Accessing expert-curated pharmacological data in the IUPHAR/BPS Guide to PHARMACOLOGY

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Current Protocols</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>CP-17-0310.R1</td>
</tr>
<tr>
<td>Wiley - Manuscript type:</td>
<td>Protocol</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>n/a</td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>Sharman, Joanna; University of Edinburgh, Deanery of Biomedical Sciences Harding, Simon; University of Edinburgh, Deanery of Biomedical Sciences Southan, Christopher; University of Edinburgh, Deanery of Biomedical Sciences Faccenda, Elena; University of Edinburgh, Deanery of Biomedical Sciences Pawson, Adam; University of Edinburgh, Deanery of Biomedical Sciences Davies, Jamie; University of Edinburgh, Deanery of Biomedical Sciences</td>
</tr>
<tr>
<td>Keywords:</td>
<td>Pharmacology, immunology, medicinal chemistry, curated database, drug target, ligand, receptor</td>
</tr>
<tr>
<td>Abstract:</td>
<td>The IUPHAR/BPS Guide to PHARMACOLOGY is an expert-curated, open-access database of information on drug targets and the substances that act on them. This paper describes the procedures for searching and downloading ligand-target binding data and for finding detailed annotations and the most relevant literature. The database includes concise overviews of the properties of 1,700 data-supported human drug targets and related proteins, divided into families, and 9,000 small molecule and peptide experimental ligands and approved drugs that bind to those targets. More detailed descriptions of pharmacology, function, and pathophysiology are provided for a subset of important targets. The information is reviewed regularly by expert subcommittees of the IUPHAR Committee on Receptor Nomenclature and Drug Classification. A new immunopharmacology portal has recently been added, drawing together data on immunological targets, ligands, cell types, processes and diseases. The data are available for download and can be accessed computationally via web services.</td>
</tr>
</tbody>
</table>
Unit Title

Accessing expert-curated pharmacological data in the IUPHAR/BPS Guide to PHARMACOLOGY

Running Title

Guide to PHARMACOLOGY and Guide to IMMUNOPHARMACOLOGY

Author(s)

Joanna L Sharman, Simon D Harding, Christopher Southan, Elena Faccenda, Adam J Pawson, Jamie A Davies and NC-IUPHAR.

Contact information

Deanery of Biomedical Sciences, University of Edinburgh, Hugh Robson Building, George Square, Edinburgh, EH8 9XD, UK

Significance Statement

Online databases contain millions of chemical structures, a fraction of which have been tested for biological activity in the literature. Activity data are often hidden in dense text or supporting material making them difficult to find. The IUPHAR/BPS Guide to PHARMACOLOGY uses an expert-driven process to select the most useful data on drug targets and to recommend ligands for experimental use, presenting the information in a standard database format for easy searching.

ABSTRACT

The IUPHAR/BPS Guide to PHARMACOLOGY is an expert-curated, open-access database of information on drug targets and the substances that act on them. This paper describes the procedures for searching and downloading ligand-target binding data and for finding detailed annotations and the most relevant literature. The database includes concise overviews of the properties of 1,700 data-supported human drug targets and related proteins, divided into families, and 9,000 small molecule and peptide experimental ligands and approved drugs that bind to those targets. More detailed descriptions of pharmacology, function, and pathophysiology are provided for a subset of important targets. The information is reviewed regularly by expert subcommittees of the IUPHAR Committee on Receptor Nomenclature and Drug Classification. A new immunopharmacology portal has recently been added, drawing together data on immunological targets, ligands, cell types, processes and diseases. The data are available for download and can be accessed computationally via web services.

Keywords:

Pharmacology, immunology, medicinal chemistry, curated database, drug target, ligand, receptor
INTRODUCTION

The International Union of Basic and Clinical Pharmacology (IUPHAR)/British Pharmacological Society (BPS) Guide to PHARMACOLOGY database (GtoPdb, http://www.guidetopharmacology.org/) is unique in receiving expert contributions and review by >500 scientists in 96 subcommittees of the IUPHAR Committee on Receptor Nomenclature and Drug Classification (NC-IUPHAR) (Harding et al., 2018). The resource aims to summarise key information on drug targets and the substances that act on them from literature and present this in a searchable, user-friendly format. The scope of coverage includes any human protein whose activity can be modulated by a small molecule or peptide ligand. Ligands are either approved drugs, natural ligands, antibodies, or other recommended compounds available for experimental use and, where possible, quantitative binding data are provided along with references. Presently, GtoPdb includes ~1,700 data-supported human drug targets and some additional related proteins of emerging interest to drug discovery. These are organised hierarchically by family: Basic Protocol 1 shows how to navigate the family tree to locate a target of interest and find ligands that modulate it. Alternate Protocol 1 shows how to perform a text-based search to find the same target; this is a quick way to find a specific target, or can be useful when the target classification is unknown. Support Protocol 1 introduces the powerful functionality that can be gained by using the GtoPdb-UniProtKB reciprocal links to combine and filter across hundreds of data types and cross-references in UniProtKB (The UniProt Consortium, 2017).

The ~9,000 ligands are organised by chemical class, e.g. synthetic organic compounds and natural products. Basic Protocol 2 shows how to navigate the ligand list to find a drug and locate information about its mechanism of action and known targets. For small molecule ligands and drugs, searching by chemical structure is also possible. Alternate Protocol 2 shows how to use the GtoPdb chemical drawing tool and search function to find ligands by structural similarity. This has a number of applications, for instance, to identify possible targets for a novel structure, or to explore sets of similar compounds. Support Protocol 2 demonstrates how to use the GtoPdb cross-references to 92 million compounds in PubChem (Kim et al., 2016), to find further sets of related structures, to which filters can then be applied; for instance, to find bioactive compounds from the medicinal chemistry literature. The protocol also shows how the results can be pivoted back to the GtoPdb “small-data” resource.

Much of the data are available to download in a choice of formats. Basic Protocol 3 describes how to download CSV files with lists of ligands, targets and interactions. Alternate Protocol 3 gives an introduction to accessing data via the web services, which can be used in computer programs to give access to the data in a machine-readable format.

A recent expansion of the database has focussed on new types of data pertaining to immunopharmacology, the immune system and inflammation. GtoPdb targets have been linked to immunological cell types, diseases and processes, accessible via a new portal, the IUPHAR Guide to IMMUNOPHARMACOLOGY (GtoImmuPdb, www.guidetoimmunopharmacology.org). Immunity, inflammation and infection are priority areas for drug discovery (Carvalho et al., 2016; Ishii, 2017; Landolina et al., 2016; Tiligada et al., 2015). Many chronic diseases, including aspects of ageing, have immune and/or inflammatory components (Licastro et al., 2005). Auto-immunity is a serious clinical problem (Mcnnes et al., 2011; O’Shea et al., 2013), with its progression dependent on immune and inflammatory responses (Perry et al., 2003). A main driver of the GtoImmuPdb project is to support the discovery of more efficacious and selective drugs to modulate these types of responses. GtoImmuPdb aims to do this by facilitating a
stronger connection between the immunological and pharmacological research communities, with easy access to data in both directions. In particular, pharmacological information has hitherto been difficult for immunologists without chemoinformatics expertise to access. The development of GtoImmuPdb, with an immunologist-friendly portal, aims to address this gap in information exchange.

The first protocol focussed on GtoImmuPdb, Basic Protocol 4, shows how to navigate the immunological disease data to identify the approved drugs and primary targets for a given disease. It introduces the GtoImmuPdb portal and explains how to view and understand the disease to target association data. Alternate Protocol 4 acts as a companion to Basic Protocol 4 by illustrating how to browse the disease to ligand association data. With both protocols, the aim is to identify approved drugs for a given immunological disease, and illustrate how navigating between disease association data and ligand summary pages provides access to more details about the approved drugs. Support Protocol 3 provides a more general illustration of searching for diseases in GtoPdb/GtoImmuPdb using the search function to find associated targets/ligands.

Basic Protocols 5 and 6 deal with exploring the GtoImmuPdb cell type (Basic Protocol 5) and process (Basic Protocol 6) data. They explain the way GtoImmuPdb maps external ontology terms to the data and how this underpins the data curation and website navigation. The two protocols demonstrate how to browse and view targets associated with cell types and processes, how to navigate from these pages to find further information, and how to interpret the data.

Finally, Basic Protocol 7 illustrates the power of GtoPdb curation in connecting literature on biology and bioactive chemistry via the triad of GtoPdb-PubMed-PubChem reciprocal linking.

**BASIC PROTOCOL 1**

**FIND A DRUG TO MODULATE A KNOWN TARGET**

The GtoPdb home page is located at http://www.guidetopharmacology.org/. There are several ways to locate targets. Clicking on the Targets link in the main navigation menu on any page goes to the targets home page at http://www.guidetopharmacology.org/targets.jsp. Mousing-over the Targets menu reveals direct links to the top-level target classes: G protein-coupled receptors (GPCRs), ion channels, nuclear hormone receptors (NHRs), kinases, catalytic receptors, transporters, enzymes and other protein targets. These are also available as quick links from the left-hand side of the home page.

Necessary resources

**Software**

An up-to-date Web browser, such as Firefox or Google Chrome, with JavaScript enabled.

**Browsing the target hierarchy to locate a target of interest**


John Wiley & Sons
Targets are divided into phylogenetically-related classes, with an additional Other Protein Targets class to capture targets not classified as one of the other types. Some targets may be classified under more than one class, for example some Kinases are also found under Catalytic Receptors.

2. Under the section on Ion Channels click on the link “View the list of ion channel families”.

Targets are organised into families, whose members tend to be similar in structure and/or share an endogenous ligand or family of ligands. By default, the tree loads with two levels expanded. Intermediate level branches (indicated by a plus or minus symbol to their left and in black text), can be expanded/condensed by clicking on their names. Clicking on the family names of leaf nodes (blue links) takes users to a summary page about that family. The whole tree can be expanded by clicking on the “Expand all nodes” button at the top. Clicking on the purple “OVERVIEW” link next to top and intermediate level families (such as Ligand-gated ion channels) will take the user to a summary page with an overview of the family’s properties and a list of subfamilies.

3. In the ion channel family tree under Ligand-gated ion channels, click on “5-HT$_3$ receptors” to go to their family page (Figure 1).

The first section on the family page contains an overview with background information on the family which is a useful summary for newcomers to the field. At the top right is a link to a more detailed introduction in the style of a review article, introducing important structural, functional and pharmacological aspects.

**Finding ligands known to modulate a target**

4. Under the Channels and Subunits section, click on “Show summary” next to 5-HT$_3$AB to show basic information about this target and key recommended ligands. Ligands are listed according to their action at the target, in this case “Channel blockers”. The affinity and reference are given next to the ligand name, where available (Figure 1).

In some cases, such as 5-HT$_3$ receptors, the functional targets can consist of multiple subunits. Multimeric targets are listed under the sub-heading “Complexes”, with a separate sub-heading below for “Subunits”. All data unless otherwise stated refer to the human proteins. Other species are sometimes included where human data are unavailable or significant species differences exist. Genetic information and database links are provided for human, mouse and rat.

5. Click on a ligand name to access further information about the ligand (described in Basic Protocol 2).

6. Return to the 5-HT$_3$AB summary and click on “More detailed page” to go to the detailed target page with more information on the structure, function, clinical relevance, and pharmacology.
Detailed pages are provided for a subset of important targets. Information is provided by NC-IUPHAR
subcommittees and reviewed and updated as necessary every two years. The data cover a range of different
types, including genetics, protein structure, functional assays, tissue distribution, clinically-relevant mutations,
associated diseases, and mouse gene knockouts. Bespoke information is also provided for particular target
types, such as ion channel conductance, GPCR transduction mechanisms, and NHR target genes. Detailed
pages typically contain longer lists of ligands, which include experimentally-used compounds, approved drugs,
and endogenous ligands.

