



# RCSB Protein Data Bank tools for 3D structure-guided cancer research: human papillomavirus (HPV) case study

David S. Goodsell <sup>1,2</sup> · Stephen K. Burley <sup>1,3,4,5</sup>

Received: 5 July 2020 / Revised: 30 July 2020 / Accepted: 4 September 2020 / Published online: 16 September 2020  
© The Author(s) 2020. This article is published with open access

## Abstract

Atomic-level three-dimensional (3D) structure data for biological macromolecules often prove critical to dissecting and understanding the precise mechanisms of action of cancer-related proteins and their diverse roles in oncogenic transformation, proliferation, and metastasis. They are also used extensively to identify potentially druggable targets and facilitate discovery and development of both small-molecule and biologic drugs that are today benefiting individuals diagnosed with cancer around the world. 3D structures of biomolecules (including proteins, DNA, RNA, and their complexes with one another, drugs, and other small molecules) are freely distributed by the open-access Protein Data Bank (PDB). This global data repository is used by millions of scientists and educators working in the areas of drug discovery, vaccine design, and biomedical and biotechnology research. The US Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) provides an integrated portal to the PDB archive that streamlines access for millions of worldwide PDB data consumers worldwide. Herein, we review online resources made available free of charge by the RCSB PDB to basic and applied researchers, healthcare providers, educators and their students, patients and their families, and the curious public. We exemplify the value of understanding cancer-related proteins in 3D with a case study focused on human papillomavirus.

## Introduction

Function follows form in biology. Knowing the structure of a protein (or nucleic acid) is crucial for understanding how

it works in nature, how it may affect plant, animal, or human health, and how it could be targeted (or harnessed) to improve the human condition. The Protein Data Bank (PDB) is the global archive of three-dimensional (3D) biomolecular structure data [1, 2]. Founded in 1971 with just seven protein structures as the first open-access digital data resource in all of biology, the PDB today houses more than 168,000 structures of proteins, DNA, RNA, macromolecular machines, viruses, and virus-like particles. The US-funded Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB; [RCSB.org](https://www.rcsb.org)), a founding member of the Worldwide Protein Data Bank (wwPDB) partnership [3], integrates the entire corpus of PDB data with ~40 external biodata resources, and provides easy-to-use web-based search and molecular visualization tools to assist many millions of PDB data consumers worldwide in finding, analyzing, and visualizing 3D structures of macromolecules and their complexes with one another, drugs, antibodies, enzyme cofactors, etc.

This wealth of structural information is particularly useful in the field of cancer biology, wherein changes in DNA sequence that manifest themselves at the level of protein 3D structure and biochemical function can have

---

✉ David S. Goodsell  
David.Goodsell@rcsb.org

✉ Stephen K. Burley  
Stephen.Burley@rcsb.org

<sup>1</sup> Research Collaboratory for Structural Bioinformatics Protein Data Bank and Institute for Quantitative Biomedicine, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA

<sup>2</sup> Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA 92037, USA

<sup>3</sup> Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA

<sup>4</sup> Research Collaboratory for Structural Bioinformatics Protein Data Bank, San Diego Supercomputer Center, and the Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, CA 92093, USA

<sup>5</sup> Rutgers Cancer Institute of New Jersey, Rutgers, The State University of New Jersey, New Brunswick, NJ 08903, USA

extreme consequences for human health and disease [4]. Structural information is also central in the search for new approaches to diagnostics and therapeutic interventions, as we seek to block or modify these changes and their oncogenic effects [5, 6]. The PDB archive currently holds structural information covering all aspects of cancer biology, ranging from the molecular details of oncogenic mutations, to mechanisms of important cellular processes such as apoptosis, to structural characterization of the molecular machines underlying organism-scale processes such as neovascularization and metastasis. For the avoidance of doubt, current PDB holdings include 3D structures of the protein targets for >90% of the 79 new antineoplastic agents [54 small molecules, 25 biologics] approved by US Food and Drug Administration 2010–2018 [6]. In this review article, we use human papillomavirus (HPV) as a case study to highlight several diverse examples from this continuously growing corpus of 3D biostructure information, and the powerful tools available from the RCSB PDB for making effective use of them in the field of cancer research.

## Fundamental biology: revealing and understanding the HPV proteome

HPV are nonenveloped double-strand DNA viruses that infect epithelial cells. Infections with most HPV types lead to self-limiting benign lesions (a.k.a. warts), but several sexually transmitted genital high-risk types (e.g., HPV16 and HPV18) cause cervical carcinomas—the leading cause of death among female cancer patients worldwide [7, 8]—and some head and neck cancers [9]. The HPV genome encodes two classes of proteins: six early nonstructural regulatory proteins (denoted with “E” names), and two late structural proteins L1 and L2 [10]. Macromolecular crystallography (MX), NMR spectroscopy (NMR), and, increasingly, cryo-electron microscopy (3DEM) have all been used to determine 3D structures of viral proteins of HPV and related papillomaviruses, and their interactions with host proteins (Table 1). Some of these structures provide critical insights into the architecture of the papillomavirus capsid, composed of the L1 and L2 proteins, and explain how HPV virus-like particles can elicit an immune response and be recognized by antibodies that neutralize the virus. Other 3D structures reveal the atomic details pertaining to the function of the HPV early proteins, including those of E1 and E2 and their roles in viral replication, and those of E6 and E7 as they recognize and bind to intracellular proteins and frustrate host tumor suppressors, leading to oncogenic transformation.

