NCBI Minute: APIs @NCBI

NCBI Webinars

- Closed captioning: www.captionedtext.com and enter 3341486
- The **recording**, will be on our YouTube channel in the Webinars playlist: <youtube>/user/NCBINLM/playlists
- Use the Questions Pod to ask questions when you think of them. Don't wait until the end.
- Answers to all questions will be available after the webinar and linked to our Webinars page: <ncbi>/home/coursesandwebinars.shtml
- Materials / Q&A: https://go.usa.gov/xRQDf

<ncbi> = www.ncbi.nlm.nih.gov
<ftp> = ftp.ncbi.nlm.nih.gov
<youtube> = www.youtube.com

9/06/2017

NCBI Minute:
APIs @NCBI

NCBI Webinars

This is an introduction to NCBI APIs for people new to APIs and for everyone, including the already API-savvy, to give you an idea of the diversity and breadth of APIs available at NCBI….
….and where to look for more information.

9/06/2017

This webinar is just an introduction….

NCBI Webinars

If you have specific questions about how to use a particular API or problems with a script you have already developed, *please send a message to the HelpDesk (info@ncbi.nlm.nih.gov)* and we'll work with you directly to help you.

After this webinar, we'll be sending you a **Survey** to ask you for some feedback:

1) Which specific API topics would you like to learn more about for future webinars?
2) What functionality or access to specific types of data would you like to have added to our existing repertoire of APIs?
3) What additional suggestions do you have to help us improve our APIs?

We'd love to hear your thoughts, so that we can improve our services for you!

3

Using APIs (Application Program Interfaces)

NCBI Webinars

What is an API? What can it do for you?

- URLs are web addresses and are composed of two main components: a protocol identifier (ex: HTTP/HTTPS or FTP/SFTP) and a resource name.
  https://www.ncbi.nlm.nih.gov or https://blast.ncbi.nlm.nih.gov
- APIs are URLs (with specific syntax) designed for developers to use within their scripts/applications/programs to interact with servers across the internet.
- NCBI's APIs have been developed for 3rd-party developers to access, process and retrieve data from our backend servers and are now also used within our own internal systems, for example: PubChem, EndNote, BioGPS, Molecules, and more…

Think of APIs as an action-based sentence - with specific grammatical parts.
I'll show you how this works in just a minute…..

4

Today…

- The NCBI site-wide API: Entrez Programming Utilities (EUtilities or EUtils)
- PubMed Central (PMC)'s ID Converter
- BLAST URL API
- PubChem's PUG REST
- Entrez Direct (Edirect)
- *A couple of others and where to find more information…*

NCBI Webinars

5

The NCBI site-wide API: Entrez Programming Utilities (EUtilities or EUtils)

A set of structured URL-based calls to access the NCBI Entrez search system and databases.

These store information covering a diverse array of data including medical and scientific literature, nucleotide and protein sequences, gene records, gene expression, genotype and biological assay studies with data, chemical information, and three-dimensional molecular structures.

Examples of things you can do:

- Search PubMed with a term and retrieve PMIDs.
- Search MedGen with a SnoMedCT ID, then link over to GTR, and retrieve their Summary info.
- Post a list of Gene IDs and fetch their full records.
- Post a PubChem Compound CID and retrieve the SDF-formatted record (to show the 3D coordinates in the Molecules App).

Where to learn more?
https://www.ncbi.nlm.nih.gov/books/NBK25501/

YouTube Videos:
2012 (4 min): https://www.youtube.com/watch?v=BCG-M5k-gvE
2014 (1 hour): https://www.youtube.com/watch?v=iCFVVexp30o

EUtils Base URL:

`https://eutils.ncbi.nlm.nih.gov/entrez/eutils/`

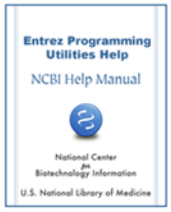
NCBI Webinars

6

Entrez Programming Utilities Help

https://www.ncbi.nlm.nih.gov/books/NBK25501/

NCBI Webinars



Entrez Programming Utilities Help

Bethesda (MD): National Center for Biotechnology Information (US); 2010-.

Copyright and Permissions

Search this book

Introduction to the E-utilities

- YouTube E-utilities Introduction
- Please see the Release Notes for details and changes.

The **Entrez Programming Utilities** (**E-utilities**) are a set of eight server-side programs that provide a stable interface into the **Entrez** query and database system at the National Center for Biotechnology Information (NCBI). The **E-utilities** use a fixed URL syntax that translates a standard set of input parameters into the values necessary for various NCBI software components to search for and retrieve the requested data. The **E-utilities** are therefore the structured interface to the **Entrez** system, which currently includes 38 databases covering a variety of biomedical data, including nucleotide and protein sequences, gene records, three-dimensional molecular structures, and the biomedical literature.

Contents

- E-utilities Quick Start
  Created: December 12, 2008; Last Update: August 9, 2013.
- A General Introduction to the E-utilities
  - Introduction
  - Usage Guidelines and Requirements
  - The Nine E-utilities in Brief
  - Understanding the E-utilities Within Entrez
  - Combining E-utility Calls to Create Entrez Applications
  - Demonstration Programs
  - For More Information
- Sample Applications of the E-utilities
  Created: April 24, 2009; Last Update: January 26, 2010.
- The E-utilities In-Depth: Parameters, Syntax and More
  Created: May 29, 2009; Last Update: January 9, 2017.
- Entrez Direct: E-utilities on the UNIX Command Line
  Created: April 23, 2013; Last Update: August 28, 2017.

7

NOUNS: Where to operate?
(which database/kind of data?)

NCBI Webinars

Literature	
Books	books and reports
MeSH	ontology used for PubMed indexing
NLM Catalog	books, journals and more in the NLM Collections
PubMed	scientific & medical abstracts/citations
PubMed Central	full-text journal articles
Health	
ClinVar	
dbGaP	
GTR	
MedGen	
OMIM	
PubMed Health	
Genomes	
Assembly	
BioProject	
BioSample	
Clone	
dbVar	
Epigenomics	epigenomic studies and display tools
Genome	genome sequencing projects by organism
GSS	genome survey sequences
Nucleotide	DNA and RNA sequences
Probe	sequence-based probes and primers
SNP	short genetic variations
SRA	high-throughput DNA and RNA sequence read archive
Taxonomy	taxonomic classification and nomenclature catalog

Genes	
EST	expressed sequence tag sequences
Gene	collected information about gene loci
GEO DataSets	functional genomics studies
GEO Profiles	gene expression and molecular abundance profiles
HomoloGene	homologous gene sets for selected organisms
	sequence sets from phylogenetic and population studies
	clusters of expressed transcripts
	conserved protein domains
	protein sequences
	sequence similarity-based protein clusters
	experimentally determined biomolecular structures
	molecular pathways with links to genes, proteins and chemicals
	bioactivity screening studies
	chemical information with structures, information and links
PubChem Substance	deposited substance and chemical information