7. Scroll down (or click on the link to Agonists in the Contents list) to the tables of ligands, which are divided by
ligand action, such as Agonists, Antagonists, Channel Blockers etc. (Figure 2). Here, there are several
agonists listed, with a comment below providing further context. These compounds are not listed on the
summary view because they are not selective for the 5-HT₃ receptor, but they are useful for characterising
the 5-HT₃ receptor family. Ligands marked with a medicine bottle icon are clinically approved drugs. Another
icon (a number one in a bullseye) indicates where this target is thought to be the primary target of the drug (i.e.
through which the drug exerts its therapeutic action).

Ligands are ordered by affinity type (pKᵣ, pKᵢ, etc), and then by decreasing potency. Tables can be reordered
by clicking on the column headings.

8. Under “Allosteric Modulators” click on the arrow at the right of the row for trichloroethanol. The row expands
to reveal further data about the interaction.

ALTERNATE PROTOCOL 1

USE SEARCH TOOLS TO FIND A DRUG TARGET

Basic Protocol 1 is useful if users know how the target they are looking for is categorised in GtoPdb. However, if a user
does not know where to find it in the family hierarchy, or simply wishes to go directly to the detailed target page, they
may wish to try a keyword search or use one of a number of recognised identifiers to search. This protocol uses the
Advanced Target Search function, but most searches can also be done using the quick search box at the top of every
page. The Target Search page includes various types of search: text-based, identifier, wildcard matches, literature-based
(e.g. by PubMed ID), and BLAST sequence searches.

Necessary resources

Software

An up-to-date Web browser, such as Firefox or Google Chrome, with JavaScript enabled.


2. Mouse over the “Advanced search” menu of the main navigation bar and click on “Target search”.

John Wiley & Sons
3. The top section is a keyword-based text search. In the search box type in “5-HT3”. Whilst typing, the search function will automatically try to match the keyword to a list of target and family names and synonyms (Figure 3). In the drop-down list click on one of the target names to go directly to its detailed target page, or on the family name to go to the family summary page.

The text search can be further restricted to specific database fields, species or target classes, and the search can be completed by clicking on the “Search the database” button.

4. Return to the search page by hitting the back button on the browser.

5. In the third search box paste in the UniProtKB accession P46098. The source database is “UniProt” by default, and this should be changed if a different type of identifier is used. Click “Search to database”.

6. On the results page click on the name “5-HT3A” to go to its detailed target page.

SUPPORT PROTOCOL 1

NAVIGATING RECIPROCAL LINKS TO UNIPROT

The provision of UniProtKB (Pundir et al., 2015; The UniProt Consortium, 2017) links in GtoPdb for protein targets (and some ligands) offers the utility to users of not only browsing the full entries but also using the protein accession for many types of query. All our human entries are Swiss-Prot (i.e. fully annotated and manually expert reviewed) but some rodent and other species are automated TrEMBL records (pending completion of their complete proteomes). Note also, all our ligand-to-target links are manually curated. For each release of GtoPdb we refresh the set of UniProtKB cross-references that point back to us (representing the set of proteins that have a recorded ligand interaction). This reciprocity has powerful advantages, just a few of which can be exemplified here. In essence, users can navigate from ligand space to target space and vice versa. This protocol will also demonstrate how to query across proteins in UniProt and then come “in” to GtoPdb.

Necessary resources

Software

An up-to-date Web browser, such as Firefox or Google Chrome, with JavaScript enabled.

A text editor, such as Microsoft WordPad, for temporary paste-across storage.

A program capable of archiving lists, such as Microsoft Excel or LibreOffice Calc.

For set intersecting, the Venny tool is useful (http://bioinfogp.cnb.csic.es/tools/venny/)

John Wiley & Sons
Getting a UniProtKB list of all the human GtoPdb targets and sorting them

1. Go to the UniProtKB website at (http://www.uniprot.org/) and open the “Advanced Query” menu by clicking on the "Advanced" button beside the search bar at the top of the page. You will be presented with the options shown in Figure 4. Set the first selection step as “Cross-references”. Then look for the subset under “Chemistry” and select “GuidetoPHARMACOLOGY”. Executing the query brings up 1999 results; the results can be filtered using the quick links on the left-hand side of the page. Restricting to reviewed (i.e. Swiss-Prot only) gives 1995, then selecting Human gives 1483.

   The combination of any selects can be copied, stored and re-used as a direct query to paste into the search bar (database:(type:guidetopharmacology) AND reviewed:yes AND organism:"Homo sapiens (Human) [9606]") (when doing this copy all the text in bold), as well as storing/sharing the execution as a URL.

   For ease of use, we suggest pasting the above query directly into the search box. If users wish to see how this query can be constructed within the Advanced query menu, clicking on Advanced after running the query will show the selections. The column display filters still need to be set manually, however.

2. For the 1483 entries, select “Columns” and a “Customize results table” page should open. For the example shown in Figure 5 the selected columns are Entry, Entry name, Protein names, Gene names, Length, EC number, Gene Ontology (molecular function) and Subcellular location [CC].

   The query and display options in UniProtKB are extensive and cover combinations of many annotation features and keywords, as well as over 100 cross-references that are mainly pointers to other databases. Note, the queries can be combined in Boolean sets and toggling which features and cross-references are shown in the table adds a second detailed level of selection. Yet a third layer of analysis is enabled by downloading the results from the display selections for sorting, sub-setting and ranking in Excel.

3. As an example, find the GtoPdb entries for proteases (with ligands) by combining GtoPdb with MEROPS (proteases and inhibitors database) (Rawlings et al., 2014; Rawlings et al., 2016) in the Advanced query section. Adding in “hydrolase” as a keyword gives a clean listing of 152, as this excludes endogenous protease inhibitor proteins from MEROPS. The storable query has this form (database:(type:guidetopharmacology) database:(type:merops) keyword:"Hydrolase [KW-0378]" AND reviewed:yes AND organism:"Homo sapiens (Human) [9606]"). The column display filters should be set as above, except that Gene Ontology can be dropped because “hydrolase” has been added as general function, and Tissue specificity (found under the Expression category) and 3D (found under the Structure category) should be added.

   For ease of use, we suggest pasting the above query directly into the search box. If users wish to see how this query can be constructed within the Advanced query menu, clicking on Advanced after running the query will show the selections. The column display filters still need to be set manually, however.

4. Use the Download button to download the result sheet for the 152 proteases to Excel. As the format select Excel and set to Uncompressed. It may be possible to rank the results in Excel by the number of Protein Data Bank (PDB) structures (Rose et al., 2017), but only after manipulating the text in the 3D column to remove the text strings and brackets. Nonetheless, for each protein, the number of 3D structures is given in parentheses after the name of each experimental method. A snapshot ranked by most populated in PDB is shown in Figure 6.

John Wiley & Sons
The examples in 2 and 3 indicate the wide flexibility to “slice and dice” the GtoPdb subset via the extensive query options in UniProtKB. This particular sequence has ordered each of the proteases by the number of PDB entries. Unsurprisingly, thrombin ranks top with 383 (but note the entry is technically prothrombin) but it is of interest that the second on the list is BACE1 with 357 because this is not yet an approved drug target. We can look at this in more detail in the next section but some caveats can be pointed out for Figure 6. Firstly there are some false negatives in the form of six missing tissue distribution annotations. These are probably just curation catch-up issues (i.e. these will eventually be added), but in the interim other cross-referenced sources, such as the Human Protein Atlas (Uhlen et al., 2015), will have expression data. The other caveat is that from the PDB counts alone, those with bound ligand structures cannot be filtered from apoprotein structures without ligands (although this splitting can be done via the PDB web portal). Nonetheless, for all the proteases in Figure 6 the vast majority of the PDB entries cross-referenced via UniProtKB will represent (mostly distinct) small-molecule ligands. GtoPdb will only curate those that are both clinically advanced and have documented activity data. The inconvenience of the Excel sheet is the absence of live links in the download. There is a workaround in the form of pasting directly from the UniProtKB result displays into an Excel sheet where the URLs are maintained in their original format.

Navigating from UniProtKB back to GtoPdb

5. Inspect the BACE1 entry at http://www.uniprot.org/uniprot/P56817 that has the most PDB structures for a protease target without any approved drugs. Using the Display menu on the left-hand side of the page, click on "Cross-references" to jump to that section of content. Scroll down and the cross-reference to GtoPdb (that was used for the query at the beginning of this section) can be found in the “Chemistry databases” section. Click on the GuidetoPHARMACOLOGY ID (2330) (http://www.guidetopharmacology.org/GRAC/ObjectDisplayForward?objectId=2330). This gives the detailed view page for BACE1; from the Contents menu click on Inhibitors (or scroll-down) to navigate to the inhibitor list.

Note that one of the ligands for BACE1, verubecestat, is used as the PubChem navigation example in Support Protocol 2.

Querying across multiple databases to find specific subsets

6. For this example, search for human proteins (with GtoPdb links) containing a domain signature for aspartyl proteases (Asp proteases). Return to the BACE1 UniProtKB entry (http://www.uniprot.org/uniprot/P56817) and select "Family & Domains" from the left-hand menu. Open up the InterPro (Finn et al., 2017) graphical display by clicking on “View protein in InterPro” under "Family and domain databases". One of the domain signatures specific for Asp proteases in general is “IPR001461 Aspartic_peptidase_A1”. Back on the UniProtKB site, run a query with the format (database:(type:interpro id:IPR001461) database:(type:guidetopharmacology) AND reviewed:yes AND organism:“Homo sapiens (Human) [9606]”) by pasting this query string into the search box. Note, a protease select is no longer needed since this is implicit within the aspartyl proteases (MEROPS class A1).

John Wiley & Sons
As above, if users wish to see how the query can be constructed, click on the Advanced menu button after running it, and check that the cross-references have been set to InterPro and GtoPdb, including the identifier for InterPro, with reviewed set to yes and the organism to human.

7. Customise the results table as before, with Entry, Entry name, Protein names, Gene names, Length, EC number, and this time include a cross-reference to the rare disease database, Orphanet (www.orpha.net) [Orphanet: an online database of rare diseases and orphan drugs. Copyright, INSERM 1997. Available at http://www.orpha.net Accessed (12/10/2017)]. Orphanet is found under the "Organism-specific" category. The results should look similar to Figure 7, which shows the table for seven Asp proteases.

This shows the use of detailed selects for specific subsets of GtoPdb targets in UniProtKB. Note, if the GtoPdb cross-reference was left out of the select, it would bring back 10 Asp proteases since NAPSA, PGA3 and PGA4 have no curated ligand interactions in GtoPdb. Another example could have used the Beta-secretase BACE PRINTERs signature as IPR009119, to retrieve just BACE1 and BACE2.

**BASIC PROTOCOL 2**

**FIND THE MECHANISM OF ACTION AND KNOWN TARGETS FOR A DRUG**

GtoPdb ligands are classified into groups based on their chemical properties, including synthetic organic compounds, natural products and their derivatives, peptides that are endogenous in mammals, peptides from other organisms or synthetically derived, metabolites, which includes small biogenic compounds such as hormones and neurotransmitters, antibodies, and inorganic compounds. In addition, ligands have also been grouped by virtue of a common property such as all approved drugs, and compounds labelled with, for example, a radioactive or fluorescent group. All these categories are available to browse via the GtoPdb ligand list available by mousing over the Ligands submenu of the main navigation bar. Quick links to these categories are also found on the left panel of the home page. This protocol describes how to find a drug via the list and locate information on its mechanism of action and known targets.

**Necessary resources**

**Software**

An up-to-date Web browser, such as Firefox or Google Chrome, with JavaScript enabled.

1. From the home page http://www.guidetopharmacology.org/, mouse over the Ligands submenu of the main navigation bar and click on “Ligand list”.

2. The ligand list opens at the first tab which is an alphabetical list of approved drugs. Click on the letter “E” to go to drug names starting with that letter.

3. Scroll down and click on the name “erlotinib” to go to its ligand page.
The ligand page opens with the Summary tab open (Figure 8), which shows the drug approval status, nomenclature, and database links.

4. Click on the “Clinical data” tab.

Information on this tab summarises the clinical use and therapeutic mechanism of action for the drug erlotinib, which targets the epidermal growth factor receptor (EGFR).

