

Bioinformatics

Lecture 2

BCH 550

Arjumand Warsy

Databases

- A **database** is a **collection of data** for one or more multiple uses. One way of classifying databases involves the **type of content**, for example: DNA sequence database; protein sequence database; SNP database; bibliographic; image; customer databases; personal database; inventory databases; accounting databases etc etc.
- Databases consist of **software-based "containers"** that are structured to **collect and store information** so users can **retrieve, add, update or remove** such information in an automatic fashion. Database programs are designed for users so that they can add or delete any information needed. The structure of a database is tabular, consisting of rows and columns of information.

Examples of Databases used in biochemistry and molecular biology

- DNA and protein databases
- Pubmed/Medline
- Entrez
- ExPASy Molecular Biology server
- GenBank
- Genome sequence database
- EMBL Nucleotide Sequence Database
- DNA Data Bank of Japan
- DBCAT
- SALSA
- 123 Genomics
- Pfam
- EMALL for DNA sequences or UNIPROT
- UniProtKB
- UniRef
- UniParc
- TIGR
- SWISS-PROT
- Protein and Associated Topics Database
- Protein 3D and Associated Topics Database

Examples of Databases (conti..)

- **Physical Maps, Genetic Maps, Transcript Maps and Integration**
- CHLC Cooperative Human Linkage Center
- An STS-Based Map of the Human Genome
- Généthon Human Genome Research Centre
- CEPH Généthon
- Marshfield
- **SNP databases**
- **Other Sequence Related Databases**
- **Genetic Disease Databases**
- OMIM
- The Human Gene Mutation Database
- **Model Organism Databases**
- Mouse Genome Database
- Drosophila Genome database
- The C. elegans Genome Project
- The Saccharomyces Genome Database
- E. coli Genome Center's Home Page
- Amos's links to organism specific databases

DNA Sites on the Internet

- - Genetics Computer Group (Accelrys) <http://www.gcg.com/>
- - National Centre for Genome Resources <http://www.ncgr.org/>
- - National Centre for Biotechnology Information
<http://www.ncbi.nlm.nih.gov/>
- - Baylor College of Medicine Human Genome Centre and Baylor
Bioinformatics Unit: <http://www.hgsc.bcm.tmc.edu/>
- - The Human Genome Database <http://www.gdb.org/>
- - European Bioinformatics Institute <http://www.ebi.ac.uk/>
- - The Sanger Centre <http://www.sanger.ac.uk/>
- - European Molecular Biology Laboratory <http://www.embl-heidelberg.de/>
- - UK Medical Research Council <http://www2.mrc-lmb.cam.ac.uk/>
- - SWISS PROJ Databank <http://www.isb-sib.ch/>
- PERL <http://bio.perl.org/>
- - Human Genome Organization <http://www.gene.ucl.ac.uk/hugo/>
- - Human Genome Resources
<http://www.ncbi.nlm.nih.gov/genome/guide/human/>
- - Human Genome Biouser
<http://genome.ucsc.edu/cgi-bin/hgGateway>

How to Use Bioinformatics- using databases

- **Databases and how to use them**
- Using **PubMed** to become quickly knowledgeable on any biological subject
- Retrieving **protein and DNA sequences** relevant to our work
- Running our first sequence/database comparison with BLAST

Lets start with a protein we know nothing about: e.g. “dUTPase” by Searching PubMed

1. Using an Internet browser, navigate to www.ncbi.nlm.nih.gov/entrez/ on the World Wide Web (WWW). The following window opens:



The screenshot shows the PubMed website interface. At the top, the NCBI logo is on the left, and the PubMed logo with the URL www.pubmed.gov is in the center. To the right of the PubMed logo, it says "A service of the National Library of Medicine and the National Institutes of Health". In the top right corner, there is a "My NCBI" button with a question mark icon, and below it, "Sign in" and "Register" links.

Below the header, there is a navigation bar with tabs for "All Databases", "PubMed", "Nucleotide", "Protein", "Genome", "Structure", "MIM", "PMC", "Journals", and "Cross".

The main search area has a search box with "PubMed" selected in a dropdown menu, followed by "for dUTPase". There are "Go" and "Clear" buttons to the right of the search box.

Below the search box, there are buttons for "Limits", "Preview/Index", "History", "Clipboard", and "Details".

On the left side, there is a vertical menu with links: "About Entrez", "Text Version", "Entrez PubMed", "Overview", "Help | FAQ", "Tutorials", "New/Noteworthy", "E Utilities", "PubMed Services", "Journals Database", and "MeSH Database".

In the center, there are two bullet points:

- To get started, enter one or more search terms.
- Search terms may be topics, authors or journals.