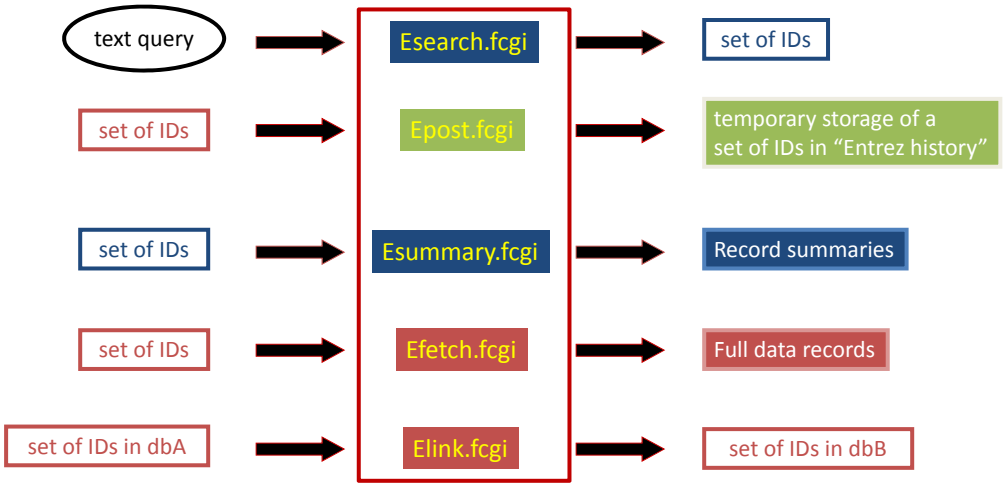
>40 Databases, >1 billion records:

- Biomedical literature (PubMed, PMC…)
- Health-related (MedGen, GTR, ClinVar…)
- Sequences (Nucleotide, SRA, Protein…)
- Genomic data (Genome, dbGaP, dbSNP…)
- Gene information (Gene, HomoloGene, BioSystems…)
- Gene expression (GEO, UniGene…)
- 3D Biomolecule Structure/function (Structure, CDD…)
- Chemicals & Biological Assays (PubChem)

8

VERBS *(do what?)* AND NOUN *(with what?)*

`https://eutils.ncbi.nlm.nih.gov/entrez/eutils/[ ].fcgi?`



NCBI Webinars

EUtils API Examples

Getting ESummary information for Gene with a Gene ID of 3269 (Human Histamine Receptor H1)

https://eutils.ncbi.nlm.nih.gov/entrez/eutils/esummary.fcgi?db=gene&id=3269

BASE URL CALL Perform this **action** on this **database** with this **term**

*(return the information in this **format**)*

NCBI Webinars

Getting ESummary information for Gene with a Gene ID of 3269 (Human Histamine Receptor H1)

https://eutils.ncbi.nlm.nih.gov/entrez/eutils/esummary.fcgi?db=gene&id=3269

NCBI Webinars


```
This XML file does not appear to have any style information associated with it. The document tree is shown below.

▼<eSummaryResult>
 ▼<DocumentSummarySet status="OK">
   <DbBuild>Build170904-2345m.1</DbBuild>
  ▼<DocumentSummary uid="3269">
    <Name>HRH1</Name>
    <Description>histamine receptor H1</Description>
    <Status>0</Status>
    <CurrentID>0</CurrentID>
    <Chromosome>3</Chromosome>
    <GeneticSource>genomic</GeneticSource>
    <MapLocation>3p25.3</MapLocation>
    <OtherAliases>H1-R, H1R, HH1R, hisH1</OtherAliases>
   ▼<OtherDesignations>
      histamine H1 receptor|histamine receptor, subclass H1
    </OtherDesignations>
    <NomenclatureSymbol>HRH1</NomenclatureSymbol>
    <NomenclatureName>histamine receptor H1</NomenclatureName>
    <NomenclatureStatus>Official</NomenclatureStatus>
   ▼<Mim>
      <int>600167</int>
    </Mim>
   ▼<GenomicInfo>
    ▼<GenomicInfoType>
       <ChrLoc>3</ChrLoc>
       <ChrAccVer>NC_000003.12</ChrAccVer>
       <ChrStart>11137092</ChrStart>
       <ChrStop>11263252</ChrStop>
       <ExonCount>8</ExonCount>
     </GenomicInfoType>
   </GenomicInfo>
   <GeneWeight>11579</GeneWeight>
   ▼<Summary>
      Histamine is a ubiquitous messenger molecule released from mast cells, enterochromaffin-like cells, and neurons. Its various actions are
      mediated by histamine receptors H1, H2, H3 and H4. The protein encoded by this gene is an integral membrane protein and belongs to the G
      protein-coupled receptor superfamily. It mediates the contraction of smooth muscles, the increase in capillary permeability due to
      contraction of terminal venules, the release of catecholamine from adrenal medulla, and neurotransmission in the central nervous system.
```


EUtils API Examples

Getting ESummary information for Gene with a Gene ID of 3269 (Human Histamine Receptor H1)

https://eutils.ncbi.nlm.nih.gov/entrez/eutils/esummary.fcgi?db=gene&id=3269

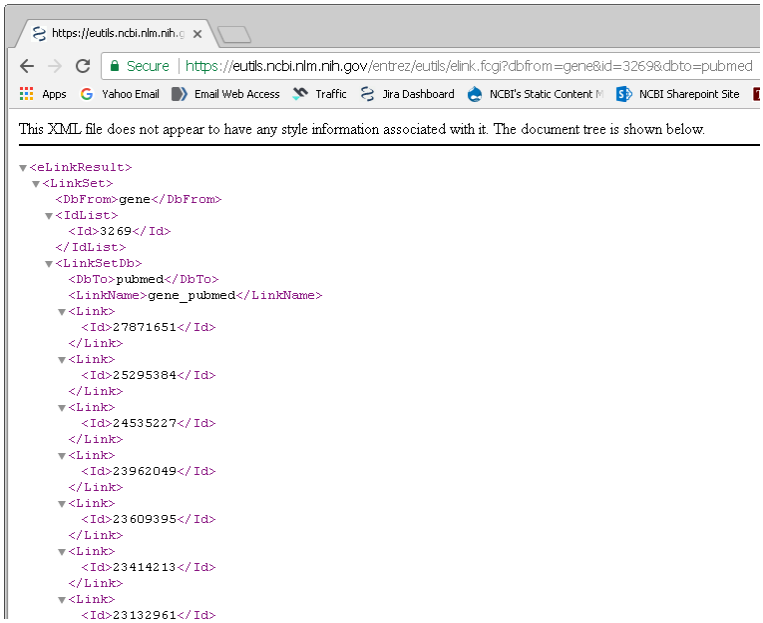
NCBI Webinars

Linking to Find PubMed Record IDs (PMIDs) for Gene ID 3269

/elink.fcgi?dbfrom=gene&id=3269&dbto=pubmed

Linking to Find PubMed Record IDs (PMIDs) for Gene ID 3269

/elink.fcgi?dbfrom=gene&id=3269&dbto=pubmed



NCBI Webinars

13

EUtils API Examples

Getting ESummary information for Gene with a Gene ID of 3269 (Human Histamine Receptor H1)

https://eutils.ncbi.nlm.nih.gov/entrez/eutils/esummary.fcgi?db=gene&id=3269

Linking to Find PubMed Record IDs (PMIDs) for Gene ID 3269

/elink.fcgi?dbfrom=gene&id=3269&dbto=pubmed

Getting ESummary or Efetch information for PMID 24236486 *(which was linked to Gene ID 3269)*