5. To access specific binding data at targets, click on the “Biological activity” tab (Figure 9). The data presented here confirm EGFR as the primary target of this compound, and this tab also provides information on other proteins erlotinib is known to interact with.

Figure 9 also shows a link near the top: “View interactive charts of activity data from ChEMBL and GtoPdb across species”. The charts available by clicking on this link provide a method of visualizing ligand activity at different targets and easily comparing across species. The tool combines data from GtoPdb and the medicinal chemistry database, ChEMBL (Bento et al., 2014), which covers a wider range of species. A table provided below the charts lists the source databases, data values and links to the references cited.

ALTERNATE PROTOCOL 2

SEARCH FOR LIGANDS SIMILAR IN STRUCTURE TO A KNOWN DRUG AND FIND THEIR TARGETS

It is also possible to search for ligands using various methods, which is useful where different isomers exist, the name-structure relationships are complex, or users wish to find ligands similar to a particular structure or substructure. This example demonstrates how to use the ligand structure search tool to find ligands sharing a substructure. It then takes one of the matching ligands and shows how to submit this structure to a further similarity search. The structure drawing tool can be accessed from the “Ligand search” link under both the “Advanced search” and the “Ligands” menus of the main navigation bar. It can also be opened with a ligand structure pre-loaded into the editor by clicking on the structure image on ligand pages (Figure 8). Ligands can also be searched by name (including wildcard searches), or chemical identifier such as PubChem CID, International Union of Pure and Applied Chemistry (IUPAC) International Chemical Identifier (InChI), or the shorter InChIKey.

Necessary resources

Software

An up-to-date Web browser, such as Firefox or Google Chrome, with JavaScript enabled.

Identifiers

A chemical structure in simplified molecular-input line-entry system (SMILES) format.
Perform a substructure search

1. Go to the home page at http://www.guidetopharmacology.org/

2. Mouse over the “Advanced search” menu of the main navigation bar and click on “Ligand search”.

3. Under “Chemical structure search” paste the SMILES string “N(c1ccccc1)c1ncnc2ccccc12” into the box and click “Import SMILES/SMARTS” (Figure 10).

   “SMILES” refers to chemical structures described in Simplified Molecular-Input Line-Entry System.
   “SMARTS” refers to SMILES Arbitrary Target Specification, a way of defining chemical structures with variable atoms of side groups.


   Chemical structures can also be drawn into the editor by hand, optionally using one of the templates shown at the bottom, and adding bonds with the button on the left and atoms from the list on the right. A user guide to Marvin.JS is available at https://docs.chemaxon.com/display/docs/Marvin+JS+User's+Guide.

5. In the first drop-down menu choose “Substructure” search. Leave the chemical class set to “Small molecules” and click on “Search the database”. Searches are performed using the Pinpoint chemical cartridge, version 2.4.2, Dotmatics (https://www.dotmatics.com/).

6. The search results are shown in a grid and the default ordering is by increasing number of rotatable bonds and molecular weight. Examine the matching structures and notice how they all contain the input structure displayed at the top of the page.

7. Change the order of results by selecting “Alphabetical” in the drop-down menu next to “Order results by”. Click on “Update”.

8. The order of results is now alphabetical by ligand name. Scroll down to find erlotinib.

Perform a similarity search

9. On the erlotinib box click on “Use in search”.

10. The erlotinib structure is loaded into the chemical drawing tool. Change the type of search to “Similarity: moderate (>70%)”. Click on “Search the database”.

John Wiley & Sons
When GtoPdb ligand structures are loaded into the drawing tool, any chiral specification is removed (i.e. chiral bonds are shown in the flat form) so that substructure or similarity searches are not confounded by the presence of different isomers.

11. The results now contain only structures that are at least 70% similar to erlotinib. The targets of each compound are summarised underneath the structures. Click on a ligand name or image to access the ligand page, or on a target name to access the target page.

Similarity searches use the Tanimoto coefficient to compare the similarity between fingerprints generated from chemical structures and return all results above a certain threshold.

SUPPORT PROTOCOL 2

NAVIGATING RECIPROCAL LINKS TO PUBCHEM: SIMILARITY SEARCHING AND SOURCE INTERSECTING

The search for ligands similar in structure to a known drug, as described in Alternate Protocol 2, is internal and consequently encompasses only those entries that GtoPdb has SMILES for. Most of these can be classified as small-molecules but include small to moderate-sized peptides up to our largest example, ligand ID 7387, lixisenatide, with 44 amino acids. Using the database links provided for ligands, users can go “out” to many other databases that are chosen for various types of complementarity (i.e. they can add value over and above what our ligand entries offer). In many cases these links are reciprocal, meaning users of the external resource can come “in” to GtoPdb. These external sources are each different and users cannot thus easily become familiar with all of their individual features and content. For bioactive chemistry, PubChem is the most comprehensive, so this will be used as an example. In addition, as explained, it is our primary source in most cases so we make considerable effort to map our ligand IDs to the correct PubChem structures. It is also important to note that we submit not just structures but also PubMed literature links. In addition, PubChem is an entry point (along with those for proteins, genes, PubMed articles, MeSH terms and others) into the NCBI Entrez system (Gibney et al., 2011). We would recommend anyone who regularly needs to engage with small molecule ligands beyond an introductory level in GtoPdb (whatever their background discipline) should become acquainted with PubChem at least to some level. The associated challenge is that just a click instantly takes a user from our modest chemical space neighbourhood of ~7,000 structures, out into one of over 90 million. There is a very wide range of operations that can be executed in PubChem but there is only space here to introduce a few basic protocols. This means there are many features and operation choices in the figures that we cannot expand on here, but these are detailed in the PubChem Help documentation (https://pubchem.ncbi.nlm.nih.gov/help.html).

Necessary resources

Software

An up-to-date Web browser, such as Firefox or Google Chrome, with JavaScript enabled.

A text editor, such as Microsoft WordPad, for temporary paste-across storage.
A program capable of archiving lists, such as Microsoft Excel or LibreOffice Calc.

For set intersecting, the Venny tool is useful (http://bioinfogp.cnb.csic.es/tools/venny/).

Identifiers

PubChem substance identifier (SID) and PubChem compound identifier (CID)

Navigating across to PubChem records

1. Find the ligand entry for “verubecestat” by searching for its name in the search box at the top right of the GtoPdb site, or go directly to http://www.guidetopharmacology.org/GRAC/LigandDisplayForward?ligandId=8699.

   This is a Merck development compound against BACE1 for Alzheimer’s disease (AD), originally declared as MK-8931 but now with the International Nonproprietary Name (INN) verubecestat. The first Phase 3 clinical study has been stopped because of lack of efficacy. Note, the searches below can be carried out with any of the other 16 ligand entries for BACE1.

2. Check the Similar ligands tab. This produces just one match (within GtoPdb) with a compound (example 20) from a patent (ligand ID 7853).

   This comes from a Roche patent (WO2010128058) for BACE2 inhibitors claimed for diabetes. As the paralogue of BACE1, GtoPdb includes nine BACE2 inhibitor ligands (http://www.guidetopharmacology.org/GRAC/ObjectDisplayForward?objectId=2330). Significantly, potent verubecestat inhibition is common to both targets.

3. Click back to the Summary tab. As shown in Figure 11, click on the link to PubChem CID within the Database Links section. This will now open the link https://pubchem.ncbi.nlm.nih.gov/compound/51352361 in a new tab.

   Many links, features and statistics are shown on the CID home page. Their density is roughly proportional to compound “popularity”. Generally, older approved drugs will have many pages whereas patent-only compounds have less information available. Verubecestat has a middle level of popularity with facet 5.3 (“Substances”) indicating 21 submissions of the same structure as SIDs from different sources (including GtoPdb).

Using PubChem pre-set similarity search and clustering

John Wiley & Sons
4. Use the Contents link “5 Related Records” to navigate the CID entry to reach the 5.2 facet for “Related Compounds” and the subheading “Similar Compounds” (Figure 12). Then click on the highlighted “526 records”. The results, features and options for this page are shown in Figure 13.

Clicking “Similar Compounds” executes, on the fly, a 90% Tanimoto similarity search. This actually shows 684 results (which is presumably because the search executes as a live-link, but the number shown, 526, was pre-computed and more compounds have since been added). It is important to note that % similarity settings (Tanimoto-based or by other methods) are not identically parameterized between different databases. This means the GtoPdb and PubChem structural “neighbours” may not give the same results but there are clearly likely to be more in the latter. Many filters and sub-sets can be accessed via the right-hand side facets shown in Figure 13.

5. Click on the “Structure Clustering” logo near the top right (Figure 13). This will generate the tree result seen in Figure 14. Note, it is sometimes necessary to expand the tree to get the full display. To do this, click on the image on the blue dots between the two blue lines representing the top and bottom branches of the tree. On the pop-up links, click on “Expand Subtree”.

There are many options that can be explored on this PubChem clustering display. One of those shown in Figure 14 is “Common substructure” that can be toggled in for any branch of the tree by clicking on the branch node blue spots. Note also the entire scoring matrix can be downloaded using the Export links.

Narrowing the search, 3D display and substructure picking

6. Return to the list of 684 similar compounds to verubecestat (CID 51352361).

Triaging 684 similar structures to CID 51352361 as a well-populated chemotype is a challenge. It can be established, by checking intersects with PubChem sources such as SureChEMBL (Papadatos et al., 2016) and/or IBM (https://www.ibm.com/), that most of these are patent-extracted structures from the many pharmaceutical companies (and some academic groups) working on BACE1 inhibitors. Filtration can of course be informative. As an example, we can select (from the right facet) “BioAssays, Active”, to view 17 entries where the activity has been extracted from published papers by ChEMBL (Bento et al., 2014). Two follow up approaches can be taken here, the first is simply to narrow down the search stringency to get a smaller set of matches. This can then be 3D clustered as an adjunct approach to 2D. The second option is to use substructure searches. While this may expand the hit list in PubChem, the protocol will show how to transfer this across to the smaller search space of GtoPdb.

7. As an alternative to the pre-set 90% similarity search used above, for any individual CID it is possible to set up a customised search as shown in Figure 15. From the PubChem Compound home page (https://pubchem.ncbi.nlm.nih.gov/), click on Structure Search on the right, and then on the tab Identity/Similarity. Underneath, click on the tab for CID, SMILES/SMARTS, InChI. Paste “51352361” into the box, and edit the Options to match those in Figure 15. The result display is the same as in Figure 13 but the result numbers in stringency order are: 93% = 288, 95% = 151, 96% = 110, 97% = 74. In the same way as

John Wiley & Sons
performed above, run the clustering for the 74, but this time set the display for 3D rather than 2D. The result is shown in Figure 16 and gives good shape-based indications on similarities within the series. Note that in most cases it is possible to find SAR data specified in the patent documents that the structures were extracted from.

8. Mouse over the red branch lines to display the “common substructure” for a set of ligands (an example of which is shown in Figure 14). A SMILES representation of this substructure can be obtained by using either the PubChem sketcher or MarvinJS in GtoPdb (described in Alternate Protocol 2) to edit the structure for verubecestat down. The example is represented by the SMILES string FC2=C(C)C=CN(C(=O)C1=NC=CC=C1)(H)C=C2. On the PubChem Structure Search, click on Substructure/Superstructure, then on CID, SMILES/SMARTS, InChI, and paste in the SMILES string. At the time of writing, the example returns 6160 CIDs, and these can be filtered for structures of particular interest.

A link to the PubChem sketcher can be found under the Draw a Structure tab, which is next to the CID, SMILES/SMARTS, InChI tab.

9. Go to the GtoPdb Chemical structure search at http://www.guidetopharmacology.org/GRAC/chemSearch.jsp. Paste FC2=C(C)C=CN(C(=O)C1=NC=CC=C1)(H)C=C2 into the input box, click “Import SMILES/SMARTS” and hit “Search the database”: The result is shown Figure 17.

Users can transition any standardised query between PubChem and GtoPdb. In this case we are running the equivalent substructure search from PubChem described above to search against the smaller, highly curated ligand set of GtoPdb. Here we see the complementarity of working between the small-data and big-data resources by exploiting different search stringencies, filtration and display options. In the PubChem 6160 results, clicking on “BioAssays, Active” reduces the number to 339, while clicking on “Protein 3D Structures” reduces this to 17, and 16, respectively. In the GtoPdb results we can see the substructure matches both BACE1 and BACE2 inhibitors.