Below the bullet points, there is a promotional box for "My NCBI" with the text "Set up an automated PubMed update in less than 5 minutes." and three numbered steps:

- (1) Get a [My NCBI account](#)
- (2) Save your search
- (3) Your PubMed updates can be e-mailed directly to you.

At the bottom of the promotional box, it says: "Read the [My NCBI Help](#) material to explore other options, such as automated updates of other databases, setting search filters, and highlighting search terms."

2. Type in dUTPase in the For window, and click the Go (enter) button.

- For the dUTPase example, we now have more than 400 references at our fingertips, more than enough to start unraveling the mysteries of dUTPase

3. For any entry in the Results list, click the associated author names.

- Titles stand out because they appear in a blue font, underlined —two signs of a clickable hyperlink.

36.
[Active site of mycobacterial dUTPase: structural characteristics and a built-in sensor.](#)
Varga B, Barabás O, Takács E, Nagy N, Nagy P, Vértessy BG.
Biochem Biophys Res Commun. 2008 Aug 15;373(1):8-13. Epub 2008 Jun 2. PMID: 18519027 [PubMed - indexed for MEDLINE] [Related articles](#)
37.
[Sequence analysis of a non-classified, non-occluded DNA virus that causes salivary gland hypertrophy of Musca domestica, MdSGHV.](#)
Garcia-Maruniak A, Maruniak JE, Farmerie W, Boucias DG.
Virology. 2008 Jul 20;377(1):184-96. Epub 2008 May 21. PMID: 18495197 [PubMed - indexed for MEDLINE] [Related articles](#) [Free article](#)
38.
[ORF018R, a highly abundant virion protein from Singapore grouper iridovirus, is involved in serine/threonine phosphorylation and virion assembly.](#)
Wang F, Bi X, Chen LM, Hew CL.
J Gen Virol. 2008 May;89(Pt 5):1169-78. PMID: 18420794 [PubMed - indexed for MEDLINE] [Related articles](#) [Free article](#)
39.
[Vaccinia virus lacking the deoxyuridine triphosphatase gene \(F2L\) replicates well in vitro and in vivo, but is hypersensitive to the antiviral drug \(N\)-methanocarbothymidine.](#)
Prichard MN, Kern ER, Quenelle DC, Keith KA, Moyer RW, Turner PC.
Virol J. 2008 Mar 5;5:39. PMID: 18321387 [PubMed - indexed for MEDLINE] [Related articles](#) [Free article](#)
40.
[Mechanism of dTTP inhibition of the bifunctional dCTP deaminase:dUTPase encoded by Mycobacterium tuberculosis.](#)
Helt SS, Thymark M, Harris P, Aagaard C, Dietrich J, Larsen S, Willemoes M.
J Mol Biol. 2008 Feb 15;376(2):554-69. Epub 2007 Dec 5. PMID: 18164314 [PubMed - indexed for MEDLINE] [Related articles](#)

4- Clicking the link calls up a page containing a rather detailed summary of the paper we chose.

- [Kinetic mechanism of human dUTPase, an essential nucleotide pyrophosphatase enzyme.](#)  Tóth J, Varga B, Kovács M, Málnási-Csizmadia A, Vértessy BG. J Biol Chem. 2007 Nov 16;282(46):33572-82. Epub 2007 Sep 11. PMID: 17848562 [PubMed - indexed for MEDLINE] [Related articles](#)  [Free article](#)

4. Save what you like to your hard drive by choosing your browser's File → Save As option.

Click here for printer-friendly format.

The screenshot displays the PubMed website interface. At the top, the NCBI logo and 'PubMed' branding are visible, along with the text 'A service of the National Library of Medicine and the National Institutes of Health'. A search bar contains the text 'dUTPase'. Below the search bar, there are navigation tabs for 'Limits', 'Preview/Index', 'History', 'Clipboard', and 'Details'. The search results are displayed in a list format, showing 'All: 323' and 'Review: 26'. A 'Send to' dropdown menu is open, listing options: 'Send to', 'Text', 'File', 'Printer', 'Clipboard', 'E-mail', 'RSS Feed', and 'Order'. The first three search results are visible, each with a checkbox and a 'Related Articles, Links' link.

