into seven subfamilies (designated A-G). Multiple MX structures of bacterial/archaeal ABC transporters are available from the PDB archive, with some bound to periplasmic binding proteins responsible for delivering substrates to the transporter.[74] These structures revealed various conformational states that cycle between inward- and outward-facing configurations. Instructive examples include an early structure of the vitamin b12 transporter (BtuCD, PDB 1l7v),[75] and three states of the *E. coli* maltose transporter MalEFGK2 (inward-open, PDB 3fh6;[76] pre-translocation, PDB 3pv0;[77] and outward-open, PDB 2r6g[78]).

ABC transporters are also relevant to human health and disease. For example, P-gp and BCRP (Breast Cancer Resistance Protein), found on the luminal surfaces of cells in the gut, modulate oral bioavailability of drug, and are, therefore, key determinants of ADME properties.[72,79] For example, increased expression of P-gp in cancer cells confers resistance to various chemotherapeutic agents. 3DEM structures, beginning with the complex of P-gp with cyclic peptide inhibitors (PDB 3g61),[80] are revealing the mechanism(s) of action of these transporters and modes of targeted inhibition.[81] The CFTR transporter (Figure 3D) is an ABC

chloride ion transporter.[82] More than 2000 CFTR gene variants have been identified in humans. Many of these differences are causative of cystic fibrosis, the most common autosomal recessive genetic disease affecting Caucasians. Inadequate chloride transport causes accumulation of mucus in lung and pancreas, leading to chronic pulmonary inflammation/infection and exocrine pancreatic insuffiency[83]. The most common disease-causing CFTR variants include deletion of Phe508, which accelerates protein degradation, and Arg117→His and Gly551→Asp, which yield transporters with gating defects. The US FDA-approved drug ivacaftor acts as a potentiator of these gating variants, yielding increase chloride transmission.[84] A recently-determined 3DEM structure of human CFTR (PDB 6o2p)[82] revealed that the drug binds at the protein-lipid interface within the transmembrane region, at a hinge site known to be involved in gating. It has been hypothesized that ivacaftor, which was discovered *via* phenotypic screening, stabilizes the open configuration of the transporter. With an EM structure in hand and a druggable site identified, the door is now open to structure-guided drug discovery of 2nd generation drugs targeting a broader spectrum of mutations causative of cystic fibrosis in affected individuals.

The SLC superfamily is highly heterogeneous, with 52 distinct human protein families that show little sequence or structural similarity, sharing simply their roles in intake and/or efflux of small molecules and inorganic ions across membranes.[85] A 2017 review[86] tabulates atomic-level 3D structure determinations for members of 23 SLC families, largely prokaryotic proteins, such as the long-awaited and much-anticipated structure of lactose permease (PDB 1pv6).[87] These structures revealed much diversity in the mechanism(s) of substrate recognition (as might have be expected), but also commonalities in the local conformational changes responsible for opening and closing “gates” on either side of the membrane to regulate transport. SLC proteins are only now being explored as drug discovery targets. A recent perspective issued a “call-to-arms” to explore this diverse and functionally important subset of transporters.[88] Successes to date include various US FDA-approved drugs, such as selective serotonin reuptake inhibitors (SSRIs) for treatment of depression, and sodium/glucose co-transporter (SGLT2) inhibitors for treatment of type 2 diabetes.[89] Structure-guided discovery of SSRIs relied largely on MX studies of a bacterial homolog LeuT,[90] first determined in 2005 (PDB 2a65).[91] Recent structures of human serotonin transporters (PDB 5i6x and others)[92] will almost certainly improve the impact of this approach for discovery and development of new pharmacologic agents targeting neuro-psychiatric disorders.

E3 Ubiquitin Ligase Structures in the PDB

Recent successes enjoyed by structural biologists studying complex multi-protein assemblies show that the PDB archive will come to represent an increasingly important source of pre-competitive information facilitating structure-guided drug discovery of other challenging targets that are not integral membrane proteins. Some of the most exciting new drug targets among the large macromolecular machines can be found within the large family of E3 ubiquitin ligases. These enzymes confer substrate selectivity on the ubiquitin-proteasome system (UPS) for degrading protein cytosolic.[93] The UPS pathway is regulated by sequential action of three classes of activating enzymes: E1 (2 human enzymes), E2 (~40 human enzymes), and E3 (>600 human enzymes) (Figure 4A).[94] The end result of this combinatorial three-step enzyme cascade is a ubiquitinated substrate that is in turn recognized and degraded by the 26S proteasome.

Table 5. **Ubiquitin-like** structures in the PDB archive. Table generated in July 2019 based on Gene Ontology or GO # 0061659, ubiquitin-like protein ligase activity.

	Ubiquitin-like protein ligase activity (All*)	NEDD* ligase activity	SUMO ligase activity	Ubiquitin protein ligase activity
Structures	984	113	5	979
X-ray	804	104	3	801
Resolution (Å)	0.8-8.3	1.1-3.3	1.7-2.4	0.8-8.3
NMR	138	9	2	136
3DEM	42	0	0	42
Resolution (Å)	2.9-16	n/a	n/a	2.9-16
Unique E3 ubiquitin ligases by gene names	123	1	4	119

* N.B.: PDB structures may appear in multiple LGIC classification categories.

