

Magdalena ŁUGOWSKA, Marcin PACHOLCZYK  
Politechnika Śląska

## **PDBrt: BAZA DANYCH KOMPLEKSÓW BIOMOLEKULARNYCH Z WYZNACZONYM CZASEM REZYDENCJI LEK-CEL MOLEKULARNY**

**Streszczenie.** Trudności w przełożeniu aktywności *in vitro* określonej w testach komórkowych na skuteczność *in vivo* w organizmach żywych, komplikują projektowanie i opracowywanie nowych leków. Czas rezydencji leku w jego celu molekularnym staje się kluczowym parametrem w projektowaniu i optymalizacji nowych leków, ponieważ ostatnie badania pokazują, że czas rezydencji może stanowić wiarygodną miarę skuteczności leku *in vivo*. Obecnie dostępne są eksperymentalne podejścia do wyznaczania parametrów kinetyki wiązania kompleksów białko-ligand, ale znane bioinformatyczne bazy danych zwykle nie podają informacji o czasie rezydencji (czas przebywania ligandu w celu molekularnym).

W celu rozszerzenia istniejących baz, opracowano bazę danych PDBrt, która przechowuje m.in. informacje o czasie rezydencji ligandów.

Baza danych PDBrt jest darmowym, niekomercyjnym repozytorium danych 3D dotyczących kompleksów białko-ligand, w tym zmierzonego czasu rezydencji ligandu w kieszeni wiążącej specyficznych makrocząsteczek biologicznych zdeponowanych w Protein Data Bank. Baza danych jest zaimplementowana przy użyciu backendu Django Rest Framework i frontendu przy użyciu HTML i JavaScript. Baza danych PDBrt zawiera informacje o białku i ligandzie oddzielnie, a także kompleksie białko-ligand i czasie rezydencji ligandu w miejscu wiązania białka.

## **PDBrt: A DATABASE OF BIOMOLECULAR COMPLEXES WITH MEASURED DRUG -TARGET RESIDENCE TIME**

**Summary.** 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.

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.

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.

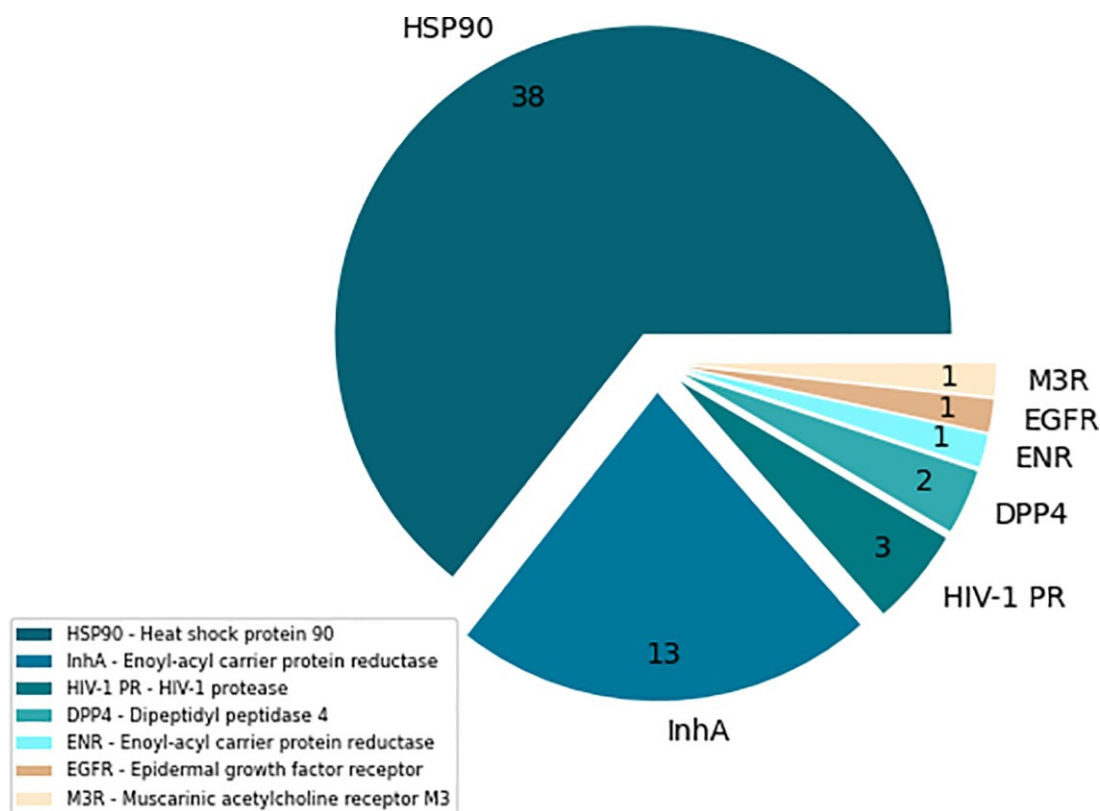
Availability: <https://pdbrt.polsl.pl>, article based on [1].