Search PubMed for dUTPase

Display Summary Show 20 Sort by Send to

All: 323 Review: 26

Items 1 - 20 of 323

Page 1 of 17 Next

1: [Folstein S, Challe M.](#) [Use of yeast for detection of endogenous abasic lesions, their repair.](#) *Methods Enzymol.* 2006;408:79-91. PMID: 16793364 [PubMed - in process] [Related Articles, Links](#)

2: [Jiang YL, Ghazizadeh S, Krasby DJ, Slivers JT.](#) [Synthesis and high-throughput evaluation of triazolin uracil libraries for inhibition of human dUTPase and UNG2.](#) *Bioorg Med Chem.* 2006 May 4; [Epub ahead of print] PMID: 16673409 [PubMed - as supplied by publisher] [Related Articles, Links](#)

3: [McCarthy CK, Schepani A, Bandaru AM, Hahn-Peters LM, Kaiser M, Bous E, Pasnowski DG, Gilbert LH.](#) [Design, synthesis and evaluation of novel uracil amino acid conjugates for the inhibition of Trypanosoma cruzi dUTPase.](#) *Bioorg Med Chem Lett.* 2006 Jul 15;16(14):3809-12. [Epub 2006 May 5.] PMID: 16677613 [PubMed - in process] [Related Articles, Links](#)

You can select many references and get their abstracts together.

Click here for printer-friendly format.

The screenshot displays the PubMed search results interface. At the top, the NCBI logo and 'PubMed' branding are visible, along with the text 'A service of the National Library of Medicine and the National Institutes of Health'. The search bar contains 'dUTPase' and the results are displayed in a list format. The 'Send to' dropdown menu is open, showing options: Text, File, Printer, Clipboard, E-mail, RSS Feed, and Order. Two checkboxes are marked with red checkmarks, indicating selected references.

Search: PubMed for dUTPase

Display: Summary Show 20 Sort by

All: 323 Review: 26

Items 1 - 20 of 323

Page 1 of 17 Next

1: [Boiteux S, Oullet M.](#) [Use of yeast for detection of endogenous abasic lesions, their repair.](#) [Methods Enzymol. 2006;408:79-91.](#) PMID: 16793364 [PubMed - in process]

2: [Jiang YL, Chang S, Krosch DL, Shivers JT.](#) [Synthesis and high-throughput evaluation of triskelion uracil boronates for inhibition of human dUTPase and UNG2.](#) [Bioorg Med Chem. 2006 May 4; \[Epub ahead of print\].](#) PMID: 16673409 [PubMed - as supplied by publisher]

3: [McCarthy CK, Schipani A, Bianchi AM, Ruiz-Ferrer LM, Kaiser M, Brian E, Papanicolaou DG, Gilbert IH.](#) [Design, synthesis and evaluation of novel uracil amino acid conjugates for the inhibition of Trypanosoma cruzi dUTPase.](#) [Bioorg Med Chem Lett. 2006 Jul 15;16\(14\):3009-12. Epub 2006 May 3.](#) PMID: 16677813 [PubMed - in process]

e.g. The abstract of the paper we selected.

- J Biol Chem. 2007 Nov 16;282(46):33572-82. Epub 2007 Sep 11.
- **Kinetic mechanism of human dUTPase, an essential nucleotide pyrophosphatase enzyme.**
- [Tóth J](#), [Varga B](#), [Kovács M](#), [Málnási-Csizmadia A](#), [Vértessy BG](#).
- Institute of Enzymology, Biological Research Center, Hungarian Academy of Sciences, Karolina út 29, 1113 Budapest, Hungary. tothj@enzim.hu
- Human dUTPase is essential in controlling relative cellular levels of dTTP/dUTP, both of which can be incorporated into DNA. The nuclear isoform of the enzyme has been proposed as a promising novel target for anticancer chemotherapeutic strategies. The recently determined three-dimensional structure of this protein in complex with an isosteric substrate analogue allowed in-depth structural characterization of the active site. However, fundamental steps of the dUTPase enzymatic cycle have not yet been revealed. This knowledge is indispensable for a functional understanding of the molecular mechanism and can also contribute to the design of potential antagonists. Here we present detailed pre-steady-state and steady-state kinetic investigations using a single tryptophan fluorophore engineered into the active site of human dUTPase. This sensor allowed distinction of the apoenzyme, enzyme-substrate, and enzyme-product complexes. We show that the dUTP hydrolysis cycle consists of at least four distinct enzymatic steps: (i) fast substrate binding, (ii) isomerization of the enzyme-substrate complex into the catalytically competent conformation, (iii) a hydrolysis (chemical) step, and (iv) rapid, nonordered release of the products. Independent quenched-flow experiments indicate that the chemical step is the rate-limiting step of the enzymatic cycle. To follow the reaction in the quenched-flow, we devised a novel method to synthesize gamma-(32)P-labeled dUTP. We also determined by indicator-based rapid kinetic assays that proton release is concomitant with the rate-limiting hydrolysis step. Our results led to a quantitative kinetic model of the human dUTPase catalytic cycle and to direct assessment of relative flexibilities of the C-terminal arm, critical for enzyme activity, in the enzyme-ligand complexes along the reaction pathway.
- PMID: 17848562 [PubMed - indexed for MEDLINE]
- [Publication Types](#), [MeSH Terms](#), [Substances](#), [Grant Support](#)
- **Publication Types:**

- When the search yields many references, the best move is to start scanning a few pages — and select the most promising papers for future use by checking; the corresponding boxes will be different.