/esummary.fcgi?db=pubmed&id=24236486 (retype= xml is the default)

/efetch.fcgi?db=pubmed&id=24236486&rettype=medline

/efetch.fcgi?db=pubmed&id=24236486&rettype=xml

NCBI Webinars

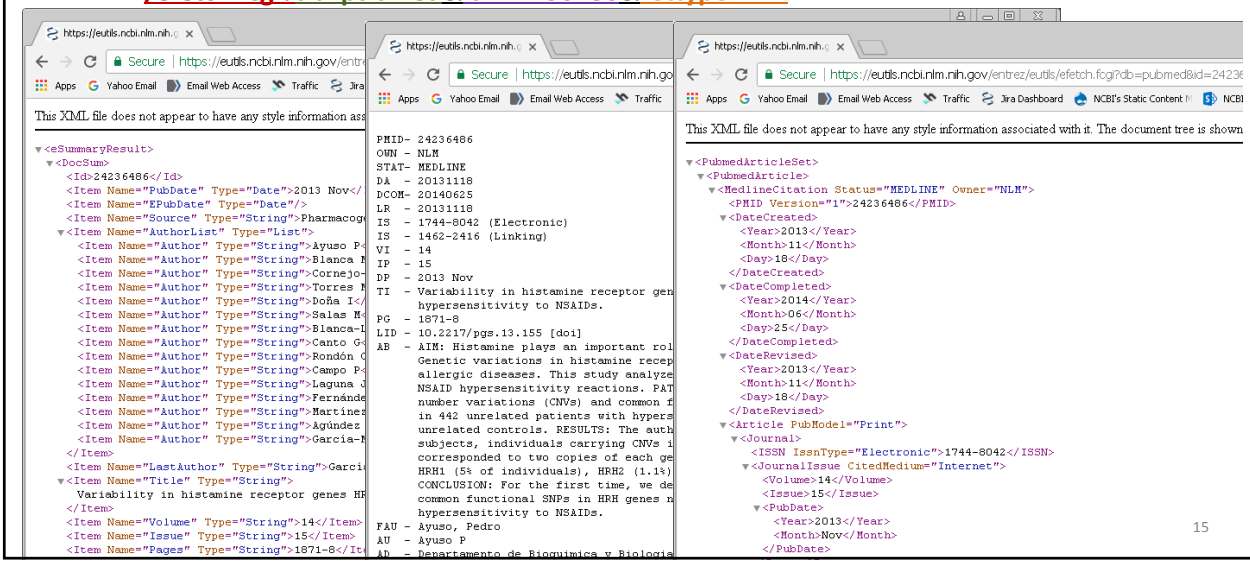
14

Getting ESummary or Efetch information for PMID 24236486 *(which was linked to Gene ID 3269)*

/esummary.fcgi?db=pubmed&id=24236486 (retype= xml is the default)

/efetch.fcgi?db=pubmed&id=24236486&rettype=medline

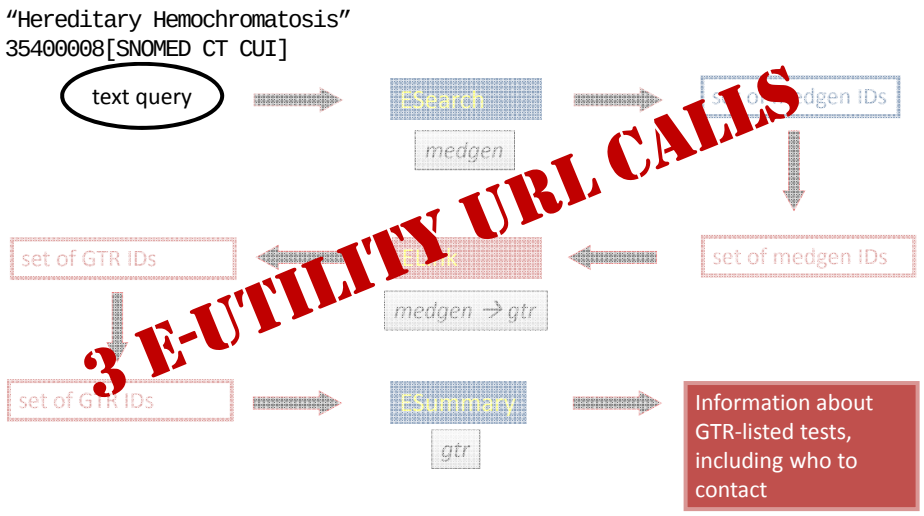
/efetch.fcgi?db=pubmed&id=24236486&rettype=xml



NCBI Webinars

15

What they are REALLY useful for: EUtil Pipelines



NCBI Webinars

16

EUtils API USAGE GUIDELINES

NCBI Webinars

Limit URL requests to less than 3 per second.

Limit large jobs to either weekends or weekdays - between 9PM and 5AM EST

Failure to comply with this policy may result in an IP address being blocked from accessing NCBI.

https://www.ncbi.nlm.nih.gov/books/NBK25497/#chapter2.Usage_Guidelines_and_Requiremen

Currently, you can register EUtil API calls that you are using by sending an Email to eutilities@ncbi.nlm.nih.gov with the name of your tool and an Email address. If there are any problems with your script, we will contact you using the registered Email before we block you.

Anticipated in the Spring of 2018, we will be initiating an **EUtilities API Registration system** to enable you to sign up to get an API Key to incorporate in your URL calls. This will enable us to better manage our server access so that we can accommodate the large and still increasing amounts of scripted access.

BONUS: Including the API Key will enable YOU to increase your EUtil URL call rates over the current standards!

17

A PMC Tool API: ID Converter

NCBI Webinars

This service enables you to find the relevant pairs of these IDs for record listed in PubMed Central (PMC):

- **PMCID** (PMC ID)
- **PMID** (PubMed ID)
- **Manuscript ID** (available *only* for articles that came in through a manuscript submission system, e.g., NIHMS, Europe PMC, PMC Canada)
- **DOI** (Digital Object Identifier, *if the article has one*)

CAVEAT: Works for all records that are listed in PMC…..
not all in PubMed, though.

Examples of things you can do:

- Finding the PMCID or PMID for a NIHMS ID
- Correlating PMIDs with PMCIDs
  (free full text articles!)
- Finding a DOI for a PMCID or PMID.

Where to learn more?

https://www.ncbi.nlm.nih.gov/pmc/tools/id-converter-api/

PMC ID Converter Base-URL:

`https://www.ncbi.nlm.nih.gov/pmc/utils/idconv/v1.0/?`

18

PMC Developer Resources & the ID Converter API

https://www.ncbi.nlm.nih.gov/pmc/tools/developers/

NCBI Webinars

NCBI Resources How To Sign in to NCBI

PMC US National Library of Medicine National Institutes of Health PMC Advanced Journal list

About PMC For Publishers Related Resources

Developer Resources

We hope that developers make use of the APIs and tools that we have built

In this page, we list the APIs and public code repositories related to PMC. V
common developer questions, and provide instructions and examples for ad

PMC APIs

The use of our APIs is entirely free, and doesn't require an API key, but we

- Do not make concurrent requests, even at off-peak times; and
- Include two parameters that help to identify your service or applicatio
  - *tool* should be the name of the application, as a string value wit
  - *email* should be the e-mail address of the maintainer of the tool,

For the sake of readability, these parameters have been omitted from t
the example pages.