BASIC PROTOCOL 3

LEARN HOW TO DOWNLOAD DATA AND VIEW IT IN MICROSOFT EXCEL

GtoPdb includes various file download and data access options. This includes data files for targets, ligands and interactions in comma-separated value (CSV) format. These represent the most requested datasets; other datasets and file formats are available by contacting the team (enquiries@guidetopharmacology.org). There is also the option of downloading an SQL file containing the whole PostgreSQL database (http://www.postgresql.org/), which would require a recent version of PostgreSQL to be installed. Basic Protocol 3 explains how to download a CSV file and open it with a program capable of viewing this file format. GtoPdb recently added web services to facilitate computational access to the data, and Alternate Protocol 3 describes how to build the URLs needed to access data on targets, ligands and interactions in a computer program.

Necessary resources

John Wiley & Sons
Software

An up-to-date Web browser, such as Firefox or Google Chrome, with JavaScript enabled.

A program capable of viewing CSV format files, such as Microsoft Excel or LibreOffice Calc.


2. Scroll down to the section called “Download data files”.

   A description of each data file can be accessed by clicking on the link above the file list. This is available at http://www.guidetopharmacology.org/DATA/file_descriptions.txt. Users are encouraged to read this document whilst looking at the downloaded files for the first time as it describes the content of each column. This document will be updated if the file format changes.

3. Click on the “Download CSV file” link next to “target and family list”.

4. Either save the file or select the program to open it with. If it is being opened with LibreOffice a Text Import wizard will open. Select “Separated by” and “comma” under Separator Options. With Microsoft Excel, if the file doesn’t open automatically, it can be saved and then loaded using the Open command under the File menu, choosing “Text Files (*.csv)” as the file type. Alternatively, if Excel does not open the CSV file correctly, try importing the file using the “From text” option in “Get External Data” under the Data menu. There, select “Delimited”, and set the file origin as “Unicode (UTF-8)”. In the second step, select “Comma” as the delimiter. In the third step, the column type can be set, which can be useful to prevent Excel from changing certain numbers (such as many identifiers) to scientific format or even to date format (to do this click on the column and choose format “Text”).

   Some users have reported problems opening CSV files with Microsoft Excel 2016 where their default list separator is a semicolon. In this case, all text after a semicolon will be lost. If this has happened, there will be no text in any of the columns after column I (“synonyms”) for several rows of the target and family CSV file (see, for example, rows 3 and 5). This is because synonyms sometimes contain semicolons as part of the HTML codes for special characters such as the Greek alpha, beta etc.. The list separator can be changed under Windows in "Control Panel / Region / Additional Settings / List separator: ".

ALTERNATE PROTOCOL 3

USE WEB SERVICES TO GAIN COMPUTATIONAL ACCESS TO TARGET AND LIGAND DATA

GtoPdb offers a range of REST (Representational State Transfer) web services described in full at http://www.guidetopharmacology.org/webServices.jsp. Requests to web services will return data in JavaScript Object Notation (JSON) format, which is both easily understood by humans and machine-readable. REST APIs (Application
Programming Interface) use HTTP to transfer data to clients, providing a simple way to incorporate external data in applications.

The GtoPdb API includes options to retrieve lists of targets, ligands, interactions, and references, as well as data for individual entries. Various parameters can be included in the URL which have the effect of filtering data in various ways. For instance, interaction lists can be filtered by ligand action (e.g. agonist, inhibitor), affinity type (e.g. $pK_a$, $pK_i$), target type (e.g. GPCR, NHR), approved drugs, or minimum affinity value. Ligand lists can be filtered by chemical property (e.g. number of rotatable bonds or hydrogen bond acceptors), type (e.g. synthetic organic, antibody), or searches performed by name, external database identifier, or SMILES string (in the same way as described in Alternate Protocol 2). Targets can be filtered by type or searched by name, gene symbol, Enzyme Commission (EC) number, or another database identifier. Alternate Protocol 3 explains how to build URLs to fetch data for sets of targets, ligands and interactions, and includes examples for accessing specific data on individual targets and ligands. Further examples are provided at http://www.guidetopharmacology.org/webServices.jsp#examples.

Necessary resources

Software

An up-to-date Web browser, such as Firefox or Google Chrome, with JavaScript enabled.

(Optional) a tool for optimally viewing JSON format data (modern Web browsers can display JSON data, but it can be helpful to view it in a “pretty” format). For example, the Firefox add-on JSONView (https://jsonview.com/) pretty-prints JSON data and provides collapsible array and object sections.

1. All URLs are built by extending the base URL: http://www.guidetopharmacology.org/services/

   Note, accessing the base URL itself doesn’t display anything.

Accessing target information

2. View the target web services documentation at http://www.guidetopharmacology.org/webServices.jsp#targets, which lists all the services that are available for targets and families.

3. To access target data append “targets” to the base URL. Thus, the full target list is available at http://www.guidetopharmacology.org/services/targets. To restrict the list to GPCRs, use http://www.guidetopharmacology.org/services/targets?type=GPCR. Type can be one of: ‘GPCR’, ‘NHR’, ‘LGIC’, ‘VGIC’, ‘OtherIC’, ‘Enzyme’, ‘CatalyticReceptor’, ‘Transporter’, ‘OtherProtein’, or ‘AccessoryProtein’.

   In this case, the URL contains a parameter (“type”) and a value (“GPCR”), which are added after a question mark at the end of the URL. Multiple parameters can be included by separating them with an ampersand character. The returned JSON results include basic information on each target, covering GtoPdb ID,
4. Search the target list by appending one of a few different parameters to the end of the URL. All possible parameters are listed in the documentation, but one useful example is to do a UniProtKB accession look-up. This is achieved using a URL of the form

http://www.guidetopharmacology.org/services/targets?database=UniProt&accession=P08908. For UniProt searches, the inclusion of the “database=UniProt” part (and thus the ampersand) is redundant, because UniProt is the default database used for accession searches. However, the database name for any other type of accession look-up would need to be specified and a full list of accepted names is given in the documentation:


*UniProtKB accession searches can normally be done using any one of the human, mouse or rat SwissProt accessions which form the triad normally mapped to each GtoPdb target (if available). Therefore, any of the 3 accessions (P08908, Q64264, P19327) would return the same result: the 5-HT₄ receptor (target ID 1).*

5. Access specific targets by appending the target ID to the URL. For example, for target ID 1 use

http://www.guidetopharmacology.org/services/targets/1. The JSON result shows a summary of the target data, which is similar to the list view.

6. Access further information on an individual target by appending the relevant service name after the target id.

For example, to access information on the interactions that target ID 1 is involved in use

http://www.guidetopharmacology.org/services/targets/1/interactions.

*The interaction data and references are listed in full, and empty fields will usually have either a null value, or an empty string. Affinity data are reported as -log₁₀ values, and raw data (where known) are reported as the molar values using HTML codes for superscripts to specify the power value. If the value for “conciseView” is true, this interaction has been selected to show on the website summary view of the target (described in Basic Protocol 1 and Figure 1).*

7. Experiment with adding different service names to the end of the URL. For example, target ID 1’s functional data can be retrieved using http://www.guidetopharmacology.org/services/targets/1/function.

8. Access family-specific information in a similar manner by adding “targets/families” to the base URL. Thus, http://www.guidetopharmacology.org/services/targets/families gives the complete family list in GtoPdb. This can also be filtered by type as described in step 3. Try obtaining a list of Enzyme families with the URL

http://www.guidetopharmacology.org/services/targets/families?type=Enzyme.

*The returned results include each family’s ID, name, the target member IDs, and the family IDs of any parent families or subfamilies in the family hierarchy, which models the display of families on the website (see Basic Protocol 1).*
**Accessing ligand information**

9. Access a description of the ligand web services at
   
   http://www.guidetopharmacology.org/webServices.jsp#ligands.

10. Ligand URLs are constructed in a similar way to target URLs; to access ligand data append “ligands” to the base URL. The full list of ligands is available at http://www.guidetopharmacology.org/services/ligands and this can be filtered and searched in several ways. For example, retrieve a list of natural products by adding the type parameter: http://www.guidetopharmacology.org/services/ligands?type=Natural product.

   The type parameter can be used with any of the following options: 'Synthetic organic', 'Metabolite', 'Natural product', 'Endogenous peptide', 'Peptide', 'Antibody', 'Inorganic', 'Approved', 'Withdrawn', 'Labelled', or 'INN'. These correspond to the subsets available via the website ligand list (see Basic Protocol 2), except for “INN” which is an additional list of all GtoPdb ligands with INNs, and “Withdrawn”, which is a set of previously approved, but now withdrawn drugs (at least by some jurisdictions; they may still be available in other countries).

11. The ligand list is searchable in a similar way to that for targets described in step 4. However, there are a few differences, as described in the documentation. This example finds ligands by InChIKey. Compare the results for a full InChIKey look-up at
   
   http://www.guidetopharmacology.org/services/ligands?inchikey=USQOVYLRWBOSQC-OAHLLOKOSA-N with a “backbone” look-up at
   
   http://www.guidetopharmacology.org/services/ligands?inchikey=USQOVYLRWBOSQC.

   The first URL uses the full InChIKey for exact matching, while the second uses the first 14 characters of the InChIKey to perform a backbone match (i.e. matches to both the isomers).

12. Ligands are also searchable by SMILES strings based on exact match, substructure or similarity in a similar way to that described in Alternate Protocol 2. For example, type the following URL to look for ligands similar to the structure of 5-fluorouridine (ID 4614):
   
   http://www.guidetopharmacology.org/services/ligands/similarity?smiles=OC[C@H]1O[C@H][C@@H](C@H)10)n1cc(F)c(-O)[nH]c1=O. This uses the default minimum of at least 85% similarity, which can be changed by adding the “similarityGt” parameter and specifying a value. Try changing this to 75% to see how it affects the results by typing this URL:
   
   http://www.guidetopharmacology.org/services/ligands/similarity?smiles=OC[C@H]1O[C@H][C@@H](C@H)10)n1cc(F)c(-O)[nH]c1=O&similarityGt=75.

13. An individual ligand entry is obtainable by adding the ligand ID at the end of the URL. As an example, retrieve the ligand with ID 4789 (5-fluorouracil) with the URL

   http://www.guidetopharmacology.org/services/ligands/4789. Looking at the results for this query, notice that there are two values under the field “prodrugIds”. These are the ligand IDs for the prodrugs capecitabine (ID 6799) and floxuridine (ID 4801), which are enzymatically converted into the active metabolite 5-fluorouracil.

John Wiley & Sons
Retrieve further data on these prodrugs by using their IDs in the ligand URL, i.e. http://www.guidetopharmacology.org/services/ligands/6799 and http://www.guidetopharmacology.org/services/ligands/4801.

14. Additional data types for ligands are accessible by appending the relevant web service name to the URL. For example, access the interaction list for ligand ID 5000 (human IL-9) using http://www.guidetopharmacology.org/services/ligands/5000/interactions. Note that one of the interaction results (interaction ID 83153) has a zero value in the “targetId” field, but 5000 in the “ligandAsTargetId” field which indicates that this ligand is actually the target of the interaction (by ligand 8934).

Accessing interactions


16. Access a list of interactions for approved drugs by using the “approved” parameter (http://www.guidetopharmacology.org/services/interactions?approved=true). The full list of interactions is available at http://www.guidetopharmacology.org/services/interactions, however, note this list is very long so can take a long time to load. There are several ways to filter the list, such as by target type (e.g. for GPCRs use http://www.guidetopharmacology.org/services/interactions?targetType=GPCR) or by ligand action (e.g. for channel blockers use http://www.guidetopharmacology.org/services/interactions?type=Channel blocker).

Interactions can also be filtered directly by ligand structure using SMILES or InChIKeys.