## 1. Wstęp

Istniejący stan wiedzy pozwalający na określenie skuteczności małych cząsteczek *in vivo* jest ograniczony. Wiadomo, że główną przyczyną niepowodzenia we wprowadzaniu kandydata na lek do praktyki klinicznej, jest brak skuteczności spowodowany słabą translacją testów efektywności działania *in vitro* na aktywność *in vivo* u pacjentów. Eksperymenty *in vitro* odnoszą się do warunków systemu zamkniętego, w których cząsteczka leku i jej cel molekularny (np. białko) są obecne w niezmiennych stężeniach podczas całego eksperymentu [2]. Jednak w organizmach żywych, procesy przebiegają w otwartych, nierównowagowych warunkach, w których lek stale wchodzi w interakcję, oprócz właściwego sobie celu, z różnymi cząsteczkami podczas wielu procesów fizjologicznych związanych z eliminacją obcej substancji przez żywy organizm. W celu poprawy predykcji skuteczności leku *in vivo* kompleksów lek - cel molekularny, należy wziąć pod uwagę czas rezydencji (czas przebywania leku w celu molekularnym, czas trwania kompleksu lek - cel), zdefiniowany jako odwrotność stałej dysocjacji ( $1/k_{\text{off}}$ ). Czas rezydencji leku w miejscu docelowym ma kluczowe znaczenie, ponieważ aktywność farmakologiczna wynika ze związania leku z jego celem molekularnym. Kiedy lek ulegnie dysocjacji od miejsca wiązania, cel molekularny może swobodnie kontynuować swoją patofizjologiczną funkcję. W ciągu 10 lat od opracowania koncepcji czasu rezydencji leku w celu molekularnym [3, 4], parametr stał się niezwykle ważnym czynnikiem w procesie optymalizacji struktur wiodących wykorzystywanym we wspomaganym komputerowo projektowaniu leków. Tradycyjne podejścia obliczeniowe (takie jak dokowanie molekularne) uwzględniają tylko powinowactwo równowagowe leku do jego celu molekularnego (np.  $K_i$ ). Koncepcja czasu rezydencji uwzględnia również dynamikę konformacyjną cząsteczek, która może mieć istotny wpływ na wiązanie i dysocjację leku. Istniejące bezpłatne i komercyjne oprogramowanie nie zawiera modelu czasu rezydencji zdolnego praktycznego do oszacowania tej wielkości, a zatem efekty modyfikacji ligandów istotne dla projektowania i optymalizacji kandydatów na lek nie są łatwo dostępne.

Aby umożliwić dalsze badania i opracowanie użytecznych modeli czasu rezydencji leku w celu molekularnym, utworzono bazę danych zawierającą eksperymentalnie zmierzone czasami rezydencji kompleksów biomolekularnych zdeponowanych w Protein Data Bank (PDB). Dane te, stanowią powiązanie między informacją strukturalną i kinetyczną kompleksów, która może być pomocna

w badaniach obliczeniowych i przy tworzeniu modeli uczenia maszynowego dla organicznych cząsteczek lekopodobnych. Obecna wersja bazy PDBrt, zawiera 59 kompleksów z eksperymentalnie zmierzonym czasem rezydencji. Białka pochodzą z siedmiu rodzin. Podsumowanie zawartości pokazuje rysunek 1.



Rys. 1. Rodziny białkowe w bazie PDBrt. Źródło: [1]

## 2. Implementacja bazy danych PDBrt

### 2.1. Projekt i struktura bazy danych PDBrt

Baza danych PDBrt została zaprojektowana jako interaktywny interfejs sieciowy, w którym użytkownik może przeglądać i pobierać informacje o czasie rezydencji ligandu w jego celu molekularnym. W celu przedstawienia wyników zapytania, w formie przyjaznej dla użytkownika strony internetowej, generowanej w HTML za pomocą JavaScript, do pozyskiwania wyników wykorzystano interfejs API RESTful (Django Rest Framework) wraz z silnikiem bazodanowym PostgreSQL, pracujące na serwerze sieciowym Nginx. Opracowano systemy wyszukiwania, zapytań i ekstrakcji danych oraz wizualizacji dla wyszukiwania czasu rezydencji ligandów i współczynników kinetyki wiązania. Baza danych jest dostępna bezpłatnie dla użytkowników niekomercyjnych i ułatwia dostęp do informacji o czasie rezydencji ligandu.