Here is a list of APIs that can be used to get data about PMC resources.

- NCBI E-utilities,
- File validation tools,
- OA web service,
- OAI-PMH,
- ID converter,
- FTP service.

PMC public code repositories

ID Converter API

The ID Converter API is the backend web service that is used by the PMCID - PMID - Manuscript ID - DOI Converter. This API allow you to programmatically convert between the various IDs used in our system.

The base URL of the service is https://www.ncbi.nlm.nih.gov/pmc/utils/idconv/v1.0/. This is shown as "*service-root*" in the examples below.

When using this API programmatically, we request that you use the *tool* and *email* parameters to identify the application making the request. *tool* should be the name of the application, as a string value with no internal spaces. *email* should be the e-mail address of the maintainer of the tool, and should be a valid e-mail address. For brevity, these have been left out of the displayed URLs in the examples below.

The *ids* parameter specifies the article IDs that you want to convert. By default, the service will determine the type of the IDs from their format. In any given service call, all of the IDs must be of the same type.

Here are examples, one for each type of ID that the service understands:

PMID: *service-root*?ids=23193287
PMCID: *service-root*?ids=PMC3531190
Manuscript ID: *service-root*?ids=NIHMS311352
DOI: *service-root*?ids=10.1093/nar/gks1195
Versioned identifier: *service-root*?ids=PMC2808187.1

Use commas to separate multiple IDs:

service-root?ids=PMC3531190,PMC3245039

This service allows for conversion of up to 200 IDs in a single request. If you have a larger number of IDs, split your list into smaller subsets. Another option for bulk-conversion of identifiers in the PMC system is to download the PMC-ids.csv.gz file, as described on the FTP Service page.

You can override ID type auto-detection behavior by using the *idtype* parameter. Valid values are "pmcid", "pmid", "mid", and "doi". For example, the following lets you specify PMCIDs as simple numeric values. If the *idtype* parameter is not used, PMCIDs must have the "PMC" prefix, in order to distinguish them from PMIDs.

service-root?ids=14900&idtype=pmcid

Examples

19

PMC ID Converter API Example

NCBI Webinars

Find the other IDs for the NIHMS Submission ID 311352

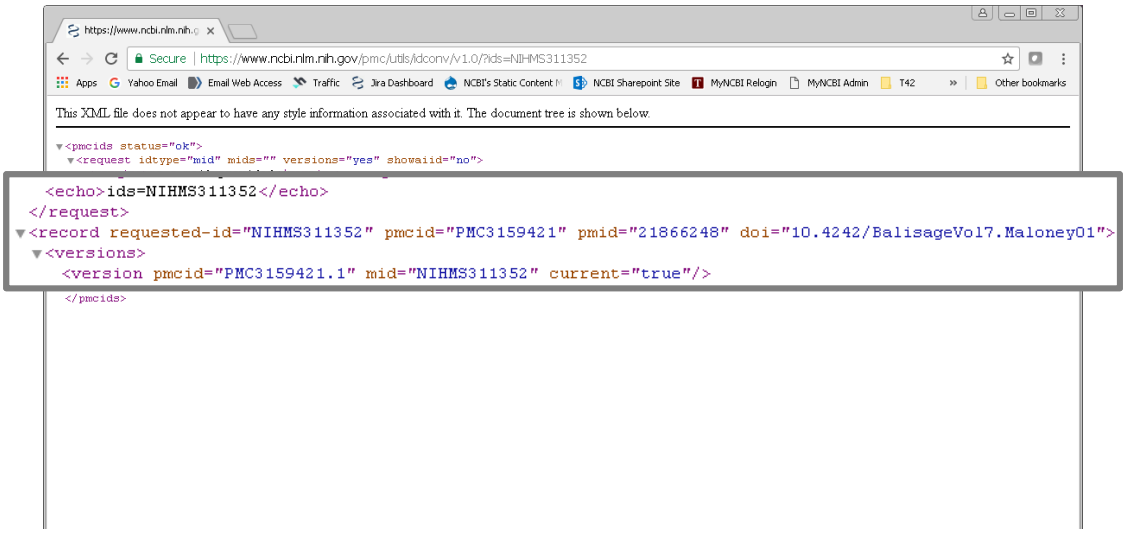
https://www.ncbi.nlm.nih.gov/pmc/utils/idconv/v1.0/?ids=NIHMS311352

BASE URL CALL Perform this **action** on this **database** with this **term**
(return the information in **this format**!)

20

Find the other IDs for the NIHMS Submission ID 311352

https://www.ncbi.nlm.nih.gov/pmc/utils/idconv/v1.0/?ids=NIHMS311352



NCBI Webinars

21

PMC ID Converter API Example

Find the other IDs for the NIHMS Submission ID 311352

https://www.ncbi.nlm.nih.gov/pmc/utils/idconv/v1.0/?ids=NIHMS311352

You can retrieve the results in several formats xml (default), html, json, or "csv".

https://www.ncbi.nlm.nih.gov/pmc/utils/idconv/v1.0/?ids=NIHMS311352&format=csv

BASE URL CALL Perform this **action** on this **database** with this **term**
(return the information in **this format**!)

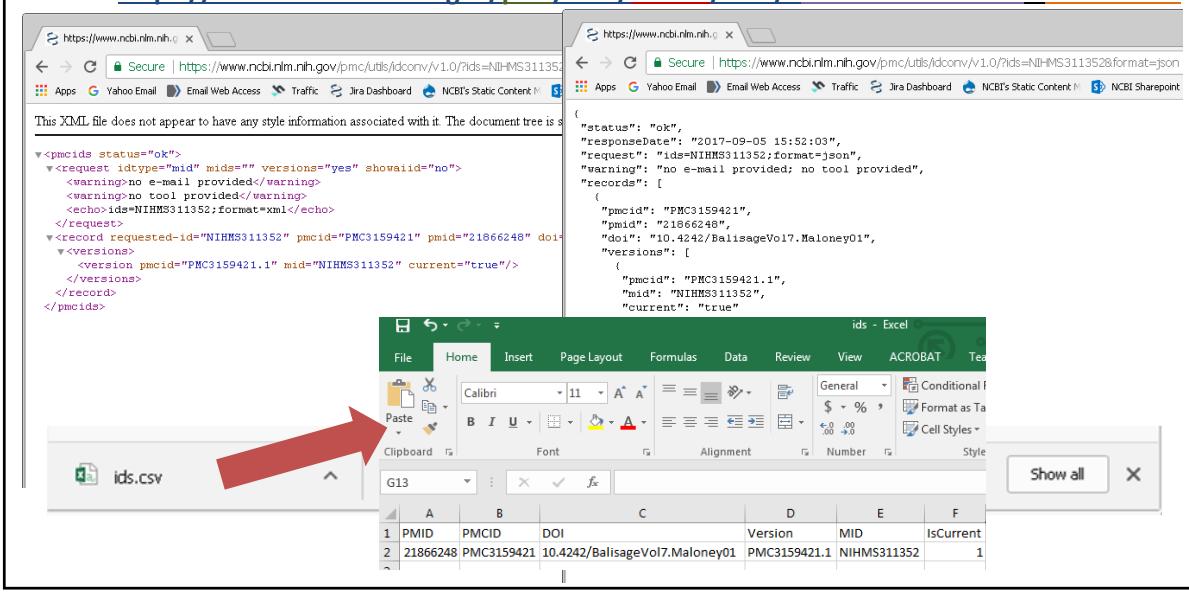
NCBI Webinars

22

You can retrieve the results in several formats xml (default), html, json, or "csv".

https://www.ncbi.nlm.nih.gov/pmc/utils/idconv/v1.0/?ids=NIHMS311352

https://www.ncbi.nlm.nih.gov/pmc/utils/idconv/v1.0/?ids=NIHMS311352&format=csv



NCBI Webinars

23

PMC ID Converter API USAGE GUIDELINES

Each lookup request can contain ≥ 1 identifiers,
but *do not mix the types of IDs in a single request*.