The interaction list results include a summary view of the interaction data for a target-ligand pair. The individual data point interaction IDs are given in the “dataPointInteractionIds” field. These can then be used to obtain the full data about an interaction by putting the interaction ID at the end of the URL. For example, interaction ID 1697 is retrievable at http://www.guidetopharmacology.org/services/interactions/1697.

BASIC PROTOCOL 4

BROWSE TO FIND APPROVED DRUGS AND THEIR PRIMARY TARGETS FOR AN IMMUNOLOGICAL DISEASE

A major data extension of the Guide to IMMUNOPHARMACOLOGY (GtoImmuPdb) has been to associate targets and ligands to diseases of immunological relevance. By doing this, users can search or browse by disease to identify the key targets and ligands involved. Information about disease associations can be readily accessed from the GtoImmuPdb portal home page, via the Disease panel. Disease data can also be accessed from anywhere on GtoImmuPdb from the Disease menu of the main navigation bar. The GtoImmuPdb disease association information contains curator comments, giving details about why an association exists between a given disease and targets or ligands, and any relevant literature references.
Wherever possible, disease terms are mapped to external disease resources. Disease mapping is a manual process identifying equivalent terms in external resources, adding the unique identifiers and using these to hyperlink to the resources. For example, the disease chronic lymphocytic leukemia (CLL) maps to the Disease Ontology (DOID:1040) term, "chronic lymphocytic leukemia", the OMIM (151400) term "Leukemia, chronic lymphocytic; CLL"; and the Orphanet (ORPHA67038) term "B-cell chronic lymphocytic leukemia". Mapping disease terms to other resources in this way is important for several reasons. Firstly, it provides clarity in understanding which disease is being referred to. In many cases, the same disease may be named differently in different papers, studies or data resources. Secondly, the mappings aid interoperability, meaning that data can be more easily linked between different resources. Finally, it enables us to provide external links allowing users to quickly navigate to other resources of disease information, bringing added value to the data. For example, GtoImmuPdb does not currently include detailed disease definitions. For CLL, clicking the external link to OMIM provides a disease description, details of clinical features, pathogenesis, and more. Linking to Orphanet also gives a disease description, and other external disease identifiers to ICD and MeSH (for example), while the Disease Ontology provides a hierarchical, ontological view of the disease, placing it into context with similar disease classifications.

The GtoImmuPdb disease section presents a list of diseases associated to targets and ligands. For targets, it also displays approved drugs that interact with the target (where the target is considered a primary target). For ligands, it includes summaries of their clinical use and distinguishes the approved drugs from those under development. The aim here is to provide a useful summary of the molecular targets and key ligands that modulate a disease, without requiring a user to navigate away from the disease page. Further details about the individual targets and ligands can be found on their database pages by clicking on their respective names.

Necessary resources

**Software**

An up-to-date Web browser, such as Firefox or Google Chrome, with JavaScript enabled.

**Finding known targets and drugs for an immunological disease (example, Chronic lymphocytic leukemia (CLL)).**

1. Begin at the GtoImmuPdb portal home page (http://www.guidetoimmunopharmacology.org) (Figure 19).

   *The GtoImmuPdb portal shares a similar style with GtoPdb, but has its own header, menu bar, colour scheme and logo to distinguish it. There are main panels that can be used as routes to browse for process, cell type, target, ligand, and disease data. GtoImmuPdb uses the same underlying database as GtoPdb, so users can find the same data in both resources. When viewing pages via GtoImmuPdb, however, different browsing options are available, data of immunological relevance are highlighted, and are also prioritised in searches.*

2. Under the section Disease, select “Immuno Disease to Target Associations”. Alternatively, select "Immuno Disease to Targets" from the "Disease" drop-down menu of the navigation bar.
All immunological diseases which are associated to targets in the database are listed, alphabetically, in the table (Figure 20). Each section contains the details for one disease. By default, the details of the associated targets are hidden (in order to provide an at-a-glance view of the diseases when the page first loads). Details for a given disease can be shown by clicking the "more details" link. The two grey buttons in the top right can be used to expand (show) or hide the details for every disease. For each disease, the total number of curated target associations is displayed, showing the names of the first few targets.

3. Scroll down the table to find the disease of interest, “Chronic lymphocytic leukemia”.

The row for Chronic lymphocytic leukemia (CLL) shows “Disease X-Refs”; these are cross-mappings to other disease resources, in this case OMIM, Orphanet and the Disease Ontology for CLL. The importance of these mappings is explained in the introduction to this protocol. Clicking on the ID next to OMIM (151400) opens a new tab displaying the OMIM entry for #151400, “Leukemia, Chronic Lymphocytic; CLL”. OMIM pages typically provide details about the disease, its clinical features, genetics, and more.

4. The summary shows that there is only 1 target associated with CLL. Click on the “more details” link to see full information about this target association (Figure 20).

In the expanded section it shows the target name(s) associated with the disease and curator comments about the association. The comments indicate that CD20 is a molecular target of ofatumumab and obinutuzumab, both of which are noted as CLL therapeutics. It also includes a list of ligands that interact with the target, where it is considered to be a primary target. There are four ligands listed against the primary target of CD20, which are associated with CLL. (Note, there may be other ligands curated as being associated with CLL not listed here, but their target would be different to CD20 and its association with CLL may be unclear, which is why it doesn’t appear in the list. Other CLL ligands not associated to CLL targets would be included under the Ligands tab, as described in Alternate Protocol 4.) The icons next to the ligand name indicate whether the ligand is an approved drug (medicine bottle) or if the target is a primary target of the ligand (number 1 in a bullseye).

5. Click on the ofatumumab link in the ligand column to view the GtoPdb ligand summary page for ofatumumab. The link opens on the Clinical data tab which reveals further information about the ligand’s therapeutic associations.

Clicking through to the ligand summary allows users to look more closely at what is known about the ligand. The GtoPdb ligand summary page presents data about the immunopharmacology ligands under five main tabs: Summary, Biological activity, Clinical data, References, and Immunopharmacology. The clinical use summary under the Clinical data tab provides details about ofatumumab’s use to treat CLL.

6. Click on the Biological activity tab for ofatumumab to view a list of targets for this drug. The data show that ofatumumab is an antibody only known to interact with one target: CD20.

The data table includes the recorded affinity of the interaction shown on the \(-\log_{10}\) scale, and the reference(s). Clicking the arrow on the right gives a bit more information about the assay including the raw affinity data.
**ALTERNATE PROTOCOL 4**

**BROWSE TO FIND APPROVED DRUGS FOR AN IMMUNOLOGICAL DISEASE VIA LIGAND-DISEASE ASSOCIATIONS**

Basic Protocol 4 takes a route whereby the user browses a list of target-disease associations, looking to identify approved drugs that modulate these targets for a known immunological disease. Similar data can be accessed by browsing lists of ligand-disease associations as an alternative, which is described in this alternate protocol.

**Necessary resources**

**Software**

An up-to-date Web browser, such as Firefox or Google Chrome, with JavaScript enabled.

**Finding known drugs for an immunological disease (example, Chronic lymphocytic leukemia (CLL)).**

1. Begin at the GtoImmuPdb portal home page (http://www.guidetoimmunopharmacology.org) (Figure 19).

2. Under the section Disease, select “Immuno Disease to Ligand Associations”. Alternatively, select "Immuno Disease to Ligand" from the "Disease" drop-down menu of the navigation bar.

   *All immunological diseases which are associated to ligands in the database are listed, alphabetically, in the table. Each section contains the details for one disease. By default, the details of the associated ligands are hidden (in order to provide an at-a-glance view of the diseases). Details for a given disease can be shown by clicking the "more details" link. The two grey buttons in the top right can be used to expand (show) or hide the detailed sections for every disease. For each disease, the total number of curated ligand associations is displayed, showing the names of the first few ligands. Users may notice that there are more disease-ligand associations than there are disease-target associations (described in Basic Protocol 4). This is because, in some cases, the actual therapeutic target and mechanism of action for the (potential) drug against the disease has yet to be confirmed, or information is not readily available in the literature or clinical trial databases.*

3. Scroll down the table to find the disease of interest, “Chronic lymphocytic leukemia”.

   *The row for Chronic lymphocytic leukemia shows “Disease X-Refs”; these are cross-mappings to other disease resources, OMIM, Orphanet and the Disease Ontology, for the same disease. The number of ligands associated to the disease is shown, with the names of the first few.*

4. Click on the “more details” link to expand the information on ligand associations to Chronic lymphocytic leukemia (Figure 21).
The expanded section for CLL shows the ligand name(s) associated with the disease in the first column. If the ligand is an approved drug this is indicated by the small medicine bottle icon and text stating which regulatory body approved the drug. Curator comments about the disease-ligand association are shown in the second column, and the ligand’s clinical use comment is summarised in column 3. References are listed in column 4.

5. Observe information on ligand associations with CLL.

When compared to the list of ligands against target CD20 which is associated with CLL as described in Basic Protocol 4, there are additional ligands listed here. Idelalisib, ibrutinib and milatuzumab are listed in addition to the four ligands for CD20. This means that they are associated with CLL, but don't interact with CD20 (as a primary target). The information shows that idelalisib and ibrutinib are additional approved drugs for CLL, and milatuzumab is an orphan drug for CLL undergoing a phase 2 trial.

6. Click on the idelalisib link in the Ligand column to view the ligand summary page for idelalisib. The link opens at the Immunopharmacology tab which shows all immunological disease associations for this ligand.

7. Click on the clinical data tab to find more information about idelalisib’s status as an approved drug for CLL in combination with rituximab (Figure 22). Click on the biological activity tab to find data for PI3Kδ as the primary target of idelalisib, as well as other secondary targets.

SUPPORT PROTOCOL 3

USE SEARCH TOOLS TO FIND TARGETS AND LIGANDS ASSOCIATED TO A DISEASE

Basic and Alternate Protocols 4 describe how to find targets and ligands associated to an immunological disease, via the GtoImmuPdb portal. Many diseases are also described in the main GtoPdb database, but there is no specific disease listing page through which to access this information directly. Many targets contain related information in their detailed pages, covering disease-causing mutations, drugs that have been developed to target them, rodent models of diseases, and changes in gene expression in disease states. Drugs that have been approved to treat a disease and those showing potential in trials will have clinical use comments mentioning the disease on their summary pages. The easiest way to locate information on targets and ligands related to a specific disease in GtoPdb is to search for it by name. This protocol shows an example of searching for Alzheimer’s disease, and explains how to interpret the results.

Necessary resources

Software

An up-to-date Web browser, such as Firefox or Google Chrome, with JavaScript enabled.

1. Go to the home page at www.guidetopharmacology.org/.

2. In the search box in the top right, enter the term “Alzheimer’s” and click “Search Database”.

John Wiley & Sons
The search function can accept various different spellings of this term: Alzheimer’s, Alzheimers and Alzheimer all work. “Alzheimer’s disease”, however, brings back 5 fewer results, as sometimes the term “disease” does not appear next to “Alzheimer’s”.

3. Examine the results of the search. The results page shows 109 results, mainly matching to targets and ligands in the database (Figure 23). The term is also found in a few family overviews and introductions (for example, muscarinic acetylcholine receptors on page 5). Each page shows 10 results. Move between pages by clicking on the numbers at the bottom of the page.

The results are ranked by how many times they match the search term and how common the term is within a particular database field. Additional weightings are also applied if a term matches to a target or ligand name or synonym, so these results appear near the top of the list. The order of results can be changed by clicking on the drop-down menu in the top right. The same search can also be run from the GtoImmuPdb home page at http://www.guidetoimmunopharmacology.org. The number of returned results is the same, but the order is slightly different, because targets and ligands that are relevant to immunopharmacology are up-weighted so they appear higher up the list.

4. Look at the matches under the target GPR3. The term has matched to text in the database tables “Consequences of altering gene expression” and “Mutations/pathophysiology”. More information can be found by clicking on the highlighted target name to go to its database page.

5. Narrow the search to just targets by clicking on the “Target search page” link at the bottom.

6. Enter “Alzheimer’s” into the Target text search box and click “Search the database”.

The results can be narrowed to certain database fields by selecting one or more field names in the list.