As the PDB archive is growing at the rate of about 10% per year, it has become increasingly challenging to navigate

**Table 1** Structures of papillomavirus biomolecules in the PDB.

| Experimental method                      |                                     |  |
|--|-------------------------------------|--|
| 63                                       | Macromolecular crystallography (MX) |  |
| 14                                       | Solution NMR (NMR)                  |  |
| 14                                       | Electron microscopy (3DEM)          |  |
| Structures of papillomavirus proteins    |                                     |  |
| 28                                       | L1                                  | 13 icosahedral capsids and 10 complexed with monoclonal antibody |
| 8  | E1                                  | 3 complexed with DNA   |
| 25                                       | E2                                  | 4 complexed with DNA   |
| 22                                       | E6                                  | 17 complexed with host protein                                   |
| 8  | E7                                  | 3 complexed with host protein                                    |
| Structures related to oncogenic subtypes |                                     |  |
| 13                                       | HPV16                               |  |
| 11                                       | HPV18                               |  |

The results from advanced search (June 08 2020) of “papillomavirus” in “Source Organism Taxonomy Name,” with Boolean AND of additional “Full Text” terms such as “L1.”

and utilize the available holdings. To ensure that the PDB archive is maximally utilized, the wwPDB and the RCSB PDB are committed to the findability, accessibility, interoperability, and reusability principles [11] emblematic of responsible data resource management. Of critical importance, PDB structural information is available open access from members of the wwPDB partnership with no limitations on usage. Building on this freely available data repository, the RCSB PDB provides a wide range of search, analysis, and molecular visualization tools to provide nimble, multimodal access to >166,000 structures. All 3D structures coming into the PDB are processed by the wwPDB global deposition, validation, and biocuration system known as OneDep [12]. Every one of these structures is validated against both experimental data and established stereochemistry to provide users with quantitative estimates of structure quality and accuracy [13, 14]. In addition, every one of these structures is annotated by a professional wwPDB biocurator to a common data standard that was established by the wwPDB in consultation with community stakeholders [15]. Strict compliance to the PDBx/mmCIF data standard [16], rigorous structure validation, and expert biocuration ensures that our data consumers who are not experts in structural biology can rely on the information they download. The PDB has been recognized as a Core Certified Repository by CoreTrustSeal ([coretrustseal.org](http://coretrustseal.org)). This international, community-based, nongovernmental, nonprofit organization promotes sustainable and trustworthy data infrastructures of which the PDB is widely regarded as a gold-standard exemplar.

Search tools are optimized to help PDB data consumers (hereafter users) find molecules that are relevant to a given research question. These tools are hierarchical, allowing users to apply increasing levels of specificity as needed. Most begin with the general search box that is prominently displayed at the top of the RCSB PDB website home page ([rcsb.org](http://rcsb.org)), which combines the open source Apache Solr platform with indexing of all PDB data. It provides a listing of hits scored and ordered by relevance to the user's search term. For example, a search of "papillomavirus" in the "Source Organism Taxonomy Name" category yields 91 entries, including HPV proteins and nucleic acids, molecules from related papillomaviruses, and host proteins associated with HPV proteins. Users may then turn to Refinements to narrow this group of search hits. In Fig. 1, filtering by "Human papillomavirus type 16" narrowed the list to structures related to this particular oncogenic subtype. A flexible Advanced Search Query Builder allows interactive construction of Boolean Operator combinations of searches on a variety of subject fields, including polymer sequence, sequence motif, structure similarity, and chemical structure. This finer-grained searching for topics was used to obtain listings of each of the viral proteins summarized in Table 1. Finally, the resultant lists may be examined using a variety of textual and graphical reports, which are linked to detailed *Structure Summary* pages for each PDB structure.

Given the state of structural biology and the enormity of the PDB archive, typical searches will return dozens of structures related to a given topic. Each PDB structure has a dedicated *Structure Summary* page that provides a telegraphic overview, which is particularly useful when filtering a selection of structures for use in a given application. These pages deliver the major features of each PDB structure, identified with a unique four-character code (e.g., PDB ID 1abc), including a static image created with the RCSB PDB Mol\* web-native molecular graphics tool [17]; structure depositor(s), release date and primary publication; structure validation and accuracy assessments; and basic information on the chemical and structural features of the entry. For example, we used these pages to filter through many entries when researching a feature on viral quasi-symmetry (<http://pdb101.rcsb.org/motm/200>). The "Global Stoichiometry" field [18] underscores one of the mysteries of papillomaviruses that was revealed by the structural biologists (e.g., PDB ID 3j6r [19]). They are "homo-360-mers," which places a surprising value of 6 subunits in the repeating unit of the icosahedral symmetric virus (Fig. 2). This multiple of 60 (*i.e.*,  $60 \times 6 = 360$ ) does not conform to the classic system of quasi-symmetry, which would require 3, 4, or 7 subunits to be consistent with the conception of a distorted triangular tessellation of a virus this size [20]. The PDB ID 3j6r structure shows that this virus, and similar polyoma viruses such as simian virus 40 [21], instead place

pentamers at locations normally occupied by hexamers, and use flexible polypeptide chain segments to resolve the inconsistencies in sites of interaction between pentamers. This information is directly relevant to the design and engineering of second-generation virus-like particles that can be formulated as HPV vaccines to prevent cervical cancers [22].

## Structure and function: HPV form and flexibility

Looking at the dozens of structures available for papillomavirus capsids and individual proteins, we see many of the general themes that underlie biomolecular structure and function. For example, mechanisms of hierarchical assembly and self-association guide the construction of icosahedral capsids, as revealed in structures like that shown in Fig. 2 (PDB ID 3j6r). Transient association of viral proteins with host proteins and nucleic acids guide each step in the viral lifecycle and have been revealed at atomic detail. For example, PDB ID 5w1o [23] includes an L1 pentamer from HPV16 bound to oligosaccharides from the cellular heparin receptor, with the surprising observation that multiple sites of virus-receptor interaction are involved in viral attachment and entry. Intrinsically disordered proteins also play central roles in several intracellular processes, notably the oncogenic interaction of E6 and E7 proteins with disordered segments of host proteins, described in more detail below.

