SOFTWARE TOOL ARTICLE

PDBrt: A free database of complexes with measured drug-target residence time [version 1; peer review: awaiting peer review]

Magdalena Ługowska, Marcin Pacholczyk

Department of Systems Biology and Engineering, Silesian University of Technology, Akademicka 16, Gliwice, 44-100, Poland

Abstract

Background: Difficulties in translating the in vitro potency determined by cellular assays into in vivo efficacy in living organisms complicates the design and development of drugs. However, the residence time of a drug in its molecular target is becoming a key parameter in the design and optimization of new drugs, as recent studies show that residence time can reliably predict drug efficacy in vivo. Experimental approaches to binding kinetics and target ligand complex solutions are currently available, but known bioinformatics databases do not usually report information about the ligand residence time in its molecular target.

Methods: To extend existing databases we developed the Protein Data Bank (PDB) residence time database (PDBrt) which reports drug residence time. The database is implemented as an open access web-based tool. The front end uses Bootstrap with Hypertext Markup Language (HTML), jQuery for the interface and 3Dmol.js to visualize the complexes. The server-side code uses Python web application framework, Django Rest Framework and backend database PostgreSQL.

Results: The PDBrt database is a free, non-commercial repository for 3D protein-ligand complex data, including the measured ligand residence time inside the binding pocket of the specific biological macromolecules as deposited in The Protein Data Bank. The PDBrt database contains information about both the protein and the ligand separately, as well as the protein-ligand complex, binding kinetics, and time of the ligand residence inside the protein binding site.

Availability: https://pdbrt.polsl.pl

Keywords

protein-ligand binding, residence time, binding kinetics, web services, database
Corresponding author: Marcin Pacholczyk (marcin.pacholczyk@polsl.pl)

Author roles: Ługowska M: Conceptualization, Data Curation, Funding Acquisition, Methodology, Project Administration, Resources, Software, Validation, Visualization, Writing – Original Draft Preparation; Pacholczyk M: Conceptualization, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Supervision, Validation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by the Polish National Science Centre [DEC-2017/01/X/ST6/00636] grant (awarded to MP), co-financed by the European Union through the European Social Fund [POWR.03.02.00-00-00-1029] and by grant No. 02/040/BKM21/1019 from Silesian University of Technology, Gliwice, Poland (awarded to ML).

Copyright: © 2021 Ługowska M and Pacholczyk M. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Ługowska M and Pacholczyk M. PDBrt: A free database of complexes with measured drug-target residence time [version 1; peer review: awaiting peer review] F1000Research 2021, 10(Chem Inf Sci):1236
https://doi.org/10.12688/f1000research.73420.1

First published: 03 Dec 2021, 10(Chem Inf Sci):1236 https://doi.org/10.12688/f1000research.73420.1
Introduction
Current knowledge allowing for the determination of the effectiveness of small molecules in vivo is limited. It is known that the leading reason for drug candidate failure is the lack of efficacy caused by a poor translation of in vitro potency assays into in vivo activity in humans (Copeland, 2016a, 2016b; Swinney, 2009). In vitro experimentation refers to closed system conditions in which the drug molecule and its target are present at unchanging concentrations throughout the experiment (Tummino & Copeland, 2008). However, in living organisms, processes run under open, non-equilibrium conditions where the drug is constantly interacting with various molecules during many physiological processes in addition to its native target. To improve prediction of in vivo drug efficacy measurements of drug-target complexes, the residence time, defined as the reciprocal of the dissociation rate constant ($k_{off}$), should be considered. Drug-target residence time is crucial because pharmacological activity depends on the drug being bound to its molecular target. When the drug dissociates from the binding site, the target molecule is free to continue its pathophysiological function. Within 10 years of the development of the drug-target residence time concept, the parameter has become an extremely important factor in the process of optimizing lead structures in computer-aided drug design (Copeland et al., 2006; Copeland, 2016a, 2016b). Traditional computational approaches (such as molecular docking) take into account only the equilibrium affinity of the drug for its molecular target (e.g., $K_i$); however, the concept of residence time also takes into account conformational dynamics of molecules, which can have a significant effect on the binding and dissociation of the drug (Copeland et al., 2006; Tummino & Copeland, 2008). Existing free and commercial software does not include a residence time model that is able to estimate this quantity, and thus the effects of ligand modification relevant to the design and optimization of drug candidates is not available.

To enable further studies and the development of useful drug-target residence time models, we created a database of experimentally measured residence times for biomolecular complexes deposited in the Protein Data Bank (PDB). These data represent a link between structural and kinetic information of the complexes which may be helpful for various computational and machine learning studies on drug-like molecules in biological systems. The current version of our database, called PDBrt, contains 59 complexes with experimentally measured residence time including seven protein families, and 56 small molecules. Summary statistics are available in Figure 1 and Table 1.