When you think you have selected enough references, just do the following:

- 1. Choose Abstract from the Display drop-down menu.**
The abstracts of all the papers you selected in the various pages appear for you to read.
- 2. If you want to print this display as is, choose File ⇨ Print from your browser menu.** Click here for printer-friendly format.
- 3. If you want to print this display in a more printer-friendly format, first: select Text from the scroll-down menu on the far right of the menu bar; to display this information in a non-HTML format, and then choose File ⇨ Print from your browser menu.**
- 4. If you'd like to save the file in the format of your choice, choose File ⇨ Save As from your browser menu, enter a new filename for the file, and then choose the file format you want to save to.**

Searching PubMed using author's names

1. Point your browser to www.ncbi.nlm.nih.gov/entrez/.
2. Type Abergel — the name of a prospective author — in the For window, and then click the Go button.
3. Results appear as follows:

The screenshot shows the PubMed search results page for the author 'Abergel'. The search was performed in the 'PubMed' database. The results are displayed in a list format, showing the first three items. Each item includes a checkbox, a numbered list item, the author names, the article title, the journal name, the publication date, and the PMID. The page also shows the total number of results (192) and the current page (1 of 10).

NCBI PubMed A service of the National Library of Medicine and the National Institutes of Health [My NCBI](#) [Sign In](#) [Register](#)

All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books

Search PubMed for Abergel [Save Search](#)

Limits Preview/Index History Clipboard Details

Display Summary Show 20 Sort by Send to

All: 192 Review: 18

Items 1 - 20 of 192 Page 1 of 10 Next

1: [Loth K, Abergel D, Pelupessy P, Delarue M, Lopes P, Ouazzani J, Duclert-Savatier N, Nilges M, Bodenhausen G, Stoven V](#) Related Articles, Links
Determination of dihedral Psi angles in large proteins by combining NH(N)/C(alpha)H(alpha) dipole/dipole cross-correlation and chemical shifts.
Proteins. 2006 Jun 19; [Epub ahead of print]
PMID: 16786593 [PubMed - as supplied by publisher]

2: [Cohen A, Chauvel C, Abergel E, Raffoul H, Diebold B](#) Related Articles, Links
Pulmonary regurgitant flow and detection of dip plateau.
J Am Soc Echocardiogr. 2006 May;19(5):580. No abstract available.
PMID: 16644448 [PubMed - indexed for MEDLINE]

3: [Abergel A, Sapin V, Dif N, Chassard C, Darcha C, Marcand-Sauvant J, Gaillard-Martinie B, Rock E, Dechelotte P, Sauvart P](#) Related Articles, Links
Growth Arrest and Decrease of alpha-SMA and Type I Collagen Expression by Palmitic Acid in the Rat Hepatic Stellate Cell Line PAV-1.
Dig Dis Sci. 2006 Apr 27; [Epub ahead of print]
PMID: 16642426 [PubMed - as supplied by publisher]

About Entrez
Text Version
Entrez PubMed
Overview
Help | FAQ
Tutorials
New/Noteworthy
E-Utilities
PubMed Services
Journals Database
MeSH Database
Single Citation
Matcher
Batch Citation Matcher
Clinical Queries
Special Queries
LinkOut
My NCBI
Related Resources
Order Documents

The result of a standard combined search for Abergel and dUTPase.

Click here for full text.

The screenshot shows a PubMed search result for the query 'Abergel dUTPase'. The search was performed on the PubMed database. The results list one article: '1. J Virol. 1999 Jan;73(1):751-3.' The article title is '"Hidden" dUTPase sequence in human immunodeficiency virus type 1 gp120.' The authors are Abergel C, Robertson DL, Claverie JM. The article is from the Laboratory of Structural and Genetic Information, CNRS EP-91, Marseille F-13402, France. The abstract text describes the discovery of a coding region homologous to the sequence for essential eukaryotic enzyme dUTPase in different genomic regions of several viral lineages. The abstract concludes that an ancestral dUTPase gene has evolved into the present primate lentivirus CD4 and cytokine receptor interacting region of gp 120. The PMID is 9847382, indexed for MEDLINE. The page also includes navigation links for full text, related articles, and a sidebar with various PubMed services and resources.