E3 ubiquitin ligases include components respectively responsible for catalysis (i.e., ubiquitination) and substrate recognition. In some cases, both functionalities are embedded within a single polypeptide chain. In many others, E3 is made up of multiple protein chains. As might be expected from the many types of substrate proteins that undergo targeted ubiquitination, E3 substrate-recognition components are highly heterogenous, and variously recognize short substrate peptides called “degrons” or larger protein surface features. Ubiquitination machines have been classified into three major families: RING (Really Interesting New Gene), HECT (Homologous with E6-associated protein C-Terminus), and the, more recently described, family of hybrid RBR (RING-IBR-RING) E3s. Extensive structural studies (Table 5) have revealed the central role played by flexibility in influencing interactions of E3s with E2-ubiquitin conjugates. This work has also explored the mechanisms by which multiple bacterial and viral proteins hijack the UPS.[95]

An exciting new development in this arena is the prospect of targeting the UPS to proteins of pharmacologic interest. This work was inspired, at least in part, by natural products (e.g., auxin, a well-characterized small-molecule plant hormone) that stabilize interactions between an E3 ligase and its degradation target.[96] Similarly, immunomodulatory imide drugs (IMiDs), such as the teratogen thalidomide, are being repurposed on the strength of their recently discovered

ability to stabilize binding of the E3 ligase CRL4-DDBI-CRBN to several lymphoid transcription factors.[97] A new strategy is also being explored to design bifunctional molecules, called PROTACs or degraders, that recognize a common site on the surface of an E3 and a specific epitope on the surface of a target protein, bringing them together to promote targeted degradation.[98] Recently deposited PDB MX structures have revealed at the atomic level how such a degrader links the chromatin-reader protein Brd4 (bromodomain-containing protein 4), a target for cancer therapy, with ubiquitin ligase complexes such as pVHL (von Hippel-Lindau disease tumor suppressor):ElonginC:ElonginB (Figure 4B, PDB 5t35)[99] and DDBI:CRBN (PDB 6bn7).[100] These proof-of-concept structures open the door to structure-guided discovery of similar degraders selective for other protein targets, *via* engineering of linkers based on structures of specific ligands bound to each of the partners.

Conclusion/Perspective

The success of the discipline we have come to know as structural biology and the relentless growth of the open-access PDB archive bode well for the continued impact of 3D biostructure data on basic and applied research across the biological and medical sciences. Of particular importance looking ahead will be the explosive growth of 3DEM depositions to the archive. Since 2016, annual 3DEM depositions have exceeded those coming from NMR spectroscopy. Notwithstanding whispers to the contrary in some quarters, MX is actually "alive and well" and remains the mainstay experimental method for atomic-level 3D structure determinations of macromolecules, accounting for ~90% of 2018 PDB depositions. The precise role that 3DEM will play in structure-guided drug discovery going forward remains to be determined. It is clear, however, from private communications received from biopharmaceutical company colleagues that they are benefiting from even lower-resolution 3DEM structures of macromolecular machines wherein tool compounds can be visualized bound to druggable surface features such as deep invaginations and protein-protein interface clefts. Knowledge of the functional groups presented by the amino acid sidechains comprising putative drug binding sites can be particularly helpful for hypothesis generation during synthesis of early lead compounds. MX is likely to remain the method of choice for any drug-discovery target that supports facile crystallization and production of higher-resolution (i.e., better than 2.2Å) co-crystal structures with pharmaceutically-acceptable lead compounds and even drug candidates. Diffraction data in these cases is typically obtained at modern synchrotron radiation sources in <1 minute of beam time and refined structures therefrom can often be generated with automated scripts within 1 hour following the experiment. It is not unusual for structural biologists

working in biopharmaceutical companies to deliver new, highly-informative co-crystal structures with one to two weeks of compound synthesis. Increased use of X-ray free electron lasers (XFELs) and serial femtosecond crystallography (SFX) in drug discovery will accelerate this process even more, while generating large numbers of structures during each experiment.[101] 3DEM will have to come a long way in terms of efficiency before it can rival the speed and relatively low cost of MX structure determination. Whatever the outcome of this "horse race", the open-access PDB archive will continue to play central roles in research and education, facilitating discovery of new biomaterials, new drugs, and new diagnostic tools around the world.

Acknowledgements

The RCSB PDB is jointly funded by the National Science Foundation (DBI-1832184), the National Institutes of Health (R01GM133198), and the United States Department of Energy (DE-SSC0019749). We gratefully acknowledge contributions from current and past members of the RCSB PDB and our wwPDB partners.

Figure 1. Growth of PDB archive, 2000-2018. A) Total number of structures publicly available each year by experimental method. B) New structures released annually by experimental method, shown using logarithmic scale.

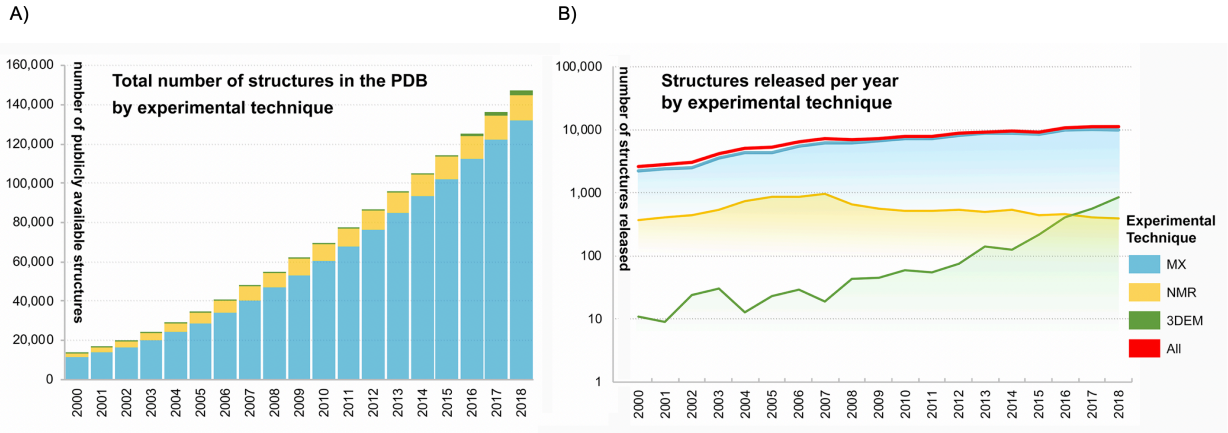


Figure 2. Growth in the complexity of PDB archival holdings 2000-2018. A) Total number of unique ligands maintained in the Chemical Component Dictionary each year. In 2018, 2498 new entries were added. B) Average molecular weight (solid purple line) and average number of polymer chains (solid orange line) of structures released each year. C) Total growth in available EM structure data, shown by accumulated number of chains and molecular weight.