≤200 IDs can be searched in a single request.
For a larger number of IDs, split your list up!

(If you have a really large set of IDs you'd like to match, consider downloading this file: ftp://ftp.ncbi.nlm.nih.gov/pub/pmc/PMC-ids.csv.gz)

NCBI Webinars

24

BLAST's URL API

The NCBI-BLAST Common URL API allows you to run sequence similarity searches remotely on the NCBI web servers (Web BLAST).

This is a 3 step process:

- ***Submission*** of the job to the servers,
- ***Checking*** to see if the job is done,
- ***Retrieval*** of the results once the job is done.

A set of parameters is available to incorporate in the calls to tailor your BLAST search and specify the format for retrieval of the search results. These are described on the NCBI BLAST GitHub page.

Examples of things you can do:

- Search for an identical or near identical sequence to find out more about it.
- Search for similar sequences in other organisms to identify potential homologs.
- Search for where the sequence may be match a region on a genome or chromosome to enable annotation for where it is encoded.
- *and more…*

Where to learn more?

https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastDocs&DOC_TYPE=DeveloperInfo

https://ncbi.github.io/blast-cloud/dev/api.html

BLAST URL API Base:

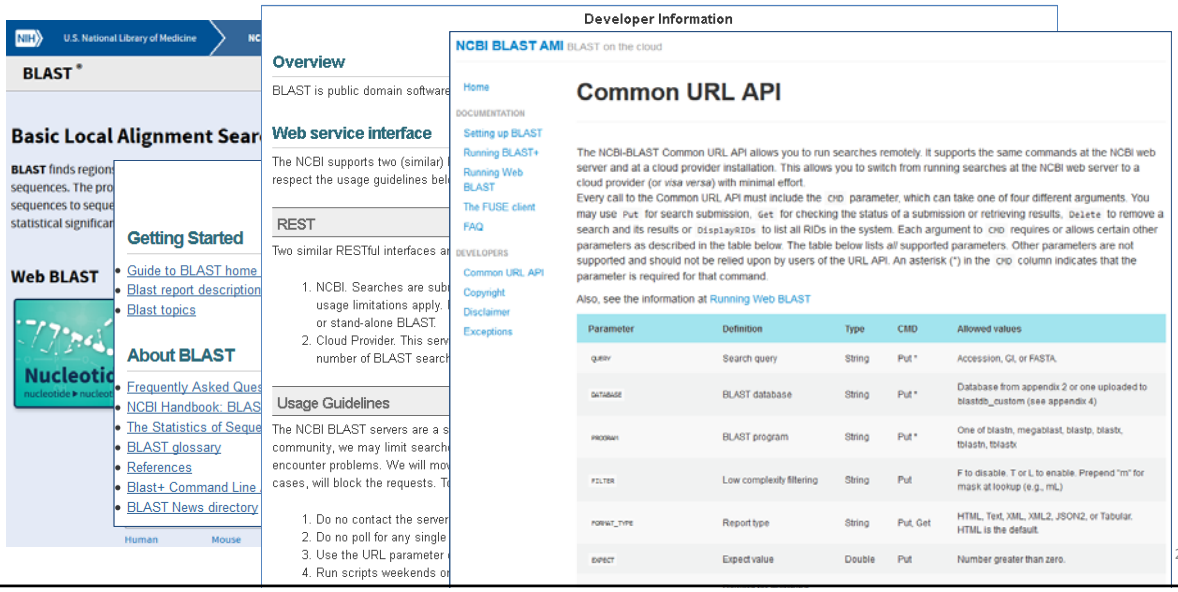
`https://blast.ncbi.nlm.nih.gov/Blast.cgi?`

NCBI Webinars

25

BLAST URL API

https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastDocs&DOC_TYPE=DeveloperInfo



NCBI Webinars

26

3-step set for BLAST URL API Calls

1) **Submit** your query sequences/IDs for comparison/calculation & get back an RID

2) **Check** on the progress of your BLAST search with the RID

3) When the Results are ready – **retrieve** them in your choice of format.

NCBI Webinars

27

3-step set for BLAST URL API Calls

1) Compare the sequence of human acetylcholinesterase with those of known structures in the PDB database.
   Submit an ID to a particular BLASTp calculation & get back an RID (Request ID).

https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Put&QUERY=P22303&PROGRAM=blastp&DATABASE=pdb

- ***Look for the RID value for use in step 2!***

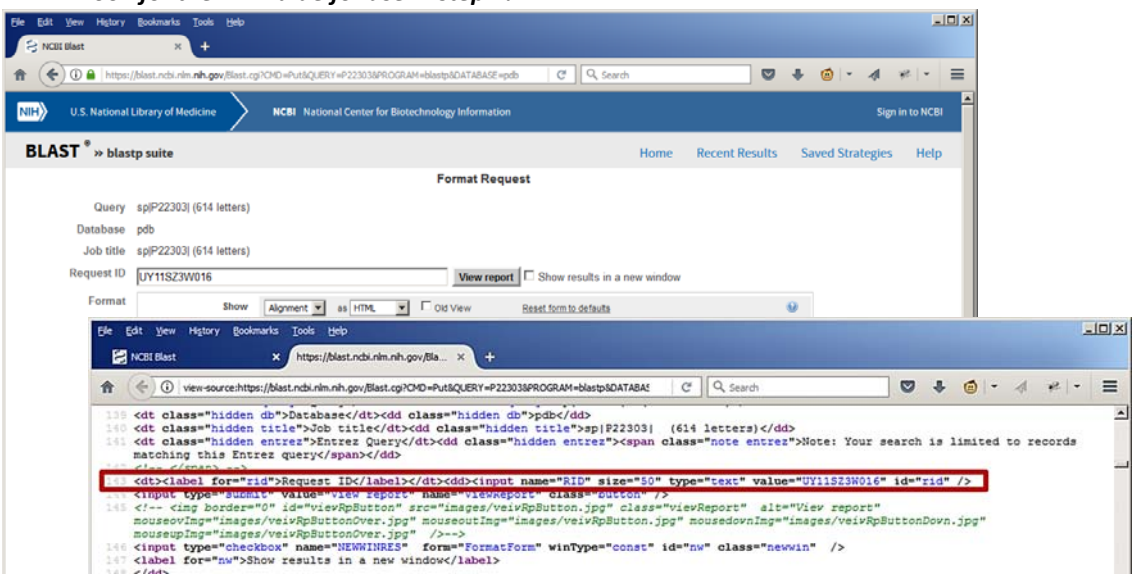
NCBI Webinars

28

1) Compare the sequence of human acetylcholinesterase with those of known structures in the PDB database.
Submit an ID to a particular BLASTp calculation & get back an RID (Request ID).