## 2.2. Pozyskiwanie i walidacja danych dotyczących kinetyki wiązania

Zbieranie danych rozpoczyna się od zebrania danych kinetyki wiązania białko-ligand z dostępnej literatury. W tym celu można zwykle wykorzystać artykuł powiązany z określonym kompleksem zdeponowanym w bazie PDB. Poszukuje się zmierzonego eksperymentalnie czasu rezydencji lub stałej dysocjacji ( $k_{\text{off}}$ ). Pozycje literaturowe zostały pobrane i dokładnie przestudiowane, aby ręcznie wyodrębnić potrzebne dane. Do końcowego zestawu danych dodano tylko kompleksy o znanym czasie rezydencji ligand-cel molekularny lub stałej dysocjacji. Dodatkowo zebrano główne współczynniki kinetyki wiązania (jeśli były dostępne): stała inhibicji ( $K_i$ ) i stała asocjacji ( $k_{\text{on}}$ ). Obecnie baza danych PDBrt obejmuje 59 kompleksów białko-ligand o znanym czasie rezydencji ligand-cel molekularny. Struktury będą dodawane na bieżąco w miarę pojawiania się odpowiednich danych.

## 2.3. Pozyskiwanie danych strukturalnych

Kompleksy z określonymi identyfikatorami PDB są pobierane z bazy danych RCSB PDB do wewnętrznego rdzenia bazy danych PDBrt. Częsteczka białka wraz z innymi składnikami, takimi jak cząsteczki wody i jony metali, jest zapisywana w formacie PDB, natomiast ligand (lek) w formacie Structure Data Format (SDF). Ani białko, ani ligand nie są poddawane żadnej strukturalnej optymalizacji lub modyfikacji po pobraniu z RCSB PDB.

## 2.4. Typy zawartości danych w PDBrt

Dane zgromadzone w PDBrt zawierają, oprócz współrzędnych i ogólnych informacji wymaganych dla wszystkich struktur zdeponowanych w bazie danych RCSB PDB, czas rezydencji i inne współczynniki kinetyki wiązania, pliki struktur wspomniane powyżej oraz podstawowe informacje dotyczące ligandów, takie jak SMILES lub ciąg InChI (sposób reprezentacji struktury w postaci ciągu znaków). Dla każdego kompleksu zapewniona jest trójwymiarowa wizualizacja struktury.

## 2.5. Dostępność i aktualizacje bazy danych PDBrt

PDBrt jest dostępna dla społeczności akademickiej za pośrednictwem strony internetowej. Ponieważ baza danych PDB szybko się rozrasta, a czas rezydencji jest parametrem o dużym znaczeniu dla projektowania i optymalizacji leków, PDBrt będzie regularnie aktualizowana, a główne wersje będą wydawane co roku. Nowe dane może dodawać administrator bazy danych, ręcznie lub przesyłając plik XLSX (MS Excel).