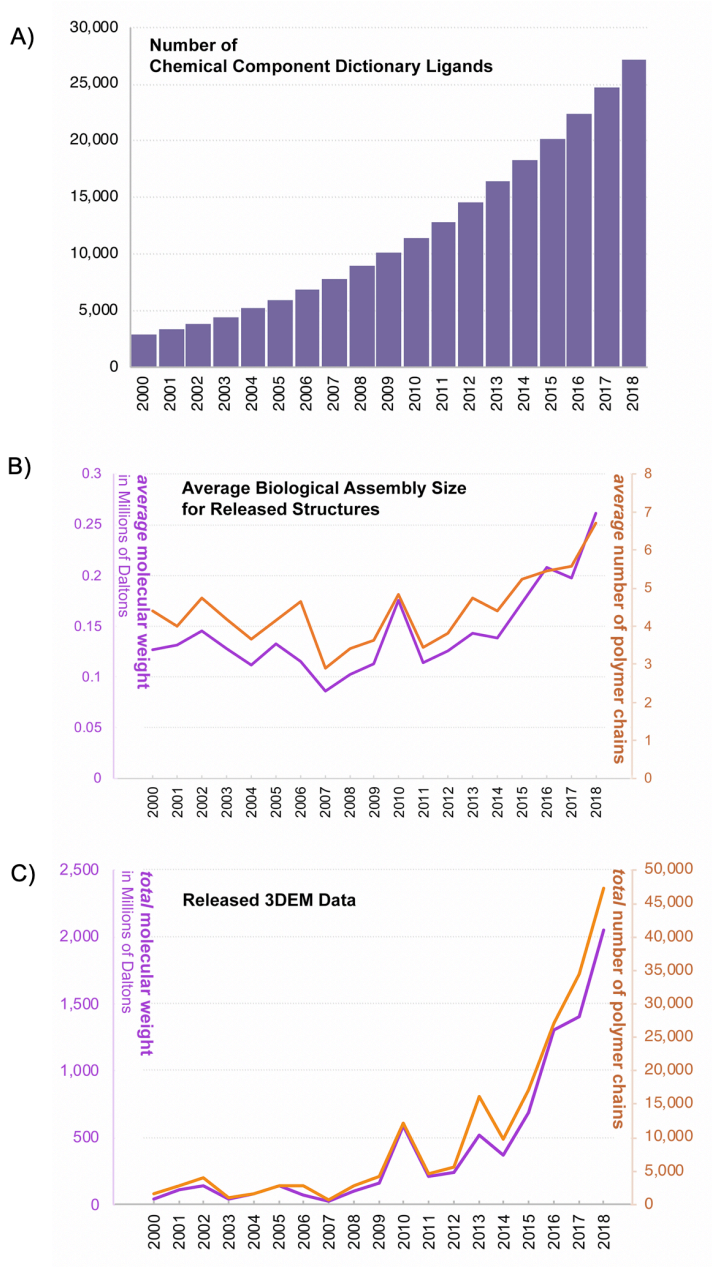


Figure 3: Ribbon drawings of exemplar structures from each of the four classes of membrane-bound proteins of pharmacologic interest, viewed parallel to the lipid bilayer (shaded grey rectangle). (A) GPCR (PDB 5vai)[35]: GLP1-R (glucagon-like peptide-1 receptor, active conformation in green) bound to GLP1 (red) and heterotrimeric G-protein (blue, yellow, magenta); (B) VGIC (PDB 6j8i)[102]: $Na_v1.7$ (green), beta1 and beta2 (blue), bound to inhibitor tetrodotoxin (yellow). Voltage-sensing helices are shown in red. (C) LGIC (PDB 5uow)[69]: NMDA receptor (blue, green, red) bound to channel blocker MK-801 (magenta). An antibody Fab (grey) was used in the structure determination. (D) Transporter (PDB 6o2p)[82]: CFTR (cystic fibrosis transmembrane conductance regulator, blue) bound to ivacaftor (yellow), which interacts with a long transmembrane helix involved in gating (red).