https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Put&QUERY=P22303&PROGRAM=blastp&DATABASE=pdb

- ***Look for the RID value for use in step 2!***



NCBI Webinars

29

3-step set for BLAST URL API Calls

1) Compare the sequence of human acetylcholinesterase with those of known structures in the PDB database.
Submit an ID to a particular BLASTp calculation & get back an RID (Request ID).

https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Put&QUERY=P22303&PROGRAM=blastp&DATABASE=pdb

- ***Look for the RID value for use in step 2!***

2) Check on the progress of your BLAST search with the RID

/Blast.cgi?CMD=Get&RID=UY11SZ3W016

- *Look for a "READY" status to indicate the BLAST search has completed and results are available for access.*

NCBI Webinars

30

3-step set for BLAST URL API Calls

NCBI Webinars

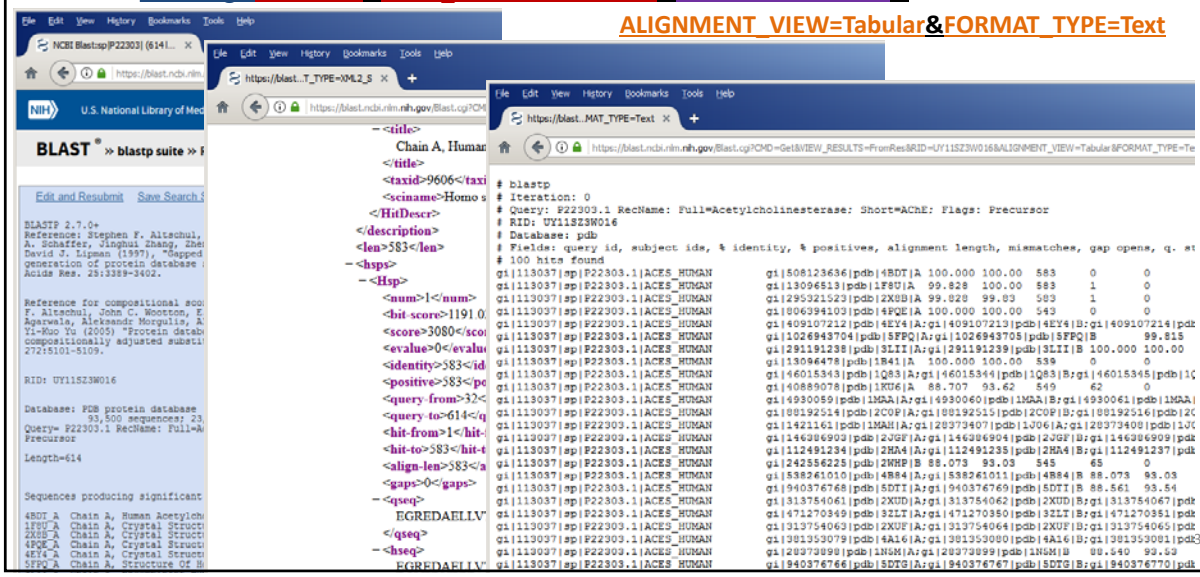
1) Compare the sequence of human acetylcholinesterase with those of known structures in the PDB database.
   Submit an ID to a particular BLASTp calculation & get back an RID (Request ID).
   https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Put&QUERY=P22303&PROGRAM=blastp&DATABASE=pdb
   - ***Look for the RID value for use in step 2!***

2) Check on the progress of your BLAST search with the RID
   /Blast.cgi?CMD=Get&RID=UY11SZ3W016
   - *Look for a "READY" status to indicate the BLAST search has completed and results are available for access*

3) When the Results are ready – **retrieve** them in your choice of format.
   /Blast.cgi?CMD=Get&VIEW_RESULTS=FromRes&RID=UY11SZ3W016&FORMAT_TYPE=TEXT
   /Blast.cgi?CMD=Get&VIEW_RESULTS=FromRes&RID=UY11SZ3W016&FORMAT_TYPE=XML2_S
   /Blast.cgi?CMD=Get&VIEW_RESULTS=FromRes&RID=UY11SZ3W016&
   ALIGNMENT_VIEW=Tabular&FORMAT_TYPE=Text

31

3) When the Results are ready – **retrieve** them in your choice of format.
   /Blast.cgi?CMD=Get&VIEW_RESULTS=FromRes&RID=UY11SZ3W016&FORMAT_TYPE=TEXT
   /Blast.cgi?CMD=Get&VIEW_RESULTS=FromRes&RID=UY11SZ3W016&FORMAT_TYPE=XML2_S
   /Blast.cgi?CMD=Get&VIEW_RESULTS=FromRes&RID=UY11SZ3W016&
   ALIGNMENT_VIEW=Tabular&FORMAT_TYPE=Text



NCBI Webinars

BLAST URL API USAGE GUIDELINES

NCBI Webinars

1) **Submit** your query sequences/IDs for comparison/calculation & get back an RID

Please submit at least 3 seconds apart.

2) **Check** on the progress of your BLAST search with the RID

You need to let the BLAST Search do its work!
So delay the initial check for at least 1 minute after submission
And at least 30 seconds for each subsequent check.

3) When the Results are ready – **retrieve** them in your choice of format.

Finally, please limit your Web BLAST searches to less than 100 per day.

You need to do more than that? ***Consider using the "BLAST in the Cloud".***

33

PubChem's Specialty API: PUG REST

NCBI Webinars

A set of structured URL-based calls to access and retrieve information from the PubChem set of databases (Compound, Substance & BioAssay).

There are many ways to access PubChem's servers, PUG REST has two facets…

PUG VIEW REST – enables the downloading of PubChem records sections from the servers

PUG REST – enabling search and retrieval of *pieces* of information from backend servers

Examples of things you can do:

- Search for chemical synonyms (other things they are called: nicknames, IUPAC, etc.)
- Find some chemical properties for a particular chemical.
- Get whole sections of PubChem Compound records (such as Pharmacology or Toxicology information).
- See the 2D structure for a chemical.