## 2.6. Architektura i model bazy danych PDBrt

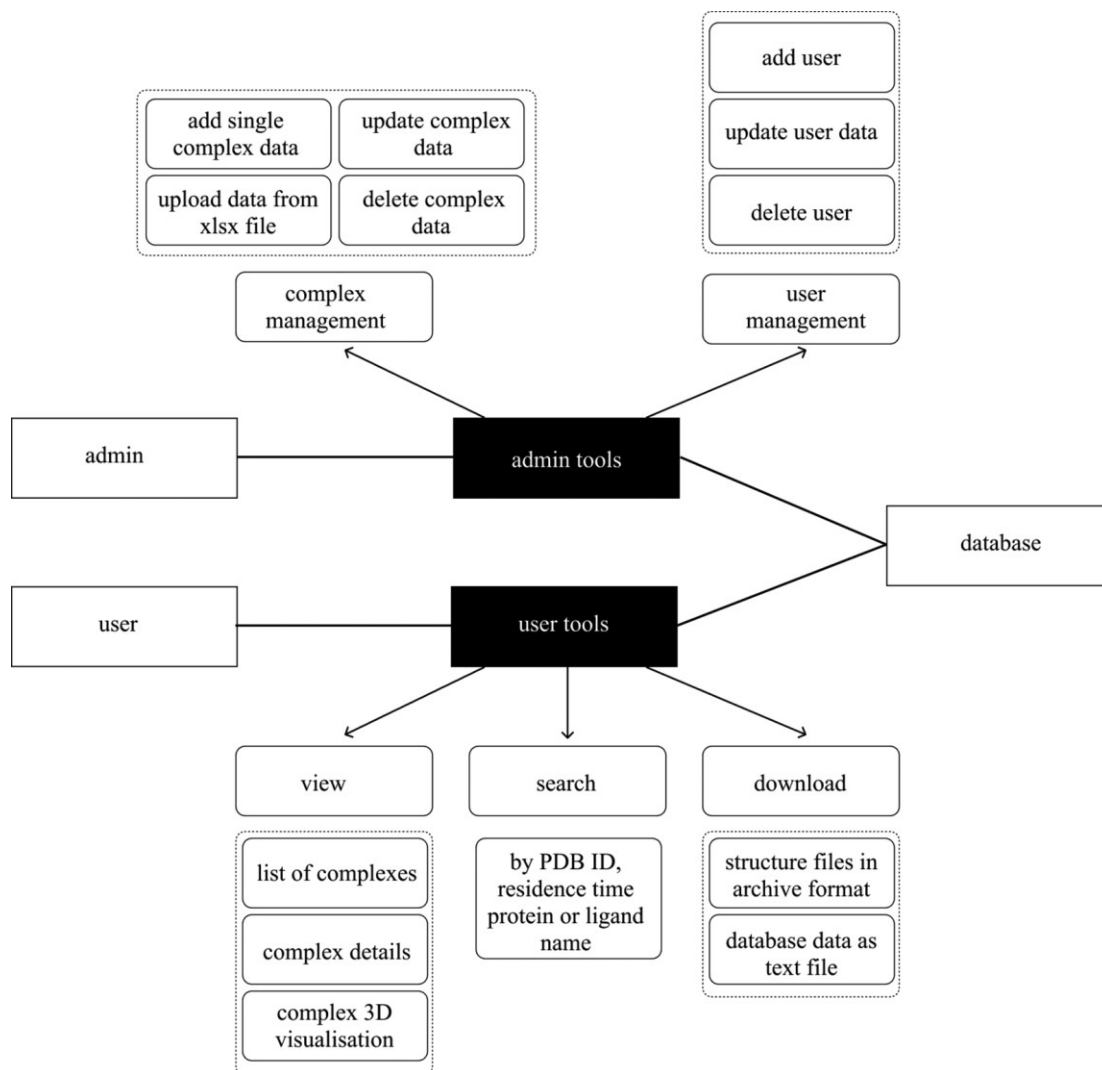
W PDBrt dane prezentowane są jako zbiór relacji – tabele. Każda kolumna (nazywana również atrybutem lub polem) w tabeli ma odrębną nazwę i przypisany do niej określony typ danych. Wszystkie informacje związane z określonym typem są przechowywane w wierszu (zwanym również rekordem) tej tabeli. Baza danych PDBrt

składa się z trzech głównych kolumn („Złożona”, „Białko”, „Ligand”) i 59 rekordów z relacjami typu jeden do jednego. Oznacza to, że jedno białko lub ligand może należeć tylko do jednego kompleksu, a jeden kompleks składa się tylko z jednej określonej cząsteczki białka i ligandu.