To explore these topics, the RCSB PDB website provides a collection of "Views" that leverage information from related sequence and structural resources, allowing users to drill deeper into the information held in each entry. The *Protein Feature* view, provided in summary form on the *Structure Summary* page and in more detail with one click, gathers data from UniProt and other external databases to assist users in understanding the context of each entry in the PDB archive. Figure 3 shows one major use of the *Protein Feature* view. Structural biologists often cut proteins into functional pieces when the full-length protein does not prove amenable to structure determination in its entirety. It can, therefore, be difficult to parse out exactly which polypeptide chain segments comprising a particular protein are present in a given PDB ID. The *Protein Feature* view for HPV16 E6 shows that the protein contains several functional domains, and structures are available for the whole protein and for two individual domains, as well as for a short peptide bound to the PDZ1 domain of cellular protein MAGI-1 [24].

Similar "views" enable exploration of other important topics. The *Small Molecule* view provides information on ligands bound to macromolecules in each structure. For example, PDB ID 2gxa [25], provides an atomic-level direct

**Fig. 1 Multiple options streamline searches of the PDB archive.** Here, HPV16 holdings were found by **(a)** searching for “papillomavirus” in the main search box, **(b)** selecting HPV16 in the refinement options, and **(c)** choosing a gallery display. Advanced searches may be built-in **(d)** for more specific queries. The search returns several L1 capsid structures, some coated with antibody Fab fragments, and structures of E2 and E6, some interacting with host cell proteins.

The screenshot displays the RCSB PDB search results page for the query "papillomavirus". The search results are displayed in a gallery view, showing 33 structures. The interface includes a search bar at the top, a navigation menu, and a search results section. The search results are filtered by "Scientific Name of Source Organism" to "Human papillomavirus type 16 (33)". The gallery shows various protein structures, including capsid structures and E2/E6 proteins, with their PDB IDs listed. The interface also includes a "Refinements" section on the left, which is used to filter the search results. The search results are displayed in a grid format, with each structure shown as a 3D ribbon model. The interface is in English and includes a "MyPDB" button in the top right corner.

look view at the nucleotide binding site of the E1 hexameric helicase assembly bound to segment of single-stranded DNA. The *Annotations* view includes third-party

annotations relating to domain structure and gene ontology (e.g., CATH defines the two domains of E1 as zinc fingers, similar in structure to that found in the large T-antigen D1

**Fig. 2 Structure Summary pages provide an overview for each entry in the PDB archive.** *Structure Summary* for a cryo-electron microscopy study of HPV16 capsid (PDB ID 3j6r [19]) provides an overview of the entry and many options to access detailed information and analysis tools.

**Structure Summary** | 3D View | Annotations | Sequence | Experiment

Biological Assembly 1

**3J6R**  
Electron cryo-microscopy of Human Papillomavirus Type 16 capsid  
DOI: 10.2210/pdb3J6R/pdb EMDDataResource: EMD-5932

Classification: **VIRUS**  
Organism(s): Human papillomavirus type 16  
Mutation(s): No

Deposition: 2014-03-20 Released: 2014-07-23  
Deposition Author(s): Cardone, G., Moyer, A.L., Cheng, N., Thompson, C.D., Dvoretzky, I., Lowy, D.R., Schiller, J.T., Steven, A.C., Buck, C.B., Trus, B.L.

Experimental Data Snapshot  
Method: ELECTRON MICROSCOPY  
Resolution: 9.10 Å  
Aggregation State: PARTICLE  
Reconstruction Method: SINGLE PARTICLE

wwPDB Validation  
Metric: Clashescore, Ramachandran outliers, Percentile Ranks, Value, 4.6%  
Percentile relative to all structures  
Percentile relative to all EM structures

This is version 1.3 of the entry. See complete history.

Literature  
Download Primary Citation

Maturation of the human papillomavirus 16 capsid.  
Cardone, G., Moyer, A.L., Cheng, N., Thompson, C.D., Dvoretzky, I., Lowy, D.R., Schiller, J.T., Steven, A.C., Buck, C.B., Trus, B.L.  
(2014) mBio 5: e01104-e01114  
PubMed: 25096873 Search on PubMed Search on PubMed Central  
DOI: 10.1128/mBio.01104-14  
Structures With Same Primary Citation

PubMed Abstract:  
Papillomaviruses are a family of nonenveloped DNA viruses that infect the skin or mucosa of their vertebrate hosts. The viral life cycle is closely tied to the differentiation of infected keratinocytes. Papillomavirus virions are released into the en ...

**Fig. 3 The Protein Feature view summarizes all PDB entries for a particular protein.** *Protein Feature* view for a complex of HPV16 E6 with guanylate kinase MAGI-1 (PDB ID 2kpl [24]) helps users understand that only a small peptide from E6 is included in the PDB ID. A similar View is available for MAGI-1, showing that only one domain from that protein is included in the entry.