Where to learn more?

https://pubchem.ncbi.nlm.nih.gov/pug_rest/PUG_REST.html

PubChem PUG REST Base URLs:

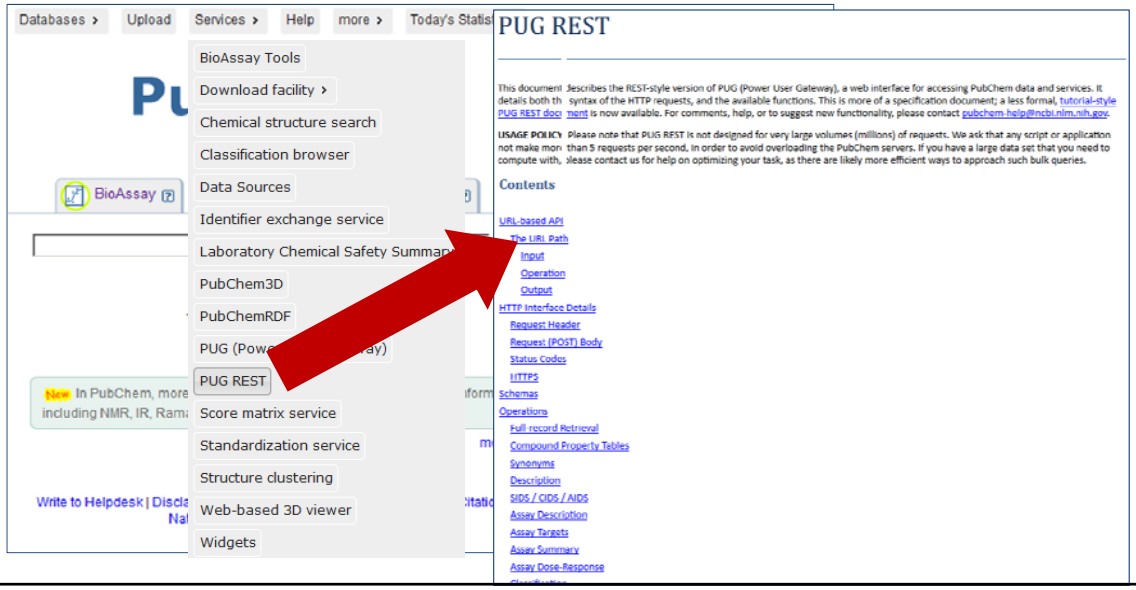
`https://pubchem.ncbi.nlm.nih.gov/rest/pug_view/data/`

`https://pubchem.ncbi.nlm.nih.gov/rest/pug/`

34

PubChem's PUG REST

https://pubchem.ncbi.nlm.nih.gov/pug_rest/PUG_REST.html



NCBI Webinars

35

PUG REST Search/Data Retrieval Calls

PUG_VIEW – Using a CID, you can retrieve ***whole sections*** of a PubChem Compound record.

https://pubchem.ncbi.nlm.nih.gov/rest/pug_view/data/compound/3100/JSON?heading=Synonyms

/compound/3100/JSON?heading=MeSH%20Pharmacological%20Classification

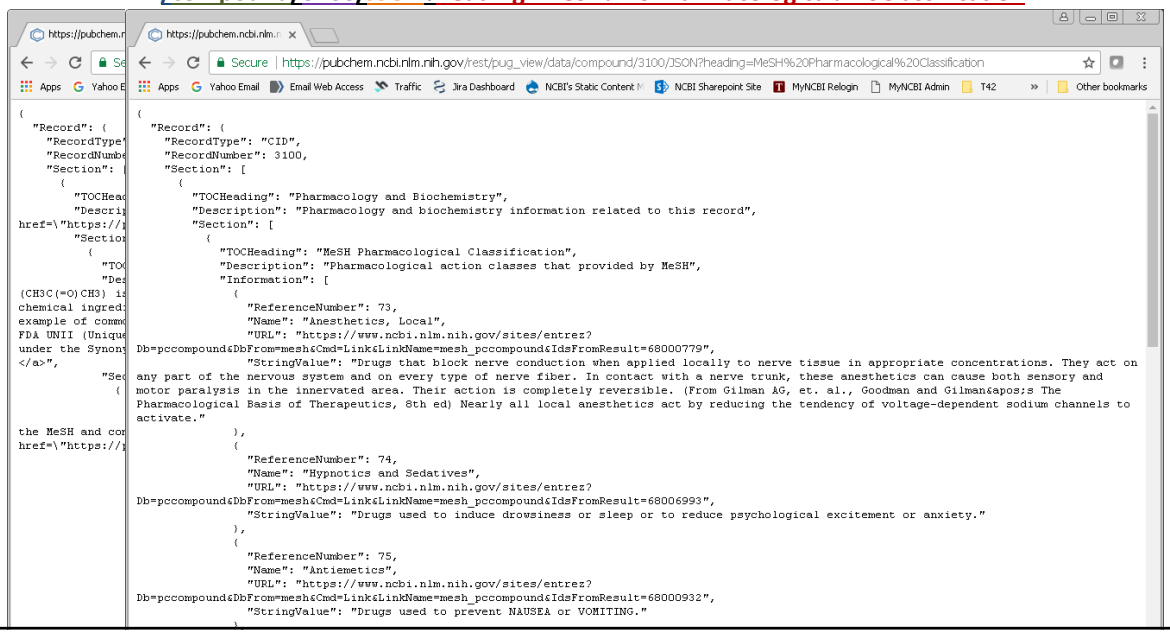
NCBI Webinars

36

PUG_VIEW – Using a CID, you can retrieve ***whole sections*** of a PubChem Compound record.

https://pubchem.ncbi.nlm.nih.gov/rest/pug_view/data/compound/3100/JSON?heading=Synonyms

/compound/3100/JSON?heading=MeSH%20Pharmacological%20Classification



NCBI Webinars

37

PUG REST Search/Data Retrieval Calls

PUG_VIEW – Using a CID, you can retrieve ***whole sections*** of a PubChem Compound record.

https://pubchem.ncbi.nlm.nih.gov/rest/pug_view/data/compound/3100/JSON?heading=Synonyms

/compound/3100/JSON?heading=MeSH%20Pharmacological%20Classification

CAVEAT: This method only works with CIDs (similar to the EUtil "Post" function),
so how do I find the CIDs?....(you *can* use the EUtil ESearch or PUG REST)

PUG - Searching with a "name" to retrieving CID(s).

https://pubchem.ncbi.nlm.nih.gov/rest/pug/compound/name/benadryl/cids/TXT

Searching with a "name" to retrieve a chemical property or even several!

/compound/name/benadryl/property/MolecularFormula/TXT

/compound/name/benadryl/property/MolecularFormula,MolecularWeight,InChIKey/XML

Searching with a "name" to retrieving a picture *(which happens to be a 2D structure!)*

/compound/name/benadryl/PNG

NCBI Webinars

38

PUG - Searching with a "name" to retrieving CID(s).

https://pubchem.ncbi.nlm.nih.gov/rest/pug/compound/name/benadryl/cids/TXT

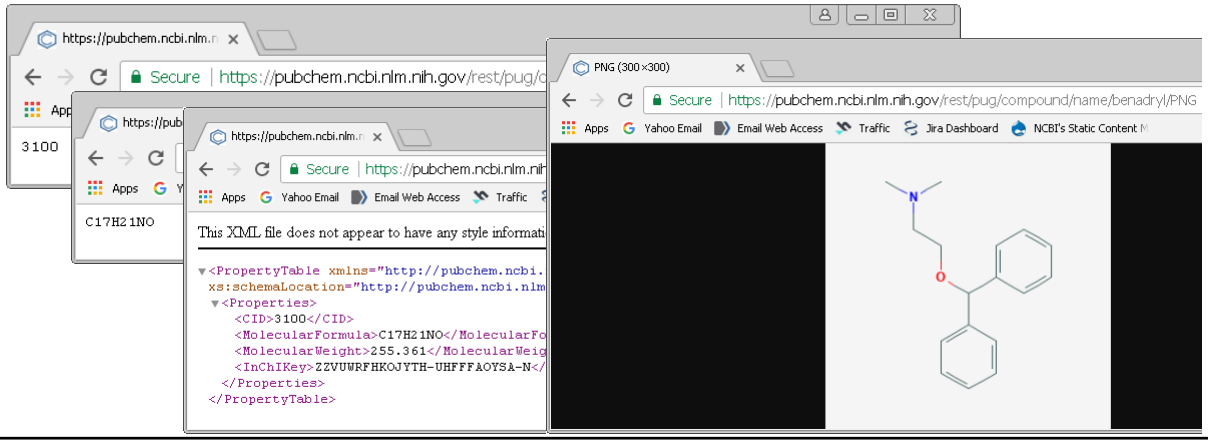
Searching with a "name" to retrieve a chemical property or even several!