NCBI PubMed A service of the National Library of Medicine and the National Institutes of Health www.pubmed.gov My NCBI [Sign In] [Registered]

All Databases PubMed PubMed Central Protein Sequences Structure OMIM PMC Journals Books

Search PubMed [Go] [Clear] [Save Search]

Limits Preview/Index History Clipboard Details

Display: Abstract Show: 20 Sort by: Send to:

All: 1 Review: 0

PubMed can now automatically show related articles: [Try it!](#)

1. J Virol. 1999 Jan;73(1):751-3. [Full Text & PDF in PubMed Central](#) [Related Articles](#) [Links](#)

"Hidden" dUTPase sequence in human immunodeficiency virus type 1 gp120.

Abergel C, Robertson DL, Claverie JM.

Laboratory of Structural and Genetic Information, CNRS EP-91, Marseille F-13402, France. charnal@igs.cnrs-mrs.fr

A coding region homologous to the sequence for essential eukaryotic enzyme dUTPase has been identified in different genomic regions of several viral lineages. Unlike the nonprimate lentiviruses (caprine arthritis-encephalitis virus, equine infectious anemia virus, feline immunodeficiency virus, and visna virus), where dUTPase is integrated into the prot coding region, this enzyme has never been demonstrated to be present in the primate lentivirus genomes (human immunodeficiency virus type 1 [HIV-1], HIV-2, or the related simian immunodeficiency virus). A novel approach allowed us to identify a weak but significant sequence similarity between HIV-1 gp120 and the human dUTPase. This finding was then extended to all of the primate lentivirus lineages. Together with the recently reported fragmentary structural similarity between the V3 loop region and the Escherichia coli dUTPase (P. D. Kwong, R. Wyatt, J. Robinson, R. W. Sweet, J. Sodroski, and W. A. Hendrickson, Nature 393:648-659, 1998), our results strongly suggest that an ancestral dUTPase gene has evolved into the present primate lentivirus CD4 and cytokine receptor interacting region of gp 120.

PMID: 9847382 [PubMed - indexed for MEDLINE]

Some journals that offers free access to the full text (as indicated by the large colored rectangles that appear above the title).

4. Click the blue rectangle on the left.

Click here for a PDF reprint.

JVI Journal of Virology
Featuring ASM Journal's Subscriber Online Tools: Table/Figure Searching

HOME HELP FEEDBACK SUBSCRIPTIONS ARCHIVE SEARCH TABLE OF CONTENTS
Institution: Bibliotheque - Univ. de la Mediterranee Sign In as Member

Journal of Virology, January 1999, p. 751-753, Vol. 73, No. 1
0022-5382/99/0104-0751-03
Copyright © 1999, American Society for Microbiology. All rights reserved.

"Hidden" dUTPase Sequence in Human Immunodeficiency Virus Type 1 gp120

Chantal Abergel,¹ David L. Robertson, and Jean-Michel Claverie

Laboratory of Structural and Genetic Information, CNRS EP-91, Marseille F-13402, France

Received 27 July 1998/Accepted 14 October 1998

ABSTRACT

A coding region homologous to the sequence for essential eukaryotic enzyme dUTPase has been identified in different genomic regions of several viral lineages. Unlike the nonprimate lentiviruses (caprine arthritis-encephalitis virus, equine infectious anemia virus, feline immunodeficiency virus, and visna virus), where dUTPase is integrated into the *pol* coding region, this enzyme has never been demonstrated to be present in the primate lentivirus genomes (human immunodeficiency virus type 1 [HIV-1], HIV-2, or the related simian immunodeficiency virus). A novel approach allowed us to identify a weak but significant sequence similarity between HIV-1 gp120 and the human dUTPase. This finding was then extended to all of the primate lentivirus lineages. Together with the recently reported fragmentary structural similarity between the V3 loop region and the *Escherichia coli* dUTPase (P. D. Kwong, R. Wyatt, J. Robinson, R. W. Sweet, J. Sodroski, and W. A. Hendrickson, *Nature* 383:648-659, 1996), our results strongly suggest that an ancestral dUTPase gene has evolved into the present primate lentivirus CD4 and cytokine receptor interacting region of gp120.

TEXT

The role of the dUTPase protein is to produce dUMP to decrease the intracellular concentration of dUTP so that uracil cannot be misincorporated into DNA (12). This enzyme, essential in eukaryotes (4), has been acquired by multiple viral lineages (11). dUTPase sequences are highly variable.