**P03126**  
Protein E6 - Human papillomavirus type 16

( Mouse scroll to zoom / Hold left click to move )

ALL

UNIPROT SEQ - P03126  
CHAIN  
LINEAR MOTIF  
ZINC FINGER  
CONFLICT  
HOMOLOGY MODELS  
PDB ENTITY ALIGN - 2FK4\_1  
PDB ENTITY ALIGN - 2KPL\_2  
PDB ENTITY ALIGN - 2LJX\_1  
PDB ENTITY ALIGN - 2LJY\_1  
PDB ENTITY ALIGN - 2LJZ\_1  
PDB ENTITY ALIGN - 4GIZ\_2  
PDB ENTITY ALIGN - 4JOP\_2  
PDB ENTITY ALIGN - 4XR8\_3  
PDB ENTITY ALIGN - 6HKS\_2  
PDB ENTITY ALIGN - 6SIV\_2  
PDB ENTITY ALIGN - 6SJA\_2  
PDB ENTITY ALIGN - 6TWQ\_2  
PDB ENTITY ALIGN - 6TWU\_2  
PDB ENTITY ALIGN - 6TWX\_2

M I K R T A F Q I K E F F R R L P L C L E L G I T T I D I I L E E V Y K D Q L L R F E M Y F A R F L C I V F R D N P Y M C K D K L Q P Y S K I S E P R M C Y S L Y G T L E Q N R P L C L L I R C I N C K P L C E E N D R L D G C R P A I R E M V G R D E D C S S F P R F E Q L

domain, and a so-called Rossmann fold [26]). Further details of the E1 protein sequence itself may be explored in the *Sequence* view, with annotations of secondary structure and other features, such as the nucleotide binding site. Details concerning the structure determination process are tabulated in the *Experiment* view. Finally, quantitative analyses of structure quality can be found in the wwPDB Validation Report, which is summarized graphically near

the top of the *Structure Summary* page and available both in 3D (click 3D Report) and downloadable pdf file (click Full Report).

Structure determinations typically provide “snapshots” of macromolecules adopting a single conformational state. For relatively simple proteins consisting of a single globular domain (e.g., sperm whale myoglobin PDB ID 1mbn [27], the first atomic-level protein structure to be determined

#### Fig. 4 Sequence and structure alignments allow analysis of conservation and flexibility.

Comparison of L1 proteins from pentamers of HPV11 (cyan, PDB ID 2r5k [31]) and HPV18 (orange, PDB ID 2r5i [31]), using the Java Web Start “Structure Alignment” tool. Arrows indicate sites with changes in the length of loops, seen as gaps in the sequence alignment at the top.

```

■EQR:415 Len1:419 Len2:423 score: 946.13 Z-score:8.03 RMSD:0.95 SeqID:63% SeqSim:78% Cov1:99% Cov2:98%
19:A  AVVATDAYVKRTNI FYHASSRLLAVGHPYYSI KKVNK T - - - - - VVPKVS GYQYRVFKVVL PDPNKFAL 82:A
20:A  AVVNTDDYVTRTSI FYHAGS RLLTVGDPYFRVP - - - - - AGGKNKQDI PKKVSATQYRVERVQLPDPNKFGL 85:A

83:A  PDS S L F D P T T Q R L V W A C T G L E V G R G Q P L G V G V S G H P L L N K Y D D V E N S G G Y G G N P G Q D N R V N V G M D Y K Q T Q 152:A
86:A  P D T S I Y N P E T Q R L V W A C A G V E I G R G Q P L G V G L S G H P F Y N K L D D T E S S H A A T S N V S E D V R D N V S V D Y K Q T Q 155:A

153:A L C M V G C A P L P G E H W K G T Q S N T S V Q N G D C P P L E L I T S V I Q D G D M V D T G F G A M N F A D L Q T N K S D V P L D I C 222:A
156:A L C I L G C A P A I G E H W A K G T A S K S R P L S Q G D C P P L E L K N T V L E D G D M V D T G Y G A M D F S T L Q D T K C E V P L D I C 225:A

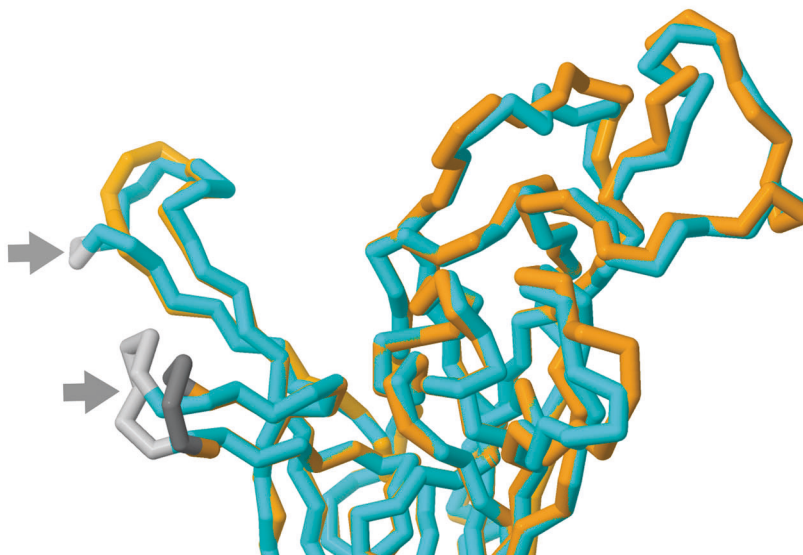
223:A G T V C K Y P D Y L Q M A A D P Y G D R L F F Y L R K E Q M F A R H F F N R A G T V G E P V P D D L L V K G G N N R S S V A S S I Y V H T P 292:A
226:A Q S I C K Y P D Y L Q M S A D P Y G D S M F F C L R R E Q L F A R H F W N R A G T M G D T V P Q S L Y I K G T G M R A S P G S C V Y S P S P 295:A

293:A S G S L V S S E A Q L F N K P Y W L Q K A Q G H N N G I C W G N H L S V T V V D T T R S T N M T L C A S V S - - K S A T Y T N S D Y K E Y M 360:A
296:A S G S I V T S D S Q L F N K P Y W L H K A Q G H N N G V C W H N Q L F V T V V D T T R S T N L T I C A S T Q S P V P G Q Y D A T K F K Q Y S 365:A

361:A R H V E E F D L Q F I F Q L C S I T L S A E V M A Y I H T M N P S V L E D W N K Q D P Y K D M S F W E V N L K E K F S S E L D Q F L G R K 463:A
366:A R H V E E Y D L Q F I F Q L C T I T L T A D V M S Y I Q S M N S S I L E D W N N K D P Y D K L K F W N V D L K E K F S L D L D Q Y L G R K 468:A

464:A F L L Q S G Y 470:A
469:A F L V Q A G L 475:A