7. Examine the results. There are 30 matches to target pages, which includes the family overviews and introductions.

A similar search can be conducted just for ligands by clicking on “Ligand search page” at the bottom of the result list (or going to Advanced search > Ligand search in the navigation bar) and entering “Alzheimer's” in the Ligand name/text search box.

BASIC PROTOCOL 5

BROWSE TO EXPLORE DATA FOR IMMUNOLOGICAL CELL TYPES

The GtoImmuPdb project has enriched GtoPdb with data on immunological cell types. Targets have been mapped to several top-level immunological cell-type categories, providing a basis for organising, navigating and searching GtoPdb data in ways that are intuitive to immunologists. Cell type data are accessible from the GtoImmuPdb portal home page,
via the Cell Types panel. They can also be reached from anywhere on the GtoImmuPdb site using the Cell Types menu in the main navigation bar. This protocol introduces how to browse and view cell type data and link this to pharmacological targets.

Necessary resources

Software

An up-to-date Web browser, such as Firefox or Google Chrome, with JavaScript enabled.

Exploring cell type data in the Guide to IMMUNOPHARMACOLOGY

1. Begin at the GtoImmuPdb portal home page (http://www.guidetoimmunopharmacology.org) (Figure 19).

2. Under the section Cell Types, select “View immuno cell types home page”. Alternatively, click on Cell Types in the menu-bar, or directly navigate to http://www.guidetoimmunopharmacology.org/GRAC/CelltypesForward.

   This page describes the top-level cell type categories used in GtoImmuPdb (Figure 24). Quick links at the top jump to the start of each cell type category section on the page. For each category, there is a brief description, alongside the Cell Ontology (CL) terms included under that category. Cell type data in GtoImmuPdb are underpinned by the use of CL terms (http://obofoundry.org/ontology/cl.html) (Bard, 2005, genome biology 6), comprising a set of cell type terms and identifiers that are mapped to top-level categories. Using the CL provides a controlled vocabulary and external identifiers, which supports interoperability of the data. Another benefit is that it supports high-resolution annotation of cell types to targets while also enhancing searches across cell type data. This means users can search using parent or child terms (which can be specific cell type terminology), and the search results will bring back all relevant data mapped to that term in GtoImmuPdb.

3. Scroll down to the section on Dendritic cells.

   The description explains that this category includes cell types that are children of the parent term "dendritic cell (CL:0000451)" from the Cell Ontology. Clicking on the CL ID links to EMBL-EBIs Ontology Lookup Service (Jupp et al., 2015), which provides more information on the term and displays it in the context of the CL hierarchy.

4. Click on the heading for the "Dendritic cells" section. Alternatively, select "Dendritic cells" from the "Cell Types" menu item of the navigation bar.

   This page shows all targets associated to the immunological cell type category "Dendritic cells" (DCs) (Figure 25 panel A). The table is split into sections for each target class (e.g. GPCRs, Enzymes), and quick links are available in the top right corner. The pull-down menu in the top left can be used to switch cell type category. The first column shows the target name (linking to its GtoPdb detailed view page) and family name (in
parentheses), the second shows curator comments about the association, the third gives Cell Ontology terms associated to the target, and the final column shows general immunopharmacology comments about the target.

5. Scroll down the page to the bottom of the GPCRs section. The final row shows the target XCR1 (Figure 25 panel B).

The information for XCR1 shows that it is part of the chemokine receptor family. The curator comments give specific details about XCR1’s expression in immunological cell types; in this case highlighting the specificity of XCR1 to a specific subset of dendritic cells (CD141-positive DCs). The third column shows Cell Ontology terms associated with XCR1, namely “CD141-positive myeloid dendritic cell (CL:0002394)”. This is an example of a specific, high resolution CL term annotated to a target.

6. Click on the 'XCR1' name in the first column to view its detailed page.

This displays the GtoPdb detailed page for the target with the GtoImmuPdb view switched on. Under this view, in the Contents section and lower down the page, tables with data of immunological relevance are highlighted in blue. A toggle button above the Contents section can be used to switch off or on the GtoImmuPdb view. There is no difference in data content between the original GtoPdb view and the new GtoImmuPdb view; the latter merely highlights content of specific immunopharmacological relevance.

7. Click on Immunopharmacology Comments in the Contents section. The page focus jumps down to the relevant table.

The general curator comments here indicate that XCR1 is one of >20 chemokine receptors. It states that it is expressed in a subset of dendritic cells and a reference is provided - click on the reference number to view details. The comments may also contain hyperlinks to relevant ligands. The Immuno Cell Type Associations section below this contains more specific comments about XCR1 expression in dendritic cells, with additional references.

BASIC PROTOCOL 6

BROWSE TO EXPLORE DATA FOR IMMUNOLOGICAL PROCESSES

The GtoImmuPdb project has enriched GtoPdb with data on biological processes of relevance to immunology. Targets have been mapped to several top-level immunological process categories, which provides another way for organising and searching immunopharmacological data. This is complementary to the cell type data described in Basic Protocol 5. Process data are accessible from the GtoImmuPdb portal home page, via the Processes/Pathways panel. They are also available from the Processes menu of the main navigation bar on every page. This protocol introduces how to browse and view immunological process data and find pharmacological targets.

Necessary resources
**Software**

An up-to-date Web browser, such as Firefox or Google Chrome, with JavaScript enabled.

**Exploring process data in the Guide to IMMUNOPHARMACOLOGY**

1. Begin at the GtoImmuPdb portal home page (http://www.guidetoimmunopharmacology.org) (Figure 19).

2. Under the section Processes/Pathways, select “View immuno processes home page”. Alternatively, click on the Processes menu at the top, or directly navigate to http://www.guidetoimmunopharmacology.org/GRAC/ProcessesForward.

   This page shows more details about the main process categories used in GtoImmuPdb (Figure 26). Quick links at the top move to the start of each process category section on the page. For each category, there is a brief description, alongside lists of the Gene Ontology (GO) terms included under that category. The process data in GtoImmuPdb are underpinned by the use of GO terms (www.geneontology.org) (Ashburner et al., 2000; The Gene Ontology Consortium, 2017), comprising a set of biological process terms and identifiers that are mapped to top-level immunological categories. Using GO provides a controlled vocabulary and external identifiers, which supports interoperability of the data. Similar to cell types, ontology mapping enables high-resolution annotation of targets to process terms, and facilitates searches by parent or child terms to bring back all the data.

3. Click on the quick link for “Chemotaxis & migration” (Figure 26).

   The summary for Chemotaxis & migration explains that this category includes (but is not limited to) processes that are children of four different GO parent terms: leukocyte cell-cell adhesion (GO:0007159), chemokine production (GO:0032602), leukocyte migration (GO:0050900), and response to chemokine (GO:1990868). Clicking on a GO term links to AmiGO 2 (Carbon et al., 2009), a GO browser, which provides more details for the term including its definition, context and annotations.

4. Click on the heading for the “Chemotaxis & migration” section. Alternatively, select this name from the "Processes" menu of the navigation bar.

   The page now shows all targets associated to the immunological process category Chemotaxis & migration (Figure 27 panel A). The table is split into sections for each target class; quick links are available in the top right corner. The pull-down menu in the top left can be used to switch process category. The first column shows the target name (linking to its detailed view page) and family name (in parentheses), the second shows curator comments about the association, the third gives GO terms associated to the target, and the final column shows general immunopharmacology comments about the target.

5. Click on "Catalytic Receptors" to move quickly to that section, and scroll down to find the Integrins (this is the family name shown in parenthesis after the target name).
There are three Integrins associated with the Chemotaxis & migration process (Figure 27 panel B). The third column lists GO terms that are annotated against these targets. Looking at the data for "integrin, alpha 4 subunit", it shows six GO terms with their IDs and 3-letter evidence codes. The evidence code is provided by the GO annotation source (the GO annotations used in GtoImmuPdb have been downloaded from UniProt (The UniProt Consortium, 2017), and it indicates the level of support for the annotation. More information on evidence codes is available from http://www.geneontology.org/page/guide-go-evidence-codes. The evidence code helps to determine the veracity of the annotation.

6. Click on the 'integrin, alpha 4 subunit' name in the first column to view its detailed view page. The detailed page provides further information on the target, with data of immunological relevance highlighted in blue.

7. Click on Immunopharmacology Comments under the Contents section to jump to this section on the page.

The comments indicate that this target promotes cell migration by interacting with a signalling adaptor protein. They also indicate that alpha 4-containing integrins are the molecular targets of the antibody natalizumab, a migration-inhibiting drug. Below this section is a detailed table of “Immuno Process Associations” listing all the processes that this target is associated with.

BASIC PROTOCOL 7

NAVIGATING RECIPROCAL LINKS TO PUBMED AND PUBCHEM

The provenance of entity records, out-links and contextual comments from curators and NC-IUPHAR members is supported by document citations. While the database does have small numbers of book references, patent numbers and URL-only links, by far the majority have PubMed Identifiers (PMIDs). GtoPdb release 2017.5 contains ~31,000 unique PMIDs, of which ~22,000 are associated with targets and families. The rest are associated with ligand data, of which ~6,000 include the quantitative binding interactions. These PMIDs form part of the substance records we submit to PubChem (SIDs). This means that users can not only query into GtoPdb with a ligand or a PMID but can also navigate PubChem-to-PubMed reciprocal linking as is shown below (Kim et al., 2015).

Necessary resources

Software

An up-to-date Web browser, such as Firefox or Google Chrome, with JavaScript enabled.

Wordpad, or equivalent for temporary paste-across storage, particularly for query strings


Identifiers
PubMed ID (PMID), PubChem substance identifier (SID) and PubChem compound identifier (CID)

Protocol steps—Step annotations

**Locating a reference and navigating to PubMed**


2. From the GtoPdb home page click on Advanced Search in the main navigation bar or go directly to the target search tools at http://www.guidetopharmacology.org/GRAC/searchPage.jsp. In the fourth box, choose PubMed Id from the field options, enter 7651358 into the search box and hit “Search the database”.

3. The result of the search is shown in Figure 28. The matches include 13 ligands and three targets. Each of these entries has a reference section with multiple PMIDs (e.g. http://www.guidetopharmacology.org/GRAC/LigandDisplayForward?tab=refs&ligandId=499), but each reference includes a link-out (with a blue URL) to PubMed. Locate the example reference in one of the linked entries or move directly to the PubMed/NCBI Entrez system at https://www.ncbi.nlm.nih.gov/pubmed/7651358. Many linking options are displayed but Figure 29 shows the three that link to GtoPdb targets directly (plus a family link), and our associated PubChem SID links (shown on the right).

The alternative literature-related query options in GtoPdb can appear complex, particularly since targets and ligands can be citation-linked not only to the same but multiple papers. However, the circularity (document < > target < > ligand) is actually useful in that users can navigate to what they need by different routes.

**Navigating from PubMed to PubChem and back to GtoPdb**

4. The right-hand (RH) facets of Figure 29 show the linking options within the NCBI Entrez system. Note, these are distinct from the external LinkOut resources that are “pushed” to PubMed (as we have done for the target links). To square the circle by navigating back to GtoPdb try the SID links first (by definition, some of these will be from GtoPdb because we have curated the PMID link). The top section of the SID link results is shown in Figure 30. These are clearly labelled as IUPHAR/BPS Guide to PHARMACOLOGY, and each entry will link back to GtoPdb. Alternatively, the corresponding CIDs can be accessed by selected this option from the PubMed RH facet (Figure 29).

Link navigation within the NCBI systems is powerful but dauntingly complex in the many presented options, most of which have elements of circularity. Notwithstanding, the examples above show users can make useful queries from both the GtoPdb and NCBI sides and join “across” both resources. The entry points can be via
PMIDs or CIDs and SIDs. Once users have crossed over to PubChem, all the options described above in Support Protocol 2 (navigating reciprocal links to PubChem) become available to extend the user's exploration of ligand relationships (Kim et al., 2015).