This Article

- Abstract (PDF)
- Full Text (PDF)
- Alert me when this article is cited
- Alert me if a correction is posted

Services

- Similar articles in this journal
- Similar articles in PubMed
- Alert me to new issues of the journal
- Download to citation manager
- Link to other online articles
- Books from ASM Press

Google Scholar

Articles citing this Article

Published

- Published Citation
- Articles by Abergel, C.
- Articles by (Robertson, D. L.)

Top
• Abstract
• Text
• References

Top
• Abstract
• Text

Searching PubMed using fields

- PubMed offers many more ways to narrow down bibliographic searches to more specific topics. To get a handle on these different ways, however, we need to know more about the internal structure of Medline entries.

1. **Click the small arrow to the right of the Display drop-down menu.**

The contents of the drop-down menu appear: These options are different ways to display information related to the current article, or are links to related information.

2. **Select the MEDLINE option**

The Medline page appears, as shown in Figure

Frequently one can get flooded with an overwhelming number of *hits* (the standard term for search results) if we formulate queries that contain common names (such as *Smith* or *Cohen*) or use search terms (such as *Down*) that can occur in different contexts. To alleviate this problem, one can query PubMed by restricting the search for each word of the query to a given field.

Changing the default abstract format to MEDLINE.

Select MEDLINE format here.

The screenshot shows the PubMed interface. At the top, there is the NCBI logo and the PubMed logo with the URL www.ncbi.nlm.nih.gov/pubmed. Below this is a navigation bar with tabs for 'All Databases', 'PubMed', 'Nucleotide', 'Protein', 'Genome', 'Structure', 'Chem', 'PMC', 'Journals', and 'Books'. The search bar contains the text 'dUTPase' and has buttons for 'Go', 'Clear', and 'Save Search'. Below the search bar, there are tabs for 'Limits', 'Preview/Index', 'History', 'Clipboard', and 'Details'. The 'Display' dropdown menu is open, showing options: 'MEDLINE' (selected), 'Summary', 'Brief', 'Abstract', 'Abstract Plus', 'Citation', 'XML', 'UI List', 'LinkOut', 'ASN.1', 'Related Articles', 'Cited Articles', 'Cited in Books', 'CancerChrom Links', 'Domain Links', '3D Domain Links', 'GEO DataSet Links', 'Gene Links', 'Gene (GeneRIF) Links', 'Genome Links', 'Project Links', 'GENSAT Links', 'GEO Profile Links', 'HomoloGene Links', 'Nucleotide Links', 'OMIA Links', 'OMIM (calculated) Links', 'OMIM (cited) Links', 'BioAssay Links', and 'Compound Links'. The search results for 'dUTPase' are displayed below the menu. The first result is a full-text article titled 'Sequence in human immunodeficiency virus type 1 gp120...'. The abstract text is partially visible: '...to the sequence for essential eukaryotic enzyme dUTPase has been identified in different genomic regions...'. The page also includes a 'Related Articles, Links' section.

Selecting a field

Select MEDLINE format here.

The screenshot shows the PubMed website interface. At the top, there is a search bar with the text "A.bergel dUTPase" and buttons for "Go", "Clear", and "Save Search". Below the search bar, there are tabs for "Limits", "Preview/index", "History", "Clipboard", and "Details". A dropdown menu is open, showing various display options: "MEDLINE" (selected), "Summary", "Brief", "Abstract", "AbstractPlus", "Citation", "XML", "VI List", "LinkOut", "ASN.1", "Related Articles", "Cited Articles", "Cited in Books", "CancerChrom Links", "Domain Links", "3D Domain Links", "GEO DataSet Links", "Gene Links", "Gene (GeneRIF) Links", "Genome Links", "Project Links", "GENSAT Links", "GEO Profile Links", "HowtoGene Links", "Nucleotide Links", "OMA Links", "OMM (calculated) Links", "OMM (cited) Links", "BioAssay Links", and "Compound Links". The main content area displays a search result for "A.bergel dUTPase" with a link to the full text article and a link to PubMed Central. The article title is "Sequence in human immunodeficiency virus type 1 gp120." and the author is "Claverie JM". The article text is partially visible, starting with "to the sequence for essential eukaryotic enzyme dUTPase has been identified in different genomic regions...".

- In Figure below, the internal structure of a Medline database record is shown. The information is spread out over separate sections, called *fields*, each one preceded by a specific abbreviation — TI for title fields, AB for abstracts, AD for the laboratory address, AU for the authors, SO for the journal abbreviation, and so forth. This structure applies to all Medline records.