```



[28]), this is not usually a major consideration. For more complicated macromolecular systems consisting of more than one globular domain (e.g., the multi-domain Abl protein kinase PDB ID 1fpu [29]) or multi-protein complexes (e.g., the CDK2/Cyclin A binary complex PDB ID 1fin [30]) a conceptual model of conformational flexibility must be built up by gathering structures in different states and comparing them. This challenge is further complicated by the fact that structural biologists often gather information from multiple viral strains and multiple host organisms, so the overall framework must be built using 3D structures from disparate sources. The *Protein Comparison Tool*, accessible using Java Web Start (<http://www.rcsb.org/pdb/workbench/workbench.do?action=menu>), is a critical tool for relating and comparing different structures. It provides several turnkey methods for pairwise sequence and structure alignments. For example, Fig. 4 includes alignment of capsid L1 protein structures from benign and high-risk strains, showing how small changes in the sequences of polypeptide chain loops on the viral surface lead to conformational differences, and ultimately to differences in the way they are recognized by the immune system [31].

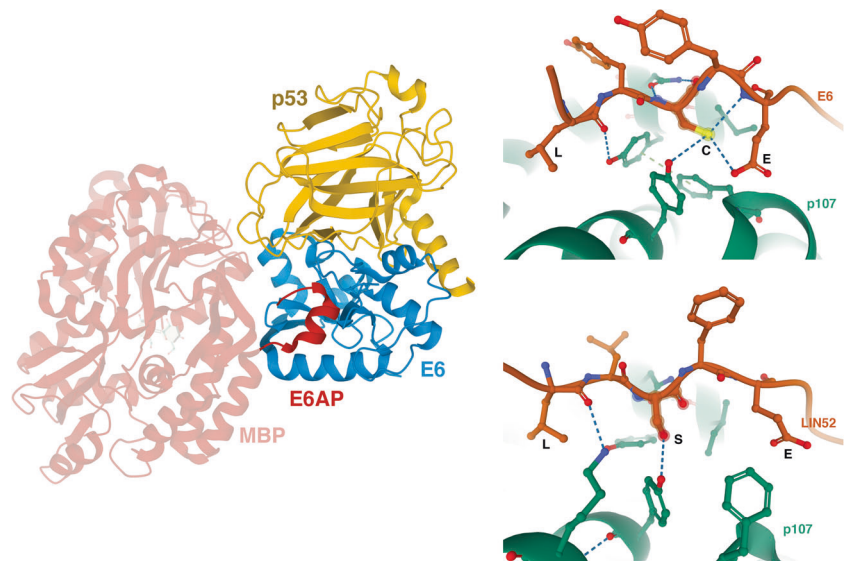
#### Visualizing functional interactions: oncogenesis and epitopes

Structural biology allows us to overcome the limitations of the human eye to “see” directly the molecular processes that underlie viral oncogenesis and immune response. A wide variety of visualization systems are available to help researchers explore and make sense of such data. These tools include highly optimized stand-alone software packages such as Chimera [32] and VMD [33], which typically have built-in options to fetch structures from the PDB archive based on the PDB ID. The principal limitation of these molecular graphics systems is the need to download software to the user’s computer and keep pace with frequent updates.

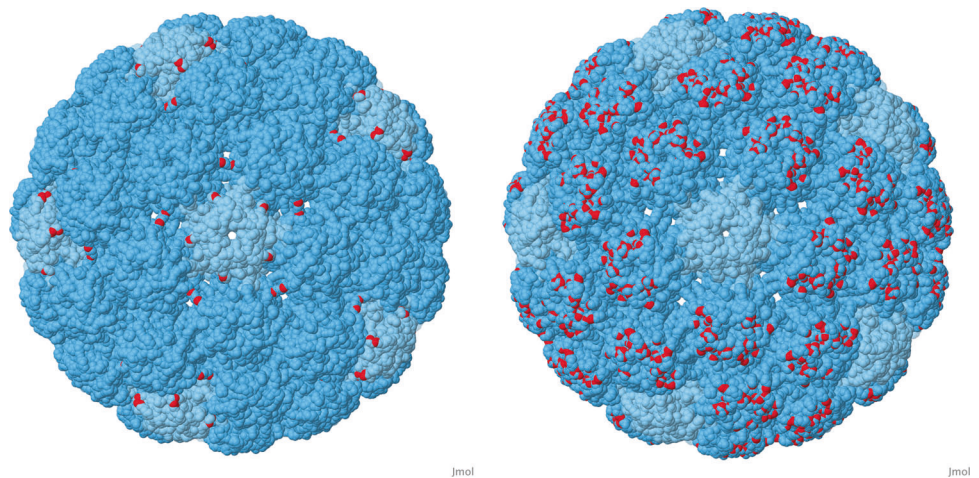
The RCSB PDB website provides several options for on-demand visualization, to simplify exploration of structures on-the-fly as the archive is being searched. The most powerful of these tools is the RCSB PDB Mol\* web-native molecular visualization tool with many options for customizing views and molecular representations [17]. It has been highly optimized to allow interactive loading and

**Fig. 5 Visualization of virus-host interactions with Mol\*.**

Left: HPV E6 brings together E6AP and p53 tumor suppressor (PDB ID 4xr8 [34]). Only portions of E6AP and p53 are included in the structure determination, and E6AP was studied as a chimera with maltose binding protein. Right: interaction of p107 with E6 and with a suboptimal host partner (PDB IDs 4yos, 4yoz [35]).