COMMENTARY
Background Information

The IUPHAR/BPS Guide to PHARMACOLOGY is intended to be reference resource for researchers, teachers and those studying pharmacology. It aims to provide information at two levels of detail: the first gives an overview of the nomenclature, key properties and recommended experimental ligands for a wide-range of targets. These can be found on the family summary pages (Figure 1 and described in Basic Protocol 1 steps 3 and 4). The summary information is published every two years as the Concise Guide to PHARMACOLOGY (CGTP, http://www.guidetopharmacology.org/concise/) (Alexander et al., 2015; Alexander et al., 2017; Alexander et al., 2013). The information is checked and updated by the CGTP collaborators (members of the NC-IUPHAR subcommittees or independent experts where no subcommittee exists) prior to publication. The CGTP provides a citable, point-in-time snapshot of the data, extracted into publication format intended as a quick desktop reference and introduction to pharmacology. It is available online as a HTML and PDF document, but the paper copies remain very popular, as it has the advantage that it can be used at the laboratory bench and kept on a shelf for rapid access. This is particularly useful in locations where there is unreliable internet access, such as in some developing countries.

The second, more comprehensive, level of detail provides additional database pages for a selection of established targets, which covers a wide range of information on function, tissue distribution, and assays, through to clinically-used drugs and variants affecting pharmacology or causing disease. Data on these pages are also reviewed biennially by the wider NC-IUPHAR subcommittees. These pages include longer lists of ligands, which may be endogenous ligands, approved drugs, or ligands currently recommended for experimental use, while some others may have been used in the past and are maintained for a historical perspective. All data tables have the option for longer free text comments, where more details about the experiments and any caveats are described. Detailed data for the GPCRs, ion channels and NHRs have been collected and revised over a 14-year period, originating in the IUPHAR-DB (which was a precursor for GtoPdb) (Harmar et al., 2009; Sharman et al., 2013; Sharman et al., 2011). Hence, these pages contain in-depth information on a range of topics, while the newer targets tend to focus on key features related to pharmacology, function and clinical relevance. Some of the data types originally included, such as variants and tissue expression, are now readily available from other resources, so we do not tend to curate these ourselves. Much of this information is available via the database link-outs provided on target pages. An example is the Human Protein Atlas (Uhlen et al., 2015) which maps protein expression in cells, tissues, and organs, and now includes the Pathology Atlas (Uhlen et al., 2017).

The IUPHAR Guide to IMMUNOPHARMACOLOGY has been created as an extension to the existing GtoPdb site, with a new portal and data types, which draws together information on immunological cell types, processes and diseases, with data on pharmacological targets and their ligands. In doing so, we have deliberately not created an entirely new resource, but have extended the existing database schema and website for GtoPdb. The immunopharmacology data (and all other data types) are available both from the original GtoPdb site and the GtoImmuPdb site. The latter alerts users to specific immunopharmacology data and relevant targets and ligands through
content highlighting. The GtoImmuPdb portal provides new paths to access the data via the cell type, process and disease pages described in Protocols 4, 5 and 6. These have been developed in collaboration with immunologists to capture the relevant categories and data fields. The site is currently under development, and is undergoing rounds of user testing, which will identify any areas for improvement. Therefore, although the main data types and navigation systems are in place, the interface and data content may change slightly from the description presented here.

The latest content statistics and a detailed breakdown of each target/ligand category are described in our latest *Nucleic Acids Research* Database Issue article (Harding et al., 2018). This two-yearly update includes a detailed explanation of new features and content added to the database since the previous descriptions were published in 2014 and 2016 (Pawson et al., 2014; Southan et al., 2016). We also regularly publish technical reports, announcements of database updates and invited commentaries on hot topics in pharmacology on our blog (https://blog.guidetopharmacology.org/). These and other news items, such as IUPHAR review articles (see publication list at http://www.guidetopharmacology.org/nciupharPublications.jsp), are promoted via our social media accounts on Twitter, Facebook and LinkedIn. We also maintain sets of public slides on SlideShare (https://www.slideshare.net/GuidetoPHARM), which include generic slides for use in teaching with GtoPdb.

### Critical Parameters and Troubleshooting

The GtoPdb and GtoImmuPdb websites generally do not require direct user input of parameters, apart from web service access described in Alternate Protocol 3. The search terms will accept any alphanumeric character, and recognise many special characters such as Greek letters commonly used in target names (e.g. β1-adrenoceptor). Keyword searches use the full text search algorithm provided as part of PostgreSQL. This involves converting user-provided search terms to their lexemes and performing searches against pre-calculated lexemes for the text fields in the database, with a ranking applied based on the strength of the match to a particular field. In practice, this means that searches for “melatonin receptors” will return results matching to “melatonin receptors”, “melatonin receptor”, or any other possible variant. However, this type of search does not always work well for some types of names used in biology and chemistry, particularly those including punctuation. This commonly affects compounds, since research code names are often recorded in literature or online databases in different formats, some containing punctuation with a hyphen or comma.

For example, ligand ID 2592 in GtoPdb is L735821, however this compound has also been referred to in the literature as L-735,821. Searches of GtoPdb using the latter term unfortunately fail because of the punctuation, which causes the name to be split into separate parts. Therefore, if users encounter a ligand name with punctuation, and initially get no results from a name search, they are encouraged to try a search without punctuation, or to try a wildcard search for part of the name. Wildcard searches are available via the Advanced Target Search or Ligand Search pages. Queries should be entered in the relevant box with a percentage (%) sign used to represent the missing characters. For example, a ligand name wildcard search for “%821” returns 4 results, one of which is L735821.

The GtoPdb web services do require URLs and parameters to be constructed correctly. Misspelling a parameter name in an otherwise correct web service call would lead to the default behaviour of all results being returned. If an incorrect target or ligand ID is provided, this would usually lead to a 404 Not Found error response code. The same is true for trying to call a web service that does not exist, for example by misspelling the name. If a correct call is made to web service that does not return any results because that particular database table is empty (for example, asking for a list of variants for target ID 400 using http://www.guidetopharmacology.org/services/targets/400/variants) a 204 No Content
response code would be returned. The list of response codes and their meanings are described in the documentation at http://www.guidetopharmacology.org/webServices.jsp#responses.

Whilst older entries are generally not removed from the database as they serve as a useful historical record of pharmacological development, occasionally data are removed if they are found to be incorrect or new information has come to light which reduces the value of their inclusion. Sometimes links to the deleted data will still exist on external sites, so users may encounter a “not found” error for deprecated entries. Occasionally, new information results in reclassification of targets and/or reorganisation of the family hierarchy. The database tracks changes in target and family nomenclature and deprecated URLs will redirect to the new pages where possible.

GtoPdb is manually compiled by a team of curators with advice from a network of over 500 expert scientists organised into 96 subcommittees overseen by NC-IUPHAR. The majority of references have thus been recommended by experts with the data extracted and normalised into a database format by the curators. In some newer areas, curators select the initial ligand reports while the subcommittees are being set up. Despite this level of double checking, as with any database errors can occur, and we encourage users to contact us by email if they are unsure about any data. The best way to do this is via the email address enquiries@guidetopharmacology.org. If the query is complex, we often refer this on to the relevant subcommittee chair for comment.

We provide Support Protocols 1 and 2 and Basic Protocol 7 as examples of the powerful data mining possibilities that arise from combining GtoPdb data with other sources, via the UniProtKB, PubChem and PubMed interfaces. However, these sites have their own complexities and we cannot go into detail about their usage here. Each resource provides their own detailed help pages and support desk, and users with specific queries about these examples are welcome to contact us for help.

ACKNOWLEDGEMENT

The authors wish to thank all members of NC-IUPHAR for their continued support http://www.guidetopharmacology.org/nciuphar.jsp#membership. We also thank the global network of NC-IUPHAR subcommittees and all the CGTP contributors (a full list of subcommittee members and contributors can be viewed at http://www.guidetopharmacology.org/GRAC/ContributorListForward). We thank T. Wigglesworth for administrative support. This work was supported by IUPHAR and the BPS. In addition to our primary funding, we are grateful for sponsorship from the American Society for Pharmacology and Experimental Therapeutics (ASPET), Laboratoires Servier and The University of Edinburgh.

LITERATURE CITED


John Wiley & Sons


FIGURE LEGENDS

Figure 1. Screenshot of the 5-HT$_3$ receptor family summary page with the summary section for the target 5-HT$_3$AB open. The summary includes NC-IUPHAR’s recommended nomenclature, alternative names by which the target has been known, component subunits (clicking on these takes users to the subunit section, which includes gene and protein identifiers), key selective ligands listed by action with affinities and references, and the functional characteristics of this ion channel.

Figure 2. Screenshot of the ligand tables on the 5-HT$_3$AB detailed target page. The Channel Blockers table contains human and mouse data and species-specific tables are available by clicking on the link below it. Tables showing the chemical structures of the compounds are available by clicking on the “View all chemical structures” links.

Figure 3. The target search tools, showing a text search for “5-HT3” matching to the names of 6 targets and 1 family.

Figure 4. Making an initial select for GtoPdb cross-references using the UniProtKB Advanced search menu bar.

Figure 5. The results of changing the column display settings for 1483 human protein GtoPdb cross-references in UniProtKB.

Figure 6. A sort on a post-download Excel sheet for GtoPdb cross-references and proteases in UniProt, ranked by number of PDB entries.

Figure 7. Seven UniProt entries matching the InterPro domain “IPR001461 Aspartic_peptidase_A1” with GtoPdb cross-references, and showing links to the rare disease resource, Orphanet.

Figure 8. Screenshot of the ligand summary page for erlotinib. Nomenclature information includes commonly-used synonyms and identifiers from the literature, the IUPAC name, English trade names, and the International Nonproprietary Name (INN). The ligand page also includes a depiction of the chemical structure, and under the Structure tab various structural identifiers such as the SMILES string, International Chemical Identifier (InChI), and InChIKey. Clicking on the structure image loads the structure into the chemical drawing tool which can be used for structure-based searches of the database.

Figure 9. Data on quantitative ligand binding at targets are displayed under the Biological activity tab of the erlotinib ligand page. The primary target is indicated as epidermal growth factor receptor, with an additional target of the OATP2B1 transporter. This drug is also known to be a substrate of the CYP3A4 enzyme.

Figure 10. A SMILES structure has been loaded into the ligand structure drawing tool. The tool can perform searches by substructure, high or moderate similarity, exact match, or SMARTS pattern matching.
Figure 11. Connection between GtoPdb and PubChem. The GtoPdb ligand entry is 8699. The arrow indicates the click crossing to PubChem CID 51352361.

Figure 12. Navigation from CID 51352361 “Contents” (left hand facets) across to “Similar Compounds”, highlighted in yellow.

Figure 13. Showing the result of the “Similar Compounds” link from CID 51352361. Additional options and filters are shown in the right-hand facets.

Figure 14. Clustering of all 684 “Similar Compounds”. One CID is highlighted to show the substructure common to all members of the set. Note, the button “Similarity Data” produces a CSV matrix file of all the Tanimoto scores.

Figure 15. The PubChem structure search page set up with CID 51352361. The Tanimoto options for similarity searching are shown with 96% set as the threshold.

Figure 16. 3D clustering of the 74 compounds 97% similar to CID 51352361. The options were set to display the conformer renderings and large-size 2D/3D juxtaposed displays for the mouseovers.

Figure 17. Showing the results of editing CID 51352361 and searching GtoPdb by substructure using the SMILES string FC2=C(C)C=C(N(C(=O)C1=NC=CC=C1)[H])C=C2. Six of the seven matches are shown.

Figure 18. A section of the JSON output from a web service call to generate a list of GPCRs. Each target is shown with summary information on its nomenclature, family IDs, component subunit IDs if it is a complex (e.g. the GABA_B receptor), and complex IDs if it is a component subunit (e.g. GABA_B1 and GABA_B2). Names include HTML codes to represent formatting such as subscript (the sub tag), and special characters such as the Greek alpha.

Figure 19. The Guide to IMMUNOPHARMACOLOGY portal, showing the entry points to processes, cell types, targets and ligands.

Figure 20. The GtoImmuPdb disease-target associations page. Users can switch to view disease-ligand associations using the tab at the top.