PMID - 9847382
 OWN - NLM
 STAT - MEDLINE
 DA - 19990128
 DCOM - 19990128
 LR - 20041117
 PUBM - Print
 IS - 0022-538X (Print)
 VI - 73
 IP - 1
 DP - 1999 Jan
 TI - "Hidden" dUTPase sequence in human immunodeficiency virus type 1 gp120.
 PG - 751-3
 AB - A coding region homologous to the sequence for essential eukaryotic enzyme dUTPase has been identified in different genomic regions of several viral lineages. -----
 -----, our results strongly suggest that an ancestral dUTPase gene has evolved into the present primate lentivirus CD4 and cytokine receptor interacting region of gp120.

The internal structure of a Medline record.

Conti...

AD - Laboratory of Structural and Genetic Information, CNRS EP-91, Mars:
F-12302, France. chantal@igs.cnrs-mrs.fr

FAU - Abergel, C
AU - Abergel C
FAU - Robertson, D L
AU - Robertson DL
FAU - Claverie, J M
AU - Claverie JM
LA - eng
PT - Journal Article
PL - UNITED STATES
TA - J Virol
JT - Journal of virology.
JID - 0113724
RN - 0 (HIV Envelope Protein gp120)
RN - EC 3.6.1. - (Pyrophosphatases)
RN - EC 3.6.1.23 (dUTP pyrophosphatase)
SB - IM
SB - X
MH - Amino Acid Sequence
MH - HIV Envelope Protein gp120/*chemistry
MH - HIV-1/* chemistry
MH - Humans
MH - Molecular Sequence Data
MH - Pyrophosphatases/*chemistry/genetics
MH - Research Support, Non-U.S. Gov't
EDAT - 1998/12/16
MHDA - 1998/12/16 00:01
PST - ppublish
SO - J Virol. 1999 Jan;73(1):751-3.

The internal structure of a Medline record (conti...)

- One can follow each term with the code (in brackets) that identifies the field in which to find the search term. Changing the field can totally modify the result of the searches. For instance, by entering three different queries — Down [AU], Down [TI], and Down [AD] into the For text box at the PubMed site.
- For example by running these queries, 275, 16,318, and 1,213, respectively, totally unrelated references were obtained. This narrowed the search a bit.

Using fields to find experts.

- 1. Point your browser to the PubMed site** (www.ncbi.nlm.nih.gov/entrez/).
- 2. In the For window, enter dUTPase [TIAB] Chicago [AD], and click Go.** By specifying the [TIAB] field, you'll be scanning the titles and abstracts of potential articles; the [AD] field specifies the address of the main laboratory associated with these articles. A couple of papers show up.
- 3. Click the list of authors (in blue, underlined).** You get the abstract of the corresponding article, together with the main laboratory address. Alternatively, to get all the abstracts at once, follow Steps 4 and 5.
- 4. Go back to the previous URL (after Step 2) using the Go Back button of your browser.**
- 5. In the Display drop-down menu, change the display option from *Summary* to *Abstract*.** A list of abstracts appears. At the top of each abstract, the name of the relevant laboratory in Chicago is displaced. With the information contained in the Medline records, names, street addresses, and sometimes e-mail addresses can be obtained. If necessary, use a telephone book (or a Web search engine) to supplement this information and find out how to contact these experts.

Searching for experts on dUTPase in Chicago.

Change from Summary to abstract here.

Restricted fields

The screenshot shows the PubMed search results page for the query "dUTPase [TAB] Chicago [AD]". The search results are displayed in a list format, showing 4 items. The first item is "Catalytic and structural role of the metal ion in dUTP pyrophosphatase" by Mustafa D, Bekas A, Vintony DG, Malhotra MW, published in Proc Natl Acad Sci U S A. 2003 May 13;100(10):1670-5. The second item is "Cloning and expression of the mouse deoxyuridine triphosphate nucle osidylpyrophosphatase gene: differs from the rat enzyme in that it lacks nuclear receptor interacting LXXLL motif" by Kan L, Jan S, Cook W, Cao SW, Ueda N, Yeldandi AV, Rao MS, Karwar PS, Reddy JK, published in Gene Expr. 1999;9(4):231-46. The third item is "Cloning and identification of rat deoxyuridine triphosphatase as an inhibitor of peroxisome proliferator-activated receptor alpha" by Chu B, Lin Y, Rao MS, Reddy JK, published in J Biol Chem. 1996 Nov 1;271(44):27670-6. The fourth item is "Identification of three genes nonessential for growth in cell culture near the right terminus of the unique sequences of long component of herpes simplex virus 1" by Barker DE, Roizman B, published in Virology. 1990 Aug;177(2):84-91. The page also includes a sidebar with navigation options and a search bar with the query "dUTPase [TAB] Chicago [AD]".