**Fig. 6 Visualization of epitopes with JSmol.** Epitopes (red) of antibodies U4 (left) and V5 (right) on HPV16 capsids, visualized with online JSmol scripting tools (PDB IDs 6bsp, 6bt3 [37]).



viewing of the large structures that are increasingly being deposited to the PDB archive. In Fig. 5, Mol\* was used to explore the mechanisms of oncogenesis in two PDB structures, wherein viral proteins E6 and E7 are frustrating host defenses. HPV E6 acts as an adapter protein, bringing together defense proteins such as the p53 tumor suppressor and E6AP, which leads to recruitment of the ubiquitin/proteasome system. PDB ID 4xr8 [34] illuminates how the LxxLL motif of E6AP is recognized by HPV16 E6, targeting the p53 protein for degradation. Binding of the HPV16 E7 LxCxE motif to the Rb tumor suppressor paralog p107 is seen in a structure of an E7 peptide bound to the protein (PDB ID 4yoz [35]). This interaction blocks the Rb binding site involved in cell cycle signaling, as seen in a complex with LIN52 peptide (PDB ID 4yos [35]). LIN52 has an LxSxExL motif and a phosphorylated serine.

JSmol is also provided at the RCSB PDB website as a lighter-weight visualization option with a user-friendly natural language scripting language [36]. This scripting capability was useful for creation of Fig. 6, which shows antibody-binding epitopes for two structures of HPV virus-like particles. Information from the primary reports for two PDB structures (PDB IDs 6bsp and 6bt3) of HPV16 with bound monoclonal antibodies U4 and V5, respectively, was used to define their epitopes [37], and then scripted for display using JSmol. U4 has a discontinuous epitope (red in Fig. 6) occurring in a groove between pentamers at the fivefold axis (lighter blue in Fig. 6) and one of the neighboring quasi-symmetrical pentamers. In contrast, V5 binds at several positions around the quasi-symmetrical pentamers. Antibody-capsid complex structures can be used to guide second-generation vaccine design efforts,

**Fig. 7 Outreach materials introduce structural topics to a wide audience.** Outreach materials include a feature and a foldable paper model of HPV (red) bound to antibody Fab fragments (blue) (PDB ID 6bt3 [37]\*).

### Human Papillomavirus and Vaccines

*The capsid protein of papillomavirus is used in vaccines that prevent cervical cancer.*

Papillomaviruses are annoying pests that occasionally turn into deadly dangers. They attack cells in our skin and mucous membranes. When they infect cells, they ramp up the normal growth functions, often forming warts. Usually our defenses are able to get the infection under control, but in some exceptional cases, the virus persists and the unwanted growth can turn into cancer. Alarmingly, infection by a few particularly-virulent types of papillomavirus is the leading cause of cervical cancer. Fortunately, by studying these viruses, scientists have discovered highly effective ways to fight them.

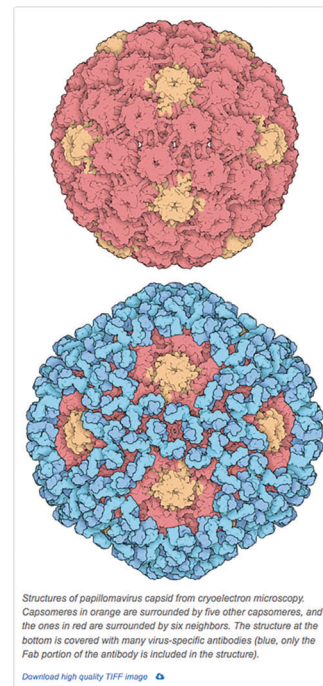
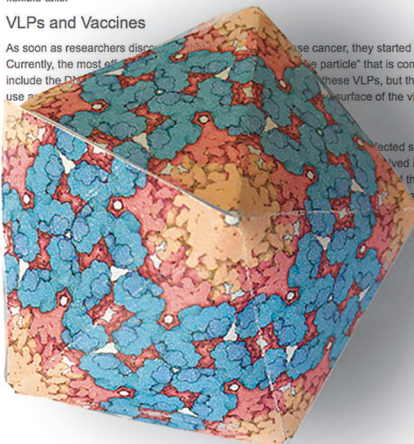
#### A Tiny Danger

Papillomavirus is a small virus, with a simple capsid surrounding a circular DNA genome. The capsid (PDB entry 3jfr) includes 360 copies of the major capsid chain, called L1. A second capsid chain, called L2, is found on the inside and may help with packaging the genome. The capsid structure, however, is not a typical quasisymmetrical virus. Instead, like simian virus 40, the L1 chains form 72 pentameric "capsomeres", which then interact with one another through long flexible tails.

#### VLPs and Vaccines

As soon as researchers discovered the structure of the virus, they started looking for vaccines to fight them. Currently, the most common vaccine is a virus-like particle (VLP) that is composed of the L1 protein, but doesn't include the DNA genome. These VLPs, but they are not infectious, so they are safe to use. The surface of the virus, making it unable to infect cells.

Scientists have developed several prophylactic vaccines to fight the virus. Two of the most common are for cervical cancer, types 16 and 18. The nonavalent vaccine adds an additional five capsids shown here are type 16



building on the success of currently approved anti-HPV vaccines [22].