Figure 21. GtoImmuPdb expanded view of the disease-ligand associations for Chronic lymphocytic leukemia. Approved drugs are indicated with the medicine bottle icon, and the main drug approval bodies (usually FDA and/or EMA, although others are given where the drug is only approved outside the US and Europe) and first approval dates are given.

Figure 22. A summary of clinical data for the drug idelalisib, from its GtoPdb ligand page.

Figure 23. A section of the search results for the disease term “Alzheimer’s”. The results are ordered by best match, with the top result (target GPR3) having 6 matches across 5 database fields.
Figure 24. The GtoImmuPdb immunological cell types home page, showing the list of top-level cell type categories.

Figure 25. (A) The list of GtoImmuPdb target associations to dendritic cells. Users can switch cell type using the drop down menu at the top left. (B) A close-up of the association between the target XCR1 and dendritic cells. The second column shows curator comments, the third column gives the Cell Ontology terms, and the final column shows an extract of the general immunopharmacology comments on XCR1.

Figure 26. The immunological processes home page, showing the definitions and mapped GO terms for the first few top-level process categories. Quick links to the categories are provided at the top.

Figure 27. (A) The list of target associations for chemotaxis and migration. The third column shows the GO immunological process terms mapped to the targets named in the left column. The final column summarises the general immunopharmacology comments for the target. (B) The process GO terms and comments for the integrins.

Figure 28. The literature search box from the GtoPdb Advanced search page is shown at the top. The initial section of the result list is in the centre, and the reference as shown in each of the linked entries is at the bottom. The PMID in blue is live-linked.

Figure 29. Link displays from https://www.ncbi.nlm.nih.gov/pubmed/7651358. The GtoPdb targets are linked via Molecular Biology Databases under LinkOut resources. PubChem links for GtoPdb ligands can be found on the right under Related information.

Figure 30. Some of the PubChem SID links from PMID 7651358 (as shown in Figure 29). All four of these structures link directly back to GtoPdb.
Figure 1. Screenshot of the 5-HT3 receptor family summary page with the summary section for the target 5-HT3AB open. The summary includes NC-IUPHAR’s recommended nomenclature, alternative names by which the target has been known, component subunits (clicking on these takes users to the subunit section, which includes gene and protein identifiers), key selective ligands listed by action with affinities and references, and the functional characteristics of this ion channel.

250x268mm (96 x 96 DPI)
Figure 2. Screenshot of the ligand tables on the 5-HT3AB detailed target page. The Channel Blockers table contains human and mouse data and species-specific tables are available by clicking on the link below it. Tables showing the chemical structures of the compounds are available by clicking on the "View all chemical structures" links.

247x406mm (96 x 96 DPI)
Figure 3. The target search tools, showing a text search for "5-HT3" matching to the names of 6 targets and 1 family.

251x271mm (96 x 96 DPI)
Figure 4. Making an initial select for GtoPdb cross-references using the UniProtKB Advanced search menu bar.

257x160mm (96 x 96 DPI)
Figure 5. The results of changing the column display settings for 1483 human protein GtoPdb cross-references in UniProtKB.

402x197mm (96 x 96 DPI)
Figure 6. A sort on a post-download Excel sheet for GtoPdb cross-references and proteases in UniProt, ranked by number of PDB entries.

284x94mm (96 x 96 DPI)
Figure 7. Seven UniProt entries matching the InterPro domain "IPR001461 Aspartic_peptidase_A1" with GtoPdb cross-references, and showing links to the rare disease resource, Orphanet.

John Wiley & Sons
Figure 8. Screenshot of the ligand summary page for erlotinib. Nomenclature information includes commonly-used synonyms and identifiers from the literature, the IUPAC name, English trade names, and the International Nonproprietary Name (INN). The ligand page also includes a depiction of the chemical structure, and under the Structure tab various structural identifiers such as the SMILES string, International Chemical Identifier (InChI), and InChIKey. Clicking on the structure image loads the structure into the chemical drawing tool which can be used for structure-based searches of the database.

249x247mm (96 x 96 DPI)
Figure 9. Data on quantitative ligand binding at targets are displayed under the Biological activity tab of the erlotinib ligand page. The primary target is indicated as epidermal growth factor receptor, with an additional target of the OATP2B1 transporter. This drug is also known to be a substrate of the CYP3A4 enzyme.
Figure 10. A SMILES structure has been loaded into the ligand structure drawing tool. The tool can perform searches by substructure, high or moderate similarity, exact match, or SMARTS pattern matching.

222x286mm (96 x 96 DPI)
Figure 11. Connection between GtoPdb and PubChem. The GtoPdb ligand entry is 8699. The arrow indicates the click crossing to PubChem CID 51352361.

259x205mm (96 x 96 DPI)
Figure 12. Navigation from CID 51352361 “Contents” (left hand facets) across to “Similar Compounds”, highlighted in yellow.

203x198mm (96 x 96 DPI)
Figure 13. Showing the result of the “Similar Compounds” link from CID 51352361. Additional options and filters are shown in the right-hand facets.
Figure 14. Clustering of all 684 “Similar Compounds”. One CID is highlighted to show the substructure common to all members of the set. Note, the button “Similarity Data” produces a CSV matrix file of all the Tanimoto scores.

186x218mm (96 x 96 DPI)
Figure 15. The PubChem structure search page set up with CID 51352361. The Tanimoto options for similarity searching are shown with 96% set as the threshold.

213x186mm (96 x 96 DPI)
Figure 16. 3D clustering of the 74 compounds 97% similar to CID 51352361. The options were set to display the conformer renderings and large-size 2D/3D juxtaposed displays for the mouseovers.

297x160mm (96 x 96 DPI)
Figure 17. Showing the results of editing CID 51352361 and searching GtoPdb by substructure using the SMILES string \texttt{FC2=C(C)C=C(N(C(\equiv O)C1=NC=CC=C1)[H])C=C2}. Six of the seven matches are shown.

184x231mm (96 x 96 DPI)
Figure 18. A section of the JSON output from a web service call to generate a list of GPCRs. Each target is shown with summary information on its nomenclature, family IDs, component subunit IDs if it is a complex (e.g. the GABAB receptor), and complex IDs if it is a component subunit (e.g. GABAB1 and GABAB2). Names include HTML codes to represent formatting such as subscript (the sub tag), and special characters such as the Greek alpha.

93x199mm (96 x 96 DPI)
Figure 19. The Guide to IMMUNOPHARMACOLOGY portal, showing the entry points to processes, cell types, targets and ligands.

258x269mm (96 x 96 DPI)
Figure 20. The GtoImmuPdb disease-target associations page. Users can switch to view disease-ligand associations using the tab at the top.

251x212mm (96 x 96 DPI)
<table>
<thead>
<tr>
<th>Ligand</th>
<th>Disease Association Comments</th>
<th>Clinical Use</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>elotuzumab</td>
<td>Approved drug for some CLL patients.</td>
<td>Approved to treat chronic lymphocytic leukemia that is refractory to fludarabine</td>
<td>[FDA (2009), EMA (2010) (Drugs.com)]</td>
</tr>
<tr>
<td>veuzumab</td>
<td>FDA and EMA orphan drug for CLL.</td>
<td>The EMA and FDA have granted veuzumab orphan designation for the treatment of the rare disease chronic lymphocytic leukemia (CLL).</td>
<td></td>
</tr>
<tr>
<td>idelalisib</td>
<td>Approved in combination with rituximab for CLL. Also used to treat relapsed small lymphocytic leukemia (SLL).</td>
<td>Idelalisib is a treatment for difficult to treat leukemia and lymphomas. The potential of this drug was highlighted by the early termination of clinical trials.</td>
<td>[FDA &amp; EMA (2014) (Drugs.com)]</td>
</tr>
<tr>
<td>rituximab</td>
<td>An anti-CD20 therapy approved for CLL and non-Hodgkin lymphomas.</td>
<td>Used to treat CD20-positive non-Hodgkin lymphoma, chronic lymphocytic leukemia, and rheumatoid arthritis. Rituximab is also used to treat follicular lymphoma and mantle cell lymphoma (MCL), a rare and aggressive type of leukemia, especially in patients with IPI ≥ 2.</td>
<td>[FDA (1997), EMA (1998) (Drugs.com)]</td>
</tr>
<tr>
<td>brentuximab</td>
<td>Approved drug for CLL.</td>
<td>Ibrutinib is approved to treat patients with mantle cell lymphoma (MCL), a rare and aggressive type of leukemia, especially in patients with IPI ≥ 2.</td>
<td>[FDA and EMA (2013) (Drugs.com)]</td>
</tr>
<tr>
<td>obinutuzumab</td>
<td>Approved for use in combination with chemotherapy in patients who have received no prior therapy.</td>
<td>Initially approved to treat chronic lymphocytic leukemia (CLL) in combination with chemotherapy in patients who have received previous treatment.</td>
<td>[FDA (2013), EMA (2014) (Drugs.com)]</td>
</tr>
</tbody>
</table>

299x218mm (96 x 96 DPI)
Figure 22. A summary of clinical data for the drug idelalisib, from its GtoPdb ligand page.

<table>
<thead>
<tr>
<th>Summary</th>
<th>Biological activity</th>
<th>Clinical data</th>
<th>References</th>
<th>Structure</th>
<th>Similar ligands</th>
<th>Immunopharmacology</th>
</tr>
</thead>
</table>

**Summary of Clinical Use**

Idelalisib is a treatment for difficult-to-treat leukemia and lymphomas. The potential of this drug was highlighted by the early termination of a phase 3 clinical trial, so that all participants could be given the drug [1]. Idelalisib was approved in July 2014, for patients with relapsed chronic lymphocytic leukemia (CLL), in combination with rituximab. The FDA also granted the drug accelerated approval for the treatment of patients with relapsed follicular B-cell non-Hodgkin lymphoma (FL) and relapsed small lymphocytic lymphoma (SLL), particularly indicated for patients who have received at least two prior systemic therapies.

To view a full list of trials registered with ClinicalTrials.gov assessing idelalisib, click here.

**Mechanism Of Action and Pharmacodynamic Effects**

PI3K signaling is critical for the activation, proliferation, survival and trafficking of B lymphocytes and is hyperactive in many B-cell malignancies. Idelalisib inhibits this activity, ultimately leading to death of cancer cells.

**External Links**

For extended ABBRE data see the following:

- Electronic Medicines Compendium (EMC)
- Drugs.com
- European Medicines Agency (EMA)
Figure 23. A section of the search results for the disease term “Alzheimer’s”. The results are ordered by best match, with the top result (target GPR3) having 6 matches across 5 database fields.

250x238mm (96 x 96 DPI)
Figure 24. The GtoImmuPdb immunological cell types home page, showing the list of top-level cell type categories.

314x212mm (96 x 96 DPI)
Figure 25. (A) The list of GtoImmuPdb target associations to dendritic cells. Users can switch cell type using the drop down menu at the top left. (B) A close-up of the association between the target XCR1 and dendritic cells. The second column shows curator comments, the third column gives the Cell Ontology terms, and the final column shows an extract of the general immunopharmacology comments on XCR1.

624x468mm (96 x 96 DPI)
Figure 26. The immunological processes home page, showing the definitions and mapped GO terms for the first few top-level process categories. Quick links to the categories are provided at the top.
Figure 27. (A) The list of target associations for chemotaxis and migration. The third column shows the GO immunological process terms mapped to the targets named in the left column. The final column summarises the general immunopharmacology comments for the target. (B) The process GO terms and comments for the integrins.
Figure 28. The literature search box from the GtoPdb Advanced search page is shown at the top. The initial section of the result list is in the centre, and the reference as shown in each of the linked entries is at the bottom. The PMID in blue is live-linked.

227x203mm (96 x 96 DPI)
Figure 29. Link displays from https://www.ncbi.nlm.nih.gov/pubmed/7651358. The GtoPdb targets are linked via Molecular Biology Databases under LinkOut resources. PubChem links for GtoPdb ligands can be found on the right under Related information.
Figure 30. Some of the PubChem SID links from PMID 7651358 (as shown in Figure 29). All four of these structures link directly back to GtoPdb.