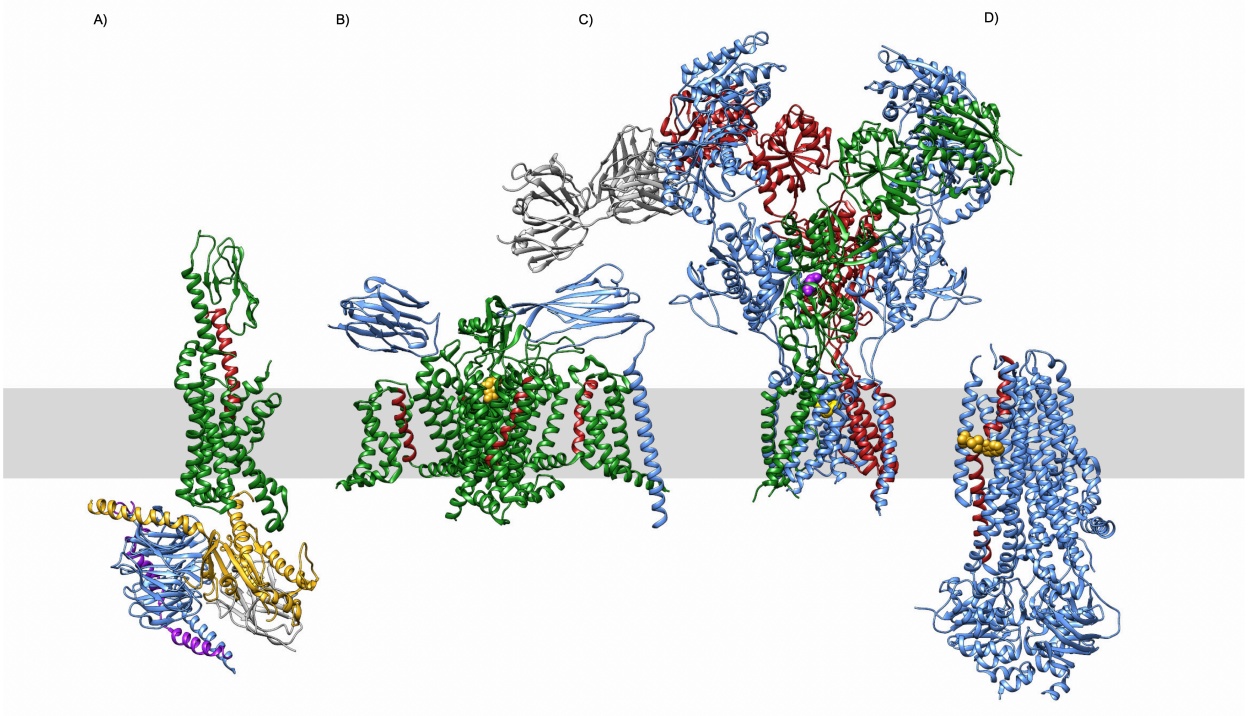
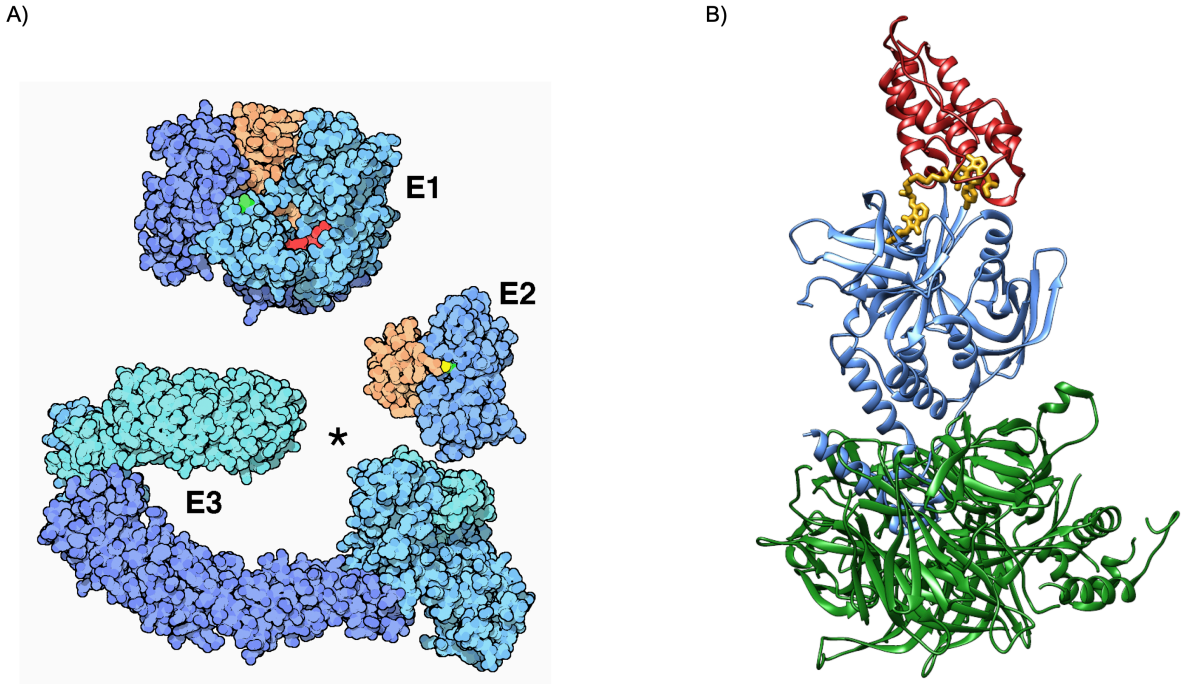


Figure 4 (A) Early structures of the components of the ubiquitin ligase system (E1,[103] E2,[104] and E3,[105,106]). Image from PDB-101 Molecule of the Month.[21] (B) Ribbon structure of the PROTAC degrader MZ1 (yellow) linking cancer target Brd4 (red) to a ubiquitin ligase complex of pVHL:ElonginC:ElonginB (blue and green) from PDB 6bn7.[100]



References

1. Protein Data Bank. (1971) Protein Data Bank. Nature New Biol. 233: 223.

2. (1972). *Cold Spring Harbor Symposia on Quantitative Biology*, 36, Cold Spring Laboratory Press.

3. Sullivan KP, Brennan-Tonetta P, Marxen LJ. (2017) Economic Impacts of the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank. doi: 10.2210/rcsb_pdb/pdb-econ-imp-2017.

4. wwPDB consortium. (2019) Protein Data Bank: the single global archive for 3D macromolecular structure data. Nucleic Acids Res 47: D520-D528.

5. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. (2000) The Protein Data Bank. Nucleic Acids Res 28: 235-242.