Search: PubMed [dUTPase [TAB] Chicago [AD]] [Go] [Clear] [Save Search]

Display: Summary [Show] 20 [Sort by] [Send to]

All: 4 [Review] [X]

Items: 1 - 4 of 4 [One page]

1: Mustafa D, Bekas A, Vintony DG, Malhotra MW. Catalytic and structural role of the metal ion in dUTP pyrophosphatase. Proc Natl Acad Sci U S A. 2003 May 13;100(10):1670-5. Epub 2003 Apr 29. PMID: 12721364 [PubMed - indexed for MEDLINE]

2: Kan L, Jan S, Cook W, Cao SW, Ueda N, Yeldandi AV, Rao MS, Karwar PS, Reddy JK. Cloning and expression of the mouse deoxyuridine triphosphate nucle osidylpyrophosphatase gene: differs from the rat enzyme in that it lacks nuclear receptor interacting LXXLL motif. Gene Expr. 1999;9(4):231-46. PMID: 10794525 [PubMed - indexed for MEDLINE]

3: Chu B, Lin Y, Rao MS, Reddy JK. Cloning and identification of rat deoxyuridine triphosphatase as an inhibitor of peroxisome proliferator-activated receptor alpha. J Biol Chem. 1996 Nov 1;271(44):27670-6. PMID: 8910338 [PubMed - indexed for MEDLINE]

4: Barker DE, Roizman B. Identification of three genes nonessential for growth in cell culture near the right terminus of the unique sequences of long component of herpes simplex virus 1. Virology. 1990 Aug;177(2):84-91. PMID: 2160730 [PubMed - indexed for MEDLINE]

Searching PubMed using limits

- PubMed offers yet another way of fine-tuning your queries: one can pre-define ranges for different attributes in different fields before running the search.

- 1. Point your browser to the PubMed site at www.ncbi.nlm.nih.gov/entrez/**
- 2. Type dUTPase in the For text box.** If you clicked the Go button at this point, you'd get a list of more than 320 articles — a bit too much reading for a nonspecialist! Restricting this list to the most recent review articles written in English would be practical.
- 3. Click the Limits tab, located just beneath the little arrow for the pulldown menu of the Search window.** The Limits screen appears, as shown in Figure. Here, below the Limited To line, you now have plenty of fields and attributes you can use for setting limits. You can go back later to this page and explore these various options.
- 4. Check the English box in the Language section.**
- 5. Check the Review box in the Type of Article section.**
- 6. Choose Title from the Default Tag drop-down menu.** Choosing Title restricts the search to articles for which the topic of dUTPase is central.
- 7. Finally, click the Go button.** Your (new and more concise) Results page appears.

Limiting a search for dUTPase to the titles of review articles in English.

Language

The screenshot shows a search filter interface with the following sections:

- Languages:** A list of languages with checkboxes. 'English' is checked. Other options include French, German, Italian, Japanese, Russian, Spanish, Afrikaans, and Albanian.
- Subsets:** A list of journal groups and topics with checkboxes. Options include Core clinical journals, Dental journals, Nursing journals, AIDS, Bioethics, Cancer, Complementary Medicine, and History of Medicine.
- Type of Article:** A list of article types with checkboxes. 'Review' is checked. Other options include Clinical Trial, Editorial, Letter, Meta-Analysis, Practice Guideline, Randomized Controlled Trial, Addresses, and Bibliography.
- Ages:** A list of age groups with checkboxes. Options include All Infant: birth-23 months, All Child: 0-18 years, All Adult: 19+ years, Newborn: birth-1 month, Infant: 1-23 months, Preschool Child: 2-5 years, Child: 6-12 years, Adolescent: 13-18 years, Adult: 19-44 years, and Middle Aged: 45-64 years.
- Tag Terms:** A section with a 'Default Tag' dropdown menu currently set to 'Title'.

Buttons for 'GO' and 'Clear All Limits' are located at the bottom right of the interface.

Type of article

Restricted field title

A few more tips about PubMed

How to get the most out of the query:

- Quoted queries (for example, “down syndrome”) behave as a single word, and are a great way to improve the relevance of your search.
- Impress your colleagues by starting using logical connectors (AND, OR, NOT) in your queries, as in dUTPase[TI] OR pyrophosphatase[TI] NOT Smith[AU]
- Adding initials to proper names (for example, “Abergel C”) can greatly reduce the number of hits.
- Write down the PubMed Identifier (the number in the PMID field) of that interesting paper you just found. It can be very useful in any subsequent searches for related items, such as associated gene and protein sequences.
- etc

Retrieving Protein Sequences