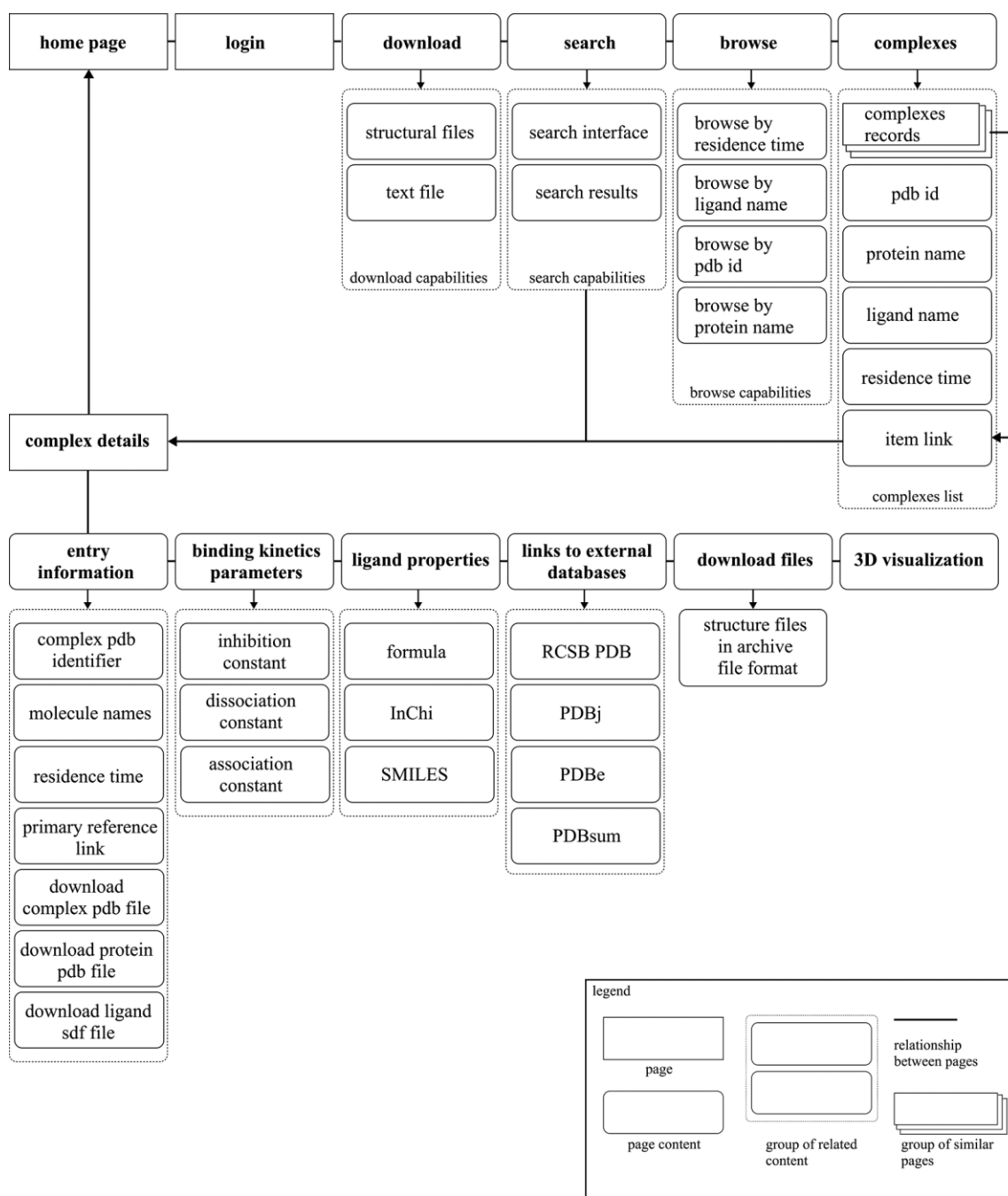
### 3. Przypadki użycia

W celu demonstracji wykorzystania bazy danych PDBrt, wykorzystano diagram przypadków użycia UML (ang. *Unified Modeling Language*). Rysunek 2 przedstawia pięć przypadków użycia bazy PDBrt oraz trzech aktorów: administratora systemu (admin), użytkownika końcowego (user) i bazę danych (database). Baza danych jest aktorem dla wszystkich przypadków użycia, administrator systemu dla dwóch z nich, użytkownik końcowy występuje w trzech przypadkach. Aktor baza danych bierze udział we wszystkich pięciu przypadkach użycia, ponieważ przechowuje dane i umożliwia operacje za pomocą których tworzone są zaawansowane funkcje obsługi danych. W przypadku użycia „zarządzanie złożone” (complex management) administrator systemu może wykonać podstawowe czynności tworzenia, odczytu, aktualizacji i usuwania (ang. *CRUD*) na tabeli „złożone”: dodawać, usuwać, aktualizować, a także przysyłać nowe dane jako plik formatu Microsoft Excel (XLSX). W przypadku użycia „zarządzanie użytkownikami” (user management) administrator systemu może dodawać, edytować i usuwać zwykłego użytkownika. W przypadku użycia „widok” (view), „wyszukiwanie” (search) i „pobieranie” (download) zaangażowany jest aktor użytkownik. W przypadku użycia „widok” użytkownikowi wyświetlana jest lista kompleksów, z której może wybrać pojedynczy kompleks i wyświetlić dotyczące go szczegóły oraz wizualizację 3D. W przypadku użycia „wyszukiwanie” użytkownik może filtrować dane z całej bazy danych według kodu PDB, czasu rezydencji i nazwy białka lub ligandu. Przypadek użycia „pobierz” umożliwia użytkownikom pobranie danych przechowywanych w bazie danych w postaci zwykłego pliku tekstowego lub plików w formacie PDB/SDF.