6. Burley SK, Berman HM, Bhikadiya C, Bi C, Chen L, Di Costanzo L, Christie C, Dalenberg K, Duarte JM, Dutta S, Feng Z, Ghosh S, Goodsell DS, Green RK, Guranovic V, Guzenko D, Hudson BP, Kalro T, Liang Y, Lowe R, Namkoong H, Peisach E, Periskova I, Prlic A, Randle C, Rose A, Rose P, Sala R, Sekharan M, Shao C, Tan L, Tao YP, Valasatava Y, Voigt M, Westbrook J, Woo J, Yang H, Young J, Zhuravleva M, Zardecki C. (2019) RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. Nucleic Acids Res 47: D464-D474.

7. Kinjo AR, Bekker GJ, Wako H, Endo S, Tsuchiya Y, Sato H, Nishi H, Kinoshita K, Suzuki H, Kawabata T, Yokochi M, Iwata T, Kobayashi N, Fujiwara T, Kurisu G, Nakamura H. (2018) New tools and functions in data-out activities at Protein Data Bank Japan (PDBj). Protein Sci. 27: 95-102.

8. Mir S, Alhroub Y, Anyango S, Armstrong DR, Berrisford JM, Clark AR, Conroy MJ, Dana JM, Deshpande M, Gupta D, Gutmanas A, Haslam P, Mak L, Mukhopadhyay A, Nadzirin N, Paysan-Lafosse T, Sehnal D, Sen S, Smart OS, Varadi M, Kleywegt GJ, Velankar S. (2018) PDBe: towards reusable data delivery infrastructure at protein data bank in Europe. Nucleic Acids Res 46: D486-D492.

9. Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J, Livny M, Mading S, Maziuk D, Miller Z, Nakatani E, Schulte CF, Tolmie DE, Kent Wenger R, Yao H, Markley JL. (2008) BioMagResBank. Nucleic Acids Res 36: D402-408.

10. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A,

Gray AJ, Groth P, Goble C, Grethe JS, Heringa J, t Hoen PA, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone SA, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B. (2016) The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3: 160018.

11. Basner J. (2017) Impact Analysis of "Berman HM et al., (2000), The Protein Data Bank". doi: 10.2210/rcsb_pdb/pdb-cit-anal-2017.

12. Markosian C, Di Costanzo L, Sekharan M, Shao C, Burley SK, Zardecki C. (2018) Analysis of impact metrics for the Protein Data Bank. Sci Data 5: 180212.

13. Gao J, Prlic A, Bi C, Bluhm WF, Dimitropoulos D, Xu D, Bourne PE, Rose PW. (2017) BioJava-ModFinder: identification of protein modifications in 3D structures from the Protein Data Bank. Bioinformatics 33: 2047-2049.

14. Gene Ontology Consortium. (2015) Gene Ontology Consortium: going forward. Nucleic Acids Res 43: D1049-1056.

15. Velankar S, Dana JM, Jacobsen J, van Ginkel G, Gane PJ, Luo J, Oldfield TJ, O'Donovan C, Martin MJ, Kleywegt GJ. (2013) SIFTS: Structure Integration with Function, Taxonomy and Sequences resource. Nucleic Acids Res 41: D483-489.

16. The UniProt Consortium. (2017) UniProt: the universal protein knowledgebase. Nucleic Acids Res 45: D158-D169.

17. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. (2016) The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res 44: D279-285.

18. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T. (2018) SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res 46: W296-W303.

19. Pieper U, Webb BM, Dong GQ, Schneidman-Duhovny D, Fan H, Kim SJ, Khuri N, Spill YG, Weinkam P, Hammel M, Tainer JA, Nilges M, Sali A. (2014) ModBase, a database of annotated comparative protein structure models and associated resources. Nucleic Acids Res 42: D336-346.

20. Rose AS, Bradley AR, Valasatava Y, Duarte JM, Prlić A, Rose PW. (2018) NGL viewer: web-based molecular graphics for large complexes. Bioinformatics: bty419.

21. Goodsell DS, Dutta S, Zardecki C, Voigt M, Berman HM, Burley SK. (2015) The RCSB PDB "Molecule of the Month": Inspiring a Molecular View of Biology. PLoS Biol. 13: e1002140.

22. Burley SK, Berman HM, Christie C, Duarte JM, Feng Z, Westbrook J, Young J, 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.

23. Westbrook JD, Burley SK. (2018) How Structural Biologists and the Protein Data Bank Contributed to Recent FDA New Drug Approvals. Structure 27: 211-217.

24. Kuhlbrandt W. (2014) Biochemistry. The resolution revolution. Science 343: 1443-1444.

25. Westbrook JD, Shao C, Feng Z, Zhuravleva M, Velankar S, 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.

26. Gilliland GL, Luo J, Vafa O, Almagro JC. (2012) Leveraging SBDD in protein therapeutic development: antibody engineering. Methods Mol Biol 841: 321-349.

27. Bollag G, Hirth P, Tsai J, Zhang J, Ibrahim PN, Cho H, Spevak W, Zhang C, Zhang Y, Habets G, Burton EA, Wong B, Tsang G, West BL, Powell B, Shellooe R, Marimuthu A, Nguyen H, Zhang KY, Artis DR, Schlessinger J, Su F, Higgins B, Iyer R, D'Andrea K, Koehler A, Stumm M, Lin PS, Lee RJ, Grippo J, Puzanov I, Kim KB, Ribas A, McArthur GA, Sosman JA, Chapman PB, Flaherty KT, Xu X, Nathanson KL, Nolop K. (2010) Clinical efficacy of a RAF inhibitor needs broad target blockade in BRAF-mutant melanoma. Nature 467: 596-599.