/compound/name/benadryl/property/MolecularFormula/TXT

/compound/name/benadryl/property/MolecularFormula,MolecularWeight,InChIKey/XML

Searching with a "name" to retrieving a picture *(which happens to be a 2D structure!)*

/compound/name/benadryl/PNG



NCBI Webinars

PubChem's PUG REST USAGE GUIDELINES

≤5 requests per second

if you want a large amount of information, we have download files including specialty "Extras" ones:
ftp://ftp.ncbi.nlm.nih.gov/pubchem/Compound/

NCBI Webinars

40

Entrez Direct (EDirect) – an alternative to Eutils

for those who need the data, not the URL….

NCBI Webinars

Entrez Direct (EDirect) provides UNIX command-line access to the data in NCBI's suite of interconnected databases. Multi-step pipelines can be quickly concatenated linking record searching, retrieval, data parsing, and results formatting.

INSTALLATION OF A FILE CONTAINING ASUITE OF APPLICATIONS IS REQUIRED

UNIX and Macintosh computers with Perl installed can run the installer and quickly start using EDirect.

For Windows/PCs, you'll need to first install the Cygwin UNIX-emulation environment and a couple of additional applications, and then the installation file.

Examples of things you can do:

- Use line commands to automatically generate EUtils APIs, then parse and reformat the data!
- Download only parts of records
- Extract and concatenate data in specific fields
- Create a custom data table
- and lots more….

Where to learn more?

https://www.ncbi.nlm.nih.gov/books/NBK179288/

Edirect Cookbook:

https://github.com/NCBI-Hackathons/EDirectCookbook

Training Programs:

Webinar & Courses (NLM, NCBI – 4 hour at NIH 9/18)

4-Part 10-15min YouTube Videos (2015):
https://www.youtube.com/watch?v=BLnYW33Mtb0

41

EDirect Example

NCBI Webinars

Create a table for genes located between position 44.9M & 45.9M on human chromosome 19.

(With this information it'll be easy to download the sequences!)

```
$esearch -db gene -query "human[ORGN] AND 19[CHR] AND 44900000:45900000[CHRPOS] AND has_refseqgene[PROP]" \
> -sort Chromosome | \
> efetch -format docsum | \
> xtract -head ' Gene ID\tSymbol\tAccession\tChr. Start\tChr. Stop' \
> -pattern DocumentSummary -element Id Name -block GenomicInfoType \
> -element ChrAccVer ChrStart ChrStop
```

Gene ID	Symbol	Accession	Chr. Start	Chr. Stop
10452	TOMM40	NC_000019.10	44891219	44903688
348	APOE	NC_000019.10	44905748	44909394
341	APOC1	NC_000019.10	44914246	44919348
346	APOC4	NC_000019.10	44942237	44945495
100533990	APOC4-APOC2	NC_000019.10	44942237	44949564
344	APOC2	NC_000019.10	44945981	44949564
79090	TRAPPC6A	NC_000019.10	45178301	45162927
388552	BLOC1S3	NC_000019.10	45178744	45202714
90332	EXOC3L2	NC_000019.10	45234210	45212620
100420902	LOC100420902	NC_000019.10	45234808	45234194
2068	ERCC2	NC_000019.10	45370646	45349836
10849	CD3EAP	NC_000019.10	45406128	45410765

42

EDirect Chapter *in Entrez Programming Utilities Help*

https://www.ncbi.nlm.nih.gov/books/NBK179288/

NCBI Webinars

EDirect Cookbook on GitHub: https://github.com/NCBI-Hackathons/EDirectCookbook

43

Two examples of the many others available…

NCBI Webinars

Variation Services (Sequence Position Deletion Insertion, SPDI)

https://api.ncbi.nlm.nih.gov/variation/v0/

- Converts genetic variation data formats (HGVS notation, VCF files)
- Searches with your own variant data to match them up with existing variants and retrieves their RefSNP information.

Genome Remapping Service (ReMap)

https://www.ncbi.nlm.nih.gov/genome/tools/remap/docs/api

- Maps genomic annotation data from one coordinate system to another (a.k.a. 'liftover').

  Three variations exist:
  - Assembly-Assembly ReMap - from one genomic assembly to another
  - Clinical ReMap - from an assembly sequence to a RefSeqGene sequence
  - and Alt Loci ReMap - for the mapping of features between the Primary assembly and the Assembly Patches and Alternate Loci available for GRC genomic assemblies

ReMap provides a Perl script to download for use of the integrated API calls.

44

Argh!!! This is so much!
Where can I find the information?

NCBI Webinars

Wrap Up

As you may imagine, with all of the many resources developed at NCBI, there are a lot of diverse APIs that are available. Other NLM groups have APIs too:

https://wwwcf.nlm.nih.gov/nlm_eresources/eresources/search_database.cfm

At NCBI, we are in the process of cataloging all of the NCBI-specific APIs and will be adding them to the Develop page shortly.

Check in on the NCBI Develop site - YOUR NCBI SCRIPTING HOME!

https://www.ncbi.nlm.nih.gov/home/develop/

Webinars & Courses are available to help! (ex: EDirect course at NIH - Sept. 18th)

Hackathons - help scripters use APIs to create specific useful tools!

(ex: upcoming Hackathon in Pittsburgh - Sept. 25-27)

Keep an eye on the NCBI Lean site for more information

https://www.ncbi.nlm.nih.gov/home/learn/

46

NCBI Webinars

- Direct URLs & Places to Learn More -

NCBI Webinars

Entrez Programming Utilities (EUtilities, EUtils)

- https://www.ncbi.nlm.nih.gov/books/NBK25501/

PMC ID Converter

- https://www.ncbi.nlm.nih.gov/pmc/tools/id-converter-api/

BLAST URL API

- https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastDocs&DOC_TYPE=DeveloperInfo

PubChem's PUG REST

- https://pubchem.ncbi.nlm.nih.gov/pug_rest/PUG_REST.html

Entrez Direct (EDirect)

- https://www.ncbi.nlm.nih.gov/books/NBK179288/

NCBI Insights Blog: ncbiinsights.ncbi.nlm.nih.gov

YouTube
User: NCBINLM

For help with NCBI resources:
info@ncbi.nlm.nih.gov

Questions?

47