W celu umożliwienia interakcji z użytkownikiem w bazie PDBrt został zaimplementowany interfejs zapytań. Rysunek 3 pokazuje sposób organizacji opcji zapytań. Wyszukiwarka udostępnia pojedyncze pole formularza umożliwiające wyszukiwanie słów kluczowych i zapewnia możliwość wyszukiwania według kodu PDB struktury kompleksu, nazwy białka i wielkości czasu rezydencji. Dwa interfejsy użytkownika dostarczają obszernych informacji o zestawach wyników uzyskanych dla konkretnego zapytania wyszukiwania. Interfejs „strony głównej” umożliwia dostęp do niektórych ogólnych informacji w formacie tabelarycznym i oferuje możliwość pobierania całych zestawów plików danych dla zestawów wyników składających się z wielu wpisów PDBrt. Interfejs „złożonych szczegółów” dostarcza informacji o poszczególnych strukturach, a także łączy internetowych do wielu zewnętrznych zasobów danych dla struktury makromolekularnej.



Rys. 2. Diagram przypadków użycia bazy PDBrt. Źródło: [1]



Rys. 3. Diagram zapytań dla PDBrt. Źródło: [1]

#### 4. Dyskusja

Wykazano, że czas rezydencji w miejscu docelowym leku odgrywa ważną rolę w przewidywaniu skuteczności leku *in vivo*. Ponieważ parametr jest niezależny od stężenia leku i enzymu, jest ważny w procesie projektowania leku i powinien być uwzględniony na jego wczesnym etapie. Dostępność informacji o kompleksach lek-cel o znanych (zmierzonych) czasach rezydencji nabiera coraz większego znaczenia. Dostępnych jest kilka baz danych zawierających informacje o kompleksach białko-ligand, np. PDB-bind [5], BindingDB [6] oraz BindingMOAD [7]. Jednak te

bazy danych (z wyjątkiem BindingDB) nie przechowują szczegółowych informacji o stałych kinetyki wiązania, a żadna z nich nie zawiera bezpośredniej informacji o czasie rezydencji ligandu (leku) w docelowej makrocząsteczce. Baza danych PDBrt jest dedykowana do raportowania budowy kompleksów białkowo-drobnocząsteczkowych wraz z ich docelowym czasem rezydencji oraz dodatkowymi parametrami wiązania ( $K_i$ ,  $k_{on}$ ,  $k_{off}$ ). Obecnie w internetowej bazie danych PDBrt zdeponowanych jest łącznie 59 kompleksów białko-ligand. Zawartość bazy będzie często aktualizowana w miarę udostępniania większej ilości danych.

Dostępność: <https://pdbrt.polsl.pl>, artykuł na podstawie [1].

## LITERATURA

1. Ługowska M., Pacholczyk M.: PDBrt: A free database of complexes with measured drug-target residence time. *F1000Research*, 2021, 10, (Chem Inf Sci):1236, (<https://doi.org/10.12688/f1000research.73420.1>).
2. Tummino P.J., Copeland R.A.: Residence time of receptor–ligand complexes and its effect on biological function. *Biochemistry* 2008, 47, p. 5481–5492.
3. Copeland R.A., Pompliano D.L., Meek T.D.: Drug-target residence time and its implications for lead optimization. *Nat. Rev. Drug Discov.* 2006, 5(2), p.730–739.
4. Copeland R.A.: The drug-target residence time model: A 10-year retrospective. *Nat. Rev. Drug Discov.* 2016, 15(2), p.87–95.
5. Liu Z., Li Y., Han L., Li J., Liu J., Zhao Z., Nie W., Liu Y., Wang R.: PDB-wide collection of binding data: current status of the PDBbind database. *Bioinformatics* 2015, 31(3), p.405-412.
6. Gilson M.K., Liu T., Baitaluk M., Nicola G., Hwang L., Chong J.: BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Research* 2016, 44, p. D1045-D1063.
7. Smith R.D., Clark J.J., Ahmed A., Orban Z.J., Dunbar Jr J.B., Carlson H.A.: Updates to Binding MOAD (Mother of All Databases): Polypharmacology Tools and Their Utility in Drug Repurposing. *J. Mol. Biol.* 2019, 431, p. 2423-2433.