28. Santos R, Ursu O, Gaulton A, Bento AP, Donadi RS, Bologa CG, Karlsson A, Al-Lazikani B, Hersey A, Oprea TI, Overington JP. (2017) A comprehensive map of molecular drug targets. Nat Rev Drug Discov 16: 19-34.

29. Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Miyano M. (2000) Crystal structure of rhodopsin: A G protein-coupled receptor. Science 289: 739-745.

30. Fredriksson R, Lagerstrom MC, Lundin LG, Schioth HB. (2003) The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. Mol Pharmacol 63: 1256-1272.

31. Caffrey M, Cherezov V. (2009) Crystallizing membrane proteins using lipidic mesophases. Nat Protoc 4: 706-731.

32. Thorsen TS, Matt R, Weis WI, Kobilka BK. (2014) Modified T4 Lysozyme Fusion Proteins Facilitate G Protein-Coupled Receptor Crystallogenesis. Structure 22: 1657-1664.

33. Katritch V, Cherezov V, Stevens RC. (2013) Structure-function of the G protein-coupled receptor superfamily. Annu Rev Pharmacol Toxicol 53: 531-556.

34. Park SH, Das BB, Casagrande F, Tian Y, Nothnagel HJ, Chu M, Kiefer H, Maier K, De Angelis AA, Marassi FM, Opella SJ. (2012) Structure of the chemokine receptor CXCR1 in phospholipid bilayers. Nature 491: 779-783.

35. Zhang Y, Sun B, Feng D, Hu H, Chu M, Qu Q, Tarrasch JT, Li S, Sun Kobilka T, Kobilka BK, Skiniotis G. (2017) Cryo-EM structure of the activated GLP-1 receptor in complex with a G protein. Nature 546: 248-253.

36. Rasmussen SG, DeVree BT, Zou Y, Kruse AC, Chung KY, Kobilka TS, Thian FS, Chae PS, Pardon E, Calinski D, Mathiesen JM, Shah ST, Lyons JA, Caffrey M, Gellman SH, Steyaert J, Skiniotis G, Weis WI, Sunahara RK, Kobilka BK. (2011) Crystal structure of the beta2 adrenergic receptor-Gs protein complex. Nature 477: 549-555.

37. Rajeev SP, Wilding J. (2016) GLP-1 as a target for therapeutic intervention. Curr Opin Pharmacol 31: 44-49.

38. Purcell RH, Hall RA. (2018) Adhesion G Protein-Coupled Receptors as Drug Targets. Annu Rev Pharmacol Toxicol 58: 429-449.

39. Folts CJ, Giera S, Li T, Piao X. (2019) Adhesion G Protein-Coupled Receptors as Drug Targets for Neurological Diseases. Trends Pharmacol Sci 40: 278-293.

40. Yu FH, Catterall WA. (2004) The VGL-chanome: a protein superfamily specialized for electrical signaling and ionic homeostasis. Sci STKE 2004: re15.

41. Zhang AH, Sharma G, Undheim EAB, Jia X, Mobli M. (2018) A complicated complex: Ion channels, voltage sensing, cell membranes and peptide inhibitors. Neurosci Lett 679: 35-47.

42. Jiang Y, Lee A, Chen J, Ruta V, Cadene M, Chait BT, MacKinnon R. (2003) X-ray structure of a voltage-dependent K+ channel. Nature 423: 33-41.

43. Doyle DA, Morais Cabral J, Pfuetzner RA, Kuo A, Gulbis JM, Cohen SL, Chait BT, MacKinnon R. (1998) The structure of the potassium channel: molecular basis of K+ conduction and selectivity. Science 280: 69-77.

44. de Lera Ruiz M, Kraus RL. (2015) Voltage-Gated Sodium Channels: Structure, Function, Pharmacology, and Clinical Indications. J Med Chem 58: 7093-7118.

45. Xu L, Ding X, Wang T, Mou S, Sun H, Hou T. (2019) Voltage-gated sodium channels: structures, functions, and molecular modeling. Drug Discov Today 24: 1389-1397.

46. Bagal SK, Marron BE, Owen RM, Storer RI, Swain NA. (2015) Voltage gated sodium channels as drug discovery targets. Channels (Austin) 9: 360-366.

47. Payandeh J, Scheuer T, Zheng N, Catterall WA. (2011) The crystal structure of a voltage-gated sodium channel. Nature 475: 353-358.

48. Pan X, Li Z, Zhou Q, Shen H, Wu K, Huang X, Chen J, Zhang J, Zhu X, Lei J, Xiong W, Gong H, Xiao B, Yan N. (2018) Structure of the human voltage-gated sodium channel Nav1.4 in complex with beta1. Science 362: eaau2486.

49. Pan X, Li Z, Huang X, Huang G, Gao S, Shen H, Liu L, Lei J, Yan N. (2019) Molecular basis for pore blockade of human Na(+) channel Nav1.2 by the mu-conotoxin KIIIA. Science 363: 1309-1313.

50. Kingwell K. (2019) Nav1.7 withholds its pain potential. Nat Rev Drug Discov 18: 321-323.

51. Yang Y, Wang Y, Li S, Xu Z, Li H, Ma L, Fan J, Bu D, Liu B, Fan Z, Wu G, Jin J, Ding B, Zhu X, Shen Y. (2004) Mutations in SCN9A, encoding a sodium channel alpha subunit, in patients with primary erythermalgia. J Med Genet 41: 171-174.

52. Cox JJ, Reimann F, Nicholas AK, Thornton G, Roberts E, Springell K, Karbani G, Jafri H, Mannan J, Raashid Y, Al-Gazali L, Hamamy H, Valente EM, Gorman S, Williams R, McHale DP, Wood JN, Gribble FM, Woods CG. (2006) An SCN9A channelopathy causes congenital inability to experience pain. Nature 444: 894-898.