### Outreach/education: disseminating the results of HPV research

The PDB archive is a unique resource for science education and outreach, given that structural information provides an intuitive and comprehensible window into difficult functional concepts in biology and medicine. RCSB PDB hosts a web portal, PDB-101 ([pdb101.rcsb.org](http://pdb101.rcsb.org)), that brings the results of structural biology to the education and lay communities [38]. As with the diverse user community of the main RCSB PDB portal, the educational and lay communities have a broad collection of needs, so PDB-101 provides multiple modalities for engaging users. A user-friendly browser is provided at the entry point that allows users to explore holdings based on common topics, such as biological energy, nanotechnology, or viruses. The highly popular *Molecule of the Month* feature presents a new topic each month, providing a short description of the structure, function, and relevance of selected molecules [39]. Links to structures in the PDB archive invite users to extend their reading by exploring the actual data. *Curriculum Modules* provide educational materials and lesson plans for popular topics in diabetes, immunology, and virology. In addition, a variety of posters, interactive animations, molecular origami paper-folding activities, and similar materials have been created to engage user communities at all levels of expertise.

For HPV, a *Molecule of the Month* feature was recently presented, describing the connection to cancer and how an understanding of HPV proteins can help discover new ways to fight viral infection. A molecular origami foldable paper model of the HPV16 virus-like particle decorated with Fab fragments of the V5 antibody (PDB ID 6bt3 [37]) is downloadable as part of this outreach effort (Fig. 7).

### The RCSB PDB and cancer research writ large

The RCSB PDB is committed to empowering cancer researchers, with advanced tools for exploring biomedically relevant structure holdings, and extensive introductory resources to lower the barrier to entry for users who are new to structural biology (<https://www.rcsb.org/pages/help/index>). For clinical researchers the RCSB PDB provides easy-to-use tools that enable discovery of cancer-related proteins and biomarkers; support in-depth 3D analyses of mutational hotspots identified via comprehensive genomic sequencing/profiling; and facilitate hypothesis generation regarding selection of targeted antineoplastic agents. The ability to combine 3D structure data with cancer-related information will complement well-established approaches that principally utilize 1D genome/protein sequence data. One of the most important challenges facing clinical researchers today is acquired resistance to targeted antineoplastic agents. Resistance is observed in many tumor types and can appear during any stage of cancer treatment through a variety of biochemical and cell biological mechanisms (reviewed in [40]). The



RCSB PDB website also provides tools for understanding acquired drug resistance in 3D and developing testable hypotheses for alternative targeted therapies, as demonstrated for osimertinib treatment failure due to emergence of a previously unreported epidermal growth factor receptor mutation (methionine 766→glutamine) [41]. PDB data and the RCSB PDB website support early-stage oncology drug discovery (reviewed in [6]). Areas of demonstrated impact include target validation; druggability assessment; characterization of screening hits; medicinal chemistry optimization of pharmaceutically acceptable leads; and design of novel proteins for diagnostic and therapeutic applications (e.g., chimeric antigen receptors, bispecific antibodies). Finally, the RCSB PDB website supports basic and applied cancer researchers, whose work can benefit significantly from “direct looks” at 3D structures of human proteins, multi-protein complexes, and protein-nucleic acid complexes as they characterize the biochemical and cell biological origins of human cancers (reviewed in [6]).

**Acknowledgements** We gratefully acknowledge discussions with Drs. X.F. Steven Zheng and Shridar Ganesan and other members of the Rutgers Cancer Institute of New Jersey, and contributions from all members of the RCSB PDB and our wwPDB partners around the world. RCSB PDB is jointly funded by grants to SKB from the National Science Foundation (DBI-1832184), the US 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.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Berman HM. The Protein Data Bank. *Nucleic Acids Res.* 2000;28:235–42.
- Burley SK, Berman HM, Bhikadiya C, Bi C, Chen L, Di Costanzo L, et al. RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic Acids Res.* 2019;47:D464–74.
- wwPDB consortium, Burley SK, Berman HM, Bhikadiya C, Bi C, Chen L, et al. Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic Acids Res.* 2019;47:D520–8.
- Goodsell DS, Zardecki C, Di Costanzo L, Duarte JM, Hudson BP, Persikova I, et al. RCSB Protein Data Bank: enabling biomedical research and drug discovery. *Protein Sci.* 2020;29:52–65.
- Westbrook JD, Burley SK. How structural biologists and the Protein Data Bank contributed to recent FDA new drug approvals. *Structure.* 2019;27:211–7.
- Westbrook JD, Soskind R, Hudson BP, Burley SK. Impact of the Protein Data Bank on antineoplastic approvals. *Drug Discov Today.* 2020;24:837–50.
- McBride AA. Mechanisms and strategies of papillomavirus replication. *Biol Chem.* 2017;398:919–27.
- Mittal S, Banks L. Molecular mechanisms underlying human papillomavirus E6 and E7 oncoprotein-induced cell transformation. *Mutat Res Mutat Res.* 2017;772:23–35.
- Kobayashi K, Hisamatsu K, Suzui N, Hara A, Tomita H, Miyazaki T. A review of HPV-related head and neck cancer. *J Clin Med.* 2018;7:241.
- Graham SV. Human papillomavirus: gene expression, regulation and prospects for novel diagnostic methods and antiviral therapies. *Future Microbiol.* 2010;5:1493–506.
- Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data.* 2016;3:160018.
- Young JY, Westbrook JD, Feng Z, Sala R, Peisach E, Oldfield TJ, et al. OneDep: Unified wwPDB system for deposition, biocuration, and validation of macromolecular structures in the PDB archive. *Structure.* 2017;25:536–45.
- Gore S, Sanz García E, Hendrickx PMS, Gutmanas A, Westbrook JD, Yang H, et al. Validation of structures in the Protein Data Bank. *Structure.* 2017;25:1916–27.
- Shao C, Yang H, Westbrook JD, Young JY, Zardecki C, Burley SK. Multivariate analyses of quality metrics for crystal structures in the PDB archive. *Structure.* 2017;25:458–68.
- Young JY, Westbrook JD, Feng Z, Peisach E, Persikova I, Sala R, et al. Worldwide Protein Data Bank biocuration supporting open access to high-quality 3D structural biology data. *Database.* 2018;2018:bay002. <https://pubmed.ncbi.nlm.nih.gov/29688351/>.
- Adams PD, Afonine PV, Baskaran K, Berman HM, Berrisford J, Bricogne G, et al. Announcing mandatory submission of PDBx/mmCIF format files for crystallographic depositions to the Protein Data Bank (PDB). *Acta Crystallogr Sect Struct Biol.* 2019;75:451–4.
- Sehnal D, Rose A, Koca J, Burley S, Velankar S. Mol\*: Towards a common library and tools for Web Molecular Graphics. In: Byska J, Krone, M, Sommer B, (eds). Workshop on Molecular Graphics and Visual Analysis of Molecular Data. The Eurographics Association; 2018. <https://doi.org/10.2312/molva.20181103>.
- Korkmaz S, Duarte JM, Prlić A, Goksuluk D, Zarsarsiz G, Saracbası O, et al. Investigation of protein quaternary structure via stoichiometry and symmetry information. *PLoS One.* 2018;13:e0197176.
- Cardone G, Moyer AL, Cheng N, Thompson CD, Dvoretzky I, Lowy DR, et al. Maturation of the human papillomavirus 16 capsid. *mBio.* 2014;5:e01104–14.
- Caspar DLD, Klug A. Physical principles in the construction of regular viruses. *Cold Spring Harb Symp Quant Biol.* 1962;27:1–24.

21. Liddington RC, Yan Y, Moulai J, Sahli R, Benjamin TL, Harrison SC. Structure of simian virus 40 at 3.8-Å resolution. *Nature*. 1991;354:278–84.
22. World Health Organization. Human papillomavirus vaccines: WHO position paper, May 2017. *Wkly Epidemiol Rec*. 2017;19:241–69.
23. Dasgupta J, Bienkowska-Haba M, Ortega ME, Patel HD, Bodevin S, Spillmann D, et al. Structural basis of oligosaccharide receptor recognition by human papillomavirus. *J Biol Chem*. 2011;286:2617–24.
24. Charbonnier S, Nominé Y, Ramírez J, Luck K, Chapelle A, Stote RH, et al. The structural and dynamic response of MAGI-1 PDZ1 with noncanonical domain boundaries to the binding of human papillomavirus E6. *J Mol Biol*. 2011;406:745–63.
25. Enemark EJ, Joshua-Tor L. Mechanism of DNA translocation in a replicative hexameric helicase. *Nature*. 2006;442:270–5.
26. Rao ST, Rossmann MG. Comparison of super-secondary structures in proteins. *J Mol Biol*. 1973;76:241–56.
27. Watson HC. The stereochemistry of the protein myoglobin. *Prog Stereochem*. 1969;4:299–333.
28. Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, Phillips DC. A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature*. 1958;181:662–6.
29. Schindler T. Structural mechanism for STI-571 inhibition of Abelson tyrosine kinase. *Science*. 2000;289:1938–42.
30. Jeffrey PD, Russo AA, Polyak K, Gibbs E, Hurwitz J, Massagué J, et al. Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex. *Nature*. 1995;376:313–20.
31. Bishop B, Dasgupta J, Klein M, Garcea RL, Christensen ND, Zhao R, et al. Crystal structures of four types of human papillomavirus L1 capsid proteins: Understanding the specificity of neutralizing monoclonal antibodies. *J Biol Chem*. 2007;282:31803–11.
32. Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH, et al. UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci*. 2018;27:14–25.
33. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph*. 1996;14:33–8.
34. Martinez-Zapien D, Ruiz FX, Poirson J, Mitschler A, Ramirez J, Forster A, et al. Structure of the E6/E6AP/p53 complex required for HPV-mediated degradation of p53. *Nature*. 2016;529(541–5).
35. Guiley KZ, Liban TJ, Felthousen JG, Ramanan P, Litovchick L, Rubin SM. Structural mechanisms of DREAM complex assembly and regulation. *Genes Dev*. 2015;29:961–74.
36. Hanson RM. *Jmol*—a paradigm shift in crystallographic visualization. *J Appl Crystallogr*. 2010;43:1250–60.
37. Guan J, Bywaters S, Brendle S, Ashley R, Makhov A, Conway J, et al. High-resolution structure analysis of antibody V5 and U4 conformational epitopes on human papillomavirus 16. *Viruses*. 2017;9:374.
38. Dutta S, Zardecki C, Goodsell DS, Berman HM. Promoting a structural view of biology for varied audiences: an overview of RCSB PDB resources and experiences. *J Appl Crystallogr*. 2010;43:1224–9.
39. Goodsell DS, Dutta S, Zardecki C, Voigt M, Berman HM, Burley SK. The RCSB PDB “Molecule of the Month”: inspiring a molecular view of biology. *PLoS Biol*. 2015;13:e1002140.
40. Holohan C, Van Schaeybroeck S, Longley DB, Johnston PG. Cancer drug resistance: an evolving paradigm. *Nat Rev Cancer*. 2013;13:714–26.
41. Castellano GM, Aisner J, Burley SK, Vallat B, Yu HA, Pine SR, et al. A novel acquired exon 20 EGFR M766Q mutation in lung adenocarcinoma mediates osimertinib resistance but is sensitive to neratinib and poziotinib. *J Thorac Oncol*. 2019;14:1982–8.