![Figure 1. Distribution of protein families stored in PDBrt.](image)
Methods

Implementation

PDBrt database design and structure

The PDBrt database has been designed as an interactive web interface where the user can browse and extract information about the ligand residence time in its molecular target. The RESTful application programming interface (API) with the Django Rest Framework (v2.2.20) and a backend PostgreSQL (v12) database running on a Nginx (v1.21.3) web server has been developed to represent the output of a query as a user-friendly web page generated in Bootstrap (v3.3.7) with Hypertext Markup Language (HTML), Cascading Style Sheets (CSS) and jQuery (v3.5.1) to report the results. Search, query, and data extraction and visualization systems were developed for searching ligand residence time and binding kinetics coefficients. The PDBrt database facilitates access to information about the ligand residence time in its molecular target.

PDBrt database management system

The Database Management System (DBMS) allows users and programmers to manipulate the data in a systematic way. The DBMS serves as an interface between the database and end users, ensuring consistent data organization and easy accessibility. The PDBrt database is a type of relational DBMS using Structured Query Language (SQL) as the standard programming language for data manipulation.

Development of the PDBrt database was a multi-step process consisting of a systematic literature search, abstract and report screening, and article review. Several sequential steps involved in data management ensure that the processed data is accessible, reliable and current for its users (Figure 2).

Binding kinetics data acquisition and extraction

Data acquisition starts with collection of protein-ligand binding kinetics data from available literature by reviewing the primary reference of each pdb file in the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB) to retrieve the experimentally measured ligand-target residence time or dissociation rate (k_off). The references were

Table 1. A list of chemical compounds in pairs with a given protein family.

<table>
<thead>
<tr>
<th>Protein name</th>
<th>Ligand code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dipeptidyl peptidase 4 (DPP4)</td>
<td>BJM, LF7</td>
</tr>
<tr>
<td>Epidermal growth factor receptor (EGFR)</td>
<td>FMM</td>
</tr>
<tr>
<td>Enoyl-acyl carrier protein reductase (ENR)</td>
<td>TCL</td>
</tr>
<tr>
<td>HIV-1 protease (HIV-1 PR)</td>
<td>AB1, 1UN, 017</td>
</tr>
<tr>
<td>Heat shock protein 90 (HSP90)</td>
<td>C4T, 72K, 72Y, 73J, 73Y, 73Z, 9EK, 2EQ, 9R8, H0T, 74E, C4N, C3Z, C4K, BSQ, PU1, P4A, BAW, B8, 9RZ, 70M, 70O, 70K, GDM, BSM, 2GG, 2GJ, YKI, YKJ, 70Z, C8W, 6GC, 6GW, 6G7, 6DL, 6FF, 6FJ</td>
</tr>
<tr>
<td>Enoyl-acyl carrier protein reductase (InhA)</td>
<td>TCU, 53K, XT3, XT, XTW, XT5, XT0, JUS, 1TN, 1US, 2TK</td>
</tr>
<tr>
<td>Muscarinic acetylcholine receptor M3, Lysozyme (M3R)</td>
<td>0HK</td>
</tr>
</tbody>
</table>

Figure 2. PDBrt database management steps.
downloaded and carefully studied to manually extract the data. Only complexes with known ligand-target residence time or dissociation rate were added into the final dataset. Additionally, major binding kinetics coefficients were collected (if available): inhibition constant ($K_i$) and association rate ($k_{on}$). Currently the PDBrt database includes 59 protein-ligand complexes with known ligand-target residence time. Structures will be added on a regular basis as respective data becomes available.

**Structural data acquisition**

Complexes with particular PDB identifiers are downloaded from the RCSB PDB database into the internal PDBrt database core. The protein molecule along with other components such as water molecules and metal ions were saved in the *pdb* format, while the ligand (drug) in Structure Data Format (*sdf*). Neither the protein nor the ligand was subjected to any structural optimization or modification after being downloaded from the RCSB PDB.

**PDBrt data content**

PDBrt database data contain, in addition to the coordinates and general information required for all deposited structures in the RCSB PDB database, target residence time and other binding kinetics coefficients, structure files mentioned above, basic ligand properties like simplified molecular-input line-entry system (SMILES) or International Chemical Identifier (InChI) string, as well as links to external databases: RCSB PDB, PDBj, PDBe, PDBsum, and the reference literature in PubMed. The database includes citations to the original sources (publications) that contain information about the experimentally measured residence time or dissociation rate. For each complex, a web-based three-dimensional rendering is provided using the free, openly available object-oriented JavaScript library 3Dmol.js, which is used for visualizing molecular data.

**PDBrt database availability and updates**

PDBrt is available to the community through its web-based interface. Since the PDB database is growing rapidly and drug–target residence time is a parameter of great interest for drug design and optimization, the PDBrt will be updated on a regular basis with major versions issued annually. New data can be added by the authors (database administrators) either by uploading an *xlsx* (MS Excel) file or manually.