53. Zhang C, Cantara W, Jeon Y, Musier-Forsyth K, Grigorieff N, Lyumkis D. (2019) Analysis of discrete local variability and structural covariance in macromolecular assemblies using Cryo-EM and focused classification. Ultramicroscopy 203: 170-180.

54. Roh SH, Hryc CF, Jeong HH, Fei X, Jakana J, Lorimer GH, Chiu W. (2017) Subunit conformational variation within individual GroEL oligomers resolved by Cryo-EM. Proc Natl Acad Sci U S A 114: 8259-8264.

55. Collingridge GL, Olsen RW, Peters J, Spedding M. (2009) A nomenclature for ligand-gated ion channels. Neuropharmacology 56: 2-5.

56. Unwin N. (2005) Refined structure of the nicotinic acetylcholine receptor at 4Å resolution. J Mol Biol 346: 967-989.

57. Nasiripourdori A, Taly V, Grutter T, Taly A. (2011) From toxins targeting ligand gated ion channels to therapeutic molecules. Toxins (Basel) 3: 260-293.

58. Olsen RW, Li GD, Wallner M, Trudell JR, Bertaccini EJ, Lindahl E, Miller KW, Alkana RL, Davies DL. (2014) Structural models of ligand-gated ion channels: sites of action for anesthetics and ethanol. Alcohol Clin Exp Res 38: 595-603.

59. Rammes G, Rupprecht R. (2007) Modulation of ligand-gated ion channels by antidepressants and antipsychotics. Mol Neurobiol 35: 160-174.

60. Fox LM. (2006) Ivermectin: uses and impact 20 years on. Curr Opin Infect Dis 19: 588-593.

61. Mayer ML. (2006) Glutamate receptors at atomic resolution. Nature 440: 456-462.

62. Traynelis SF, Wollmuth LP, McBain CJ, Menniti FS, Vance KM, Ogden KK, Hansen KB, Yuan H, Myers SJ, Dingledine R. (2010) Glutamate receptor ion channels: structure, regulation, and function. Pharmacol Rev 62: 405-496.

63. Kumar J, Mayer ML. (2013) Functional insights from glutamate receptor ion channel structures. Annu Rev Physiol 75: 313-337.

64. Sobolevsky AI, Rosconi MP, Gouaux E. (2009) X-ray structure, symmetry and mechanism of an AMPA-subtype glutamate receptor. Nature 462: 745-756.

65. Wang JX, Furukawa H. (2019) Dissecting diverse functions of NMDA receptors by structural biology. Curr Opin Struct Biol 54: 34-42.

66. Chen S, Gouaux E. (2019) Structure and mechanism of AMPA receptor - auxiliary protein complexes. Curr Opin Struct Biol 54: 104-111.

67. Carruthers NI, Lovenberg TW, Traynelis SF. (2019) Allosteric Modulation of Ionotropic Glutamate Receptors Special Issue. ACS Med Chem Lett 10: 226.

68. Alam S, Lingenfelter KS, Bender AM, Lindsley CW. (2017) Classics in Chemical Neuroscience: Memantine. ACS Chem Neurosci 8: 1823-1829.

69. Lu W, Du J, Goehring A, Gouaux E. (2017) Cryo-EM structures of the triheteromeric NMDA receptor and its allosteric modulation. Science 355.

70. Redziniak G, Leclerc M, Panijel J, Monsigny M. (1978) Separation of two different populations of axial organ cells of Asterias rubens by the use of lectins. Biochimie 60: 525-527.

71. Saier MH, Jr., Reddy VS, Tsu BV, Ahmed MS, Li C, Moreno-Hagelsieb G. (2016) The Transporter Classification Database (TCDB): recent advances. Nucleic Acids Res 44: D372-379.

72. Zhang Y. (2018) Overview of Transporters in Pharmacokinetics and Drug Discovery. Curr Protoc Pharmacol 82: e46.

73. Wilkens S. (2015) Structure and mechanism of ABC transporters. F1000Prime Rep 7: 14.

74. Beis K. (2015) Structural basis for the mechanism of ABC transporters. Biochem Soc Trans 43: 889-893.

75. Locher KP, Lee AT, Rees DC. (2002) The E. coli BtuCD structure: a framework for ABC transporter architecture and mechanism. Science 296: 1091-1098.

76. Khare D, Oldham ML, Orelle C, Davidson AL, Chen J. (2009) Alternating access in maltose transporter mediated by rigid-body rotations. Mol Cell 33: 528-536.

77. Oldham ML, Chen J. (2011) Crystal structure of the maltose transporter in a pretranslocation intermediate state. Science 332: 1202-1205.

78. Oldham ML, Khare D, Quiocho FA, Davidson AL, Chen J. (2007) Crystal structure of a catalytic intermediate of the maltose transporter. Nature 450: 515-521.

79. Hennessy M, Spiers JP. (2007) A primer on the mechanics of P-glycoprotein the multidrug transporter. Pharmacol Res 55: 1-15.

80. Aller SG, Yu J, Ward A, Weng Y, Chittaboina S, Zhuo R, Harrell PM, Trinh YT, Zhang Q, Urbatsch IL, Chang G. (2009) Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. Science 323: 1718-1722.

81. Stoll F, Goller AH, Hillisch A. (2011) Utility of protein structures in overcoming ADMET-related issues of drug-like compounds. Drug Discov Today 16: 530-538.

82. Liu F, Zhang Z, Levit A, Levring J, Touhara KK, Shoichet BK, Chen J. (2019) Structural identification of a hotspot on CFTR for potentiation. Science 364: 1184-1188.

83. Elborn JS. (2016) Cystic fibrosis. Lancet 388: 2519-2531.

84. Ramsey BW, Davies J, McElvaney NG, Tullis E, Bell SC, Drevinek P, Griese M, McKone EF, Wainwright CE, Konstan MW, Moss R, Ratjen F, Sermet-Gaudelus I, Rowe SM, Dong Q, Rodriguez S, Yen K, Ordonez C, Elborn JS, Group VXS. (2011) A CFTR potentiator in patients with cystic fibrosis and the G551D mutation. N Engl J Med 365: 1663-1672.

85. Hediger MA, Clemencon B, Burrier RE, Bruford EA. (2013) The ABCs of membrane transporters in health and disease (SLC series): introduction. Mol Aspects Med 34: 95-107.

86. Bai X, Moraes TF, Reithmeier RAF. (2017) Structural biology of solute carrier (SLC) membrane transport proteins. Mol Membr Biol 34: 1-32.

87. Abramson J, Smirnova I, Kasho V, Verner G, Kaback HR, Iwata S. (2003) Structure and mechanism of the lactose permease of Escherichia coli. Science 301: 610-615.

88. Cesar-Razquin A, Snijder B, Frappier-Brinton T, Isserlin R, Gyimesi G, Bai X, Reithmeier RA, Hepworth D, Hediger MA, Edwards AM, Superti-Furga G. (2015) A Call for Systematic Research on Solute Carriers. Cell 162: 478-487.

89. Rives ML, Javitch JA, Wickenden AD. (2017) Potentiating SLC transporter activity: Emerging drug discovery opportunities. Biochem Pharmacol 135: 1-11.

90. Immadisetty K, Geffert LM, Surratt CK, Madura JD. (2013) New design strategies for antidepressant drugs. Expert Opin Drug Discov 8: 1399-1414.

91. Yamashita A, Singh SK, Kawate T, Jin Y, Gouaux E. (2005) Crystal structure of a bacterial homologue of Na+/Cl--dependent neurotransmitter transporters. Nature 437: 215-223.

92. Coleman JA, Green EM, Gouaux E. (2016) X-ray structures and mechanism of the human serotonin transporter. Nature 532: 334-339.

93. Bulatov E, Ciulli A. (2015) Targeting Cullin-RING E3 ubiquitin ligases for drug discovery: structure, assembly and small-molecule modulation. Biochem J 467: 365-386.

94. Komander D. (2009) The emerging complexity of protein ubiquitination. Biochem Soc Trans 37: 937-953.

95. Zheng N, Shabek N. (2017) Ubiquitin Ligases: Structure, Function, and Regulation. Annu Rev Biochem 86: 129-157.

96. Shabek N, Zheng N. (2014) Plant ubiquitin ligases as signaling hubs. Nat Struct Mol Biol 21: 293-296.

97. Asatsuma-Okumura T, Ito T, Handa H. (2019) Molecular mechanisms of cereblon-based drugs. Pharmacol Ther: doi: 10.1016/j.pharmthera.2019.1006.1004.

98. Sakamoto KM, Kim KB, Kumagai A, Mercurio F, Crews CM, Deshaies RJ. (2001) Protacs: chimeric molecules that target proteins to the Skp1-Cullin-F box complex for ubiquitination and degradation. Proc Natl Acad Sci U S A 98: 8554-8559.

99. Gadd MS, Testa A, Lucas X, Chan KH, Chen W, Lamont DJ, Zengerle M, Ciulli A. (2017) Structural basis of PROTAC cooperative recognition for selective protein degradation. Nat Chem Biol 13: 514-521.

100. Nowak RP, DeAngelo SL, Buckley D, He Z, Donovan KA, An J, Safaee N, Jedrychowski MP, Ponthier CM, Ishoey M, Zhang T, Mancias JD, Gray NS, Bradner JE, Fischer ES. (2018) Plasticity in binding confers selectivity in ligand-induced protein degradation. Nat Chem Biol 14: 706-714.

101. Mishin A, Gusach A, Luginina A, Marin E, Borshchevskiy V, Cherezov V. (2019) An outlook on using serial femtosecond crystallography in drug discovery. Expert Opin Drug Discov 14: 933-945.

102. Shen H, Liu D, Wu K, Lei J, Yan N. (2019) Structures of human Nav1.7 channel in complex with auxiliary subunits and animal toxins. Science 363: 1303-1308.

103. Walden H, Podgorski MS, Huang DT, Miller DW, Howard RJ, Minor DL, Jr., Holton JM, Schulman BA. (2003) The structure of the APPBP1-UBA3-NEDD8-ATP complex reveals the basis for selective ubiquitin-like protein activation by an E1. Mol Cell 12: 1427-1437.

104. Hamilton KS, Ellison MJ, Barber KR, Williams RS, Huzil JT, McKenna S, Ptak C, Glover M, Shaw GS. (2001) Structure of a conjugating enzyme-ubiquitin thiolester intermediate reveals a novel role for the ubiquitin tail. Structure 9: 897-904.

105. Zheng N, Schulman BA, Song L, Miller JJ, Jeffrey PD, Wang P, Chu C, Koepp DM, Elledge SJ, Pagano M, Conaway RC, Conaway JW, Harper JW, Pavletich NP. (2002) Structure of the Cul1-Rbx1-Skp1-F boxSkp2 SCF ubiquitin ligase complex. Nature 416: 703-709.

106. Schulman BA, Carrano AC, Jeffrey PD, Bowen Z, Kinnucan ER, Finnin MS, Elledge SJ, Harper JW, Pagano M, Pavletich NP. (2000) Insights into SCF ubiquitin ligases from the structure of the Skp1-Skp2 complex. Nature 408: 381-386.