**PDBrt database architecture**

PDBrt is a three-tier architecture:

1. Data tier comprises database and data access layer;
2. Application tier controls application functionality by performing detailed processing;
3. Presentation tier is accessible for end-users and displays information on the website.

Core relational database managed by PostgreSQL server provides information storage for the deposited data. Back end (data access layer) was implemented in *Python* (v3.6) and the front end of the database (presentation layer) in HTML/CSS.

**PDBrt database model**

In PDBrt data is presented as a collection of relations - tables. Each column (also called attribute or field) in the table has a distinct name and a specific data type assigned to it. All the information related to a particular type is stored in a row (also called record) of that table. PDBrt database has three main columns (‘Complex’, ‘Protein’, ‘Ligand’) and 59 records with one-to-one type of relationship. This means that one protein or ligand could only belong to one complex and one complex consists of only one specific protein and ligand molecule (Figure 3).

**Operation**

PDBrt is available to the community through its web-based interface and is freely available to non-commercial users. The PDBrt database runs on all modern web browsers. See *Software availability* (Lugowska & Pacholczyk, 2021) for access to the database and code.
Use case

To show the usage of the PDBrt database, the Unified Modelling Language (UML) use case diagram was adopted. Figure 4 shows five use cases of PDBrt as well as three actors: system administrator, end user and the database. The database is an actor for all use cases, system administrator for two and end user for three of these. The database actor is involved in all five use cases because it stores the data and enables operations by which advanced data handling functions are created. In the ‘complex management’ use case the system administrator is able to perform basic Create, Read, Update and Delete (CRUD) actions on the ‘Complex’ table: add, delete, update as well as upload new data as Microsoft Excel (.xlsx) file format. In the ‘user management’ use case the system administrator can add, edit and delete a regular user. In the ‘view’, ‘search’ and ‘download’ use case a user actor is involved. In the ‘view’ use case a list of complexes is displayed to the user who can choose a single complex and view its details as well as its 3D visualization. In the ‘search’ use case the user is able to filter from the whole database by PDB code, residence time, and protein or ligand name. The ‘download’ use case allows users to either download the data stored in the database as plain text file or pdb/sdf file formats.
Query interface has been implemented for the query of data within PDBrt. Figure 5 shows how the query options are organised. The search engine provides one form field for keyword search and allows retrieval by PDB code, protein name and residence time.

Two user interfaces provide extensive information for result sets obtained for particular search query. The ‘homepage’ interface allows access to some general information in tabular format, and offers the possibility to download whole sets of data files for result sets consisting of multiple PDBrt entries. The ‘complex detail’ interface provides information about individual structures as well as cross-links to many external resources for macromolecular structure data.

**Discussion**

Drug-target residence time has been shown to play an important role in the prediction of in vivo efficacy of the drug. Since the parameter is independent of drug and enzyme concentration, it is important in the drug design process and should be taken into account at an early stage. The availability of information about drug-target complexes with known (measured) residence times is becoming more and more significant. Several databases containing information about protein–ligand complexes are available, for example PDB-bind (Liu et al., 2015), BindingDB (Gilson et al., 2016), and Binding MOAD (Smith et al., 2019). However, these databases (except for BindingDB) do not store detailed information about binding rates, and none of these contain direct information about ligand (drug) residence time in its target macromolecule. The PDBrt database is dedicated to reporting the structure of protein - small molecule complexes, along with their target residence time and additional binding parameters ($K_d$, $k_{on}$, $k_{off}$). Currently a total of 59 protein–ligand complexes are
Figure 5. Current query capabilities of the PDBrt diagram. Diagram presents the structure of the PDBrt website. Each symbol (shape or arrow) presents PDBrt web page (single web view), page content, group of related content on a single page, relationship between web pages and group of similar web pages as described in the legend.

deposited in the web–based PDBrt database and this data will be updated frequently as more data is made available. The information associated with the existing 59 protein–ligand complexes is available in Underlying data (Lugowska, 2021).

Data availability
Underlying data
This project contains the following underlying data:

- **Residence_time_data.xlsx**: (information about target residence time and other binding kinetics coefficients, basic ligand properties like simplified molecular-input line-entry system (SMILES) or International Chemical Identifier (InChI) string, as well as the reference literature in PubMed).

- **Structures.zip**: (the protein molecule along with other components such as water molecules and metal ions in the *pdb* format, the protein-ligand structure in the *pdb* format, and the ligand (drug) in both *pdb* and Structure Data Format (*sdf*)).

Data are available under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

**Software availability**

Database available from: https://pdbrt.polsl.pl/

Source code available from: https://github.com/mlugowska/residence_time

Archived source code at time of publication: https://doi.org/10.5281/zenodo.5583543 (Ługowska & Pacholczyk, 2021)

License: MIT

**References**


The benefits of publishing with F1000Research:

• Your article is published within days, with no editorial bias
• You can publish traditional articles, null/negative results, case reports, data notes and more
• The peer review process is transparent and collaborative
• Your article is indexed in PubMed after passing peer review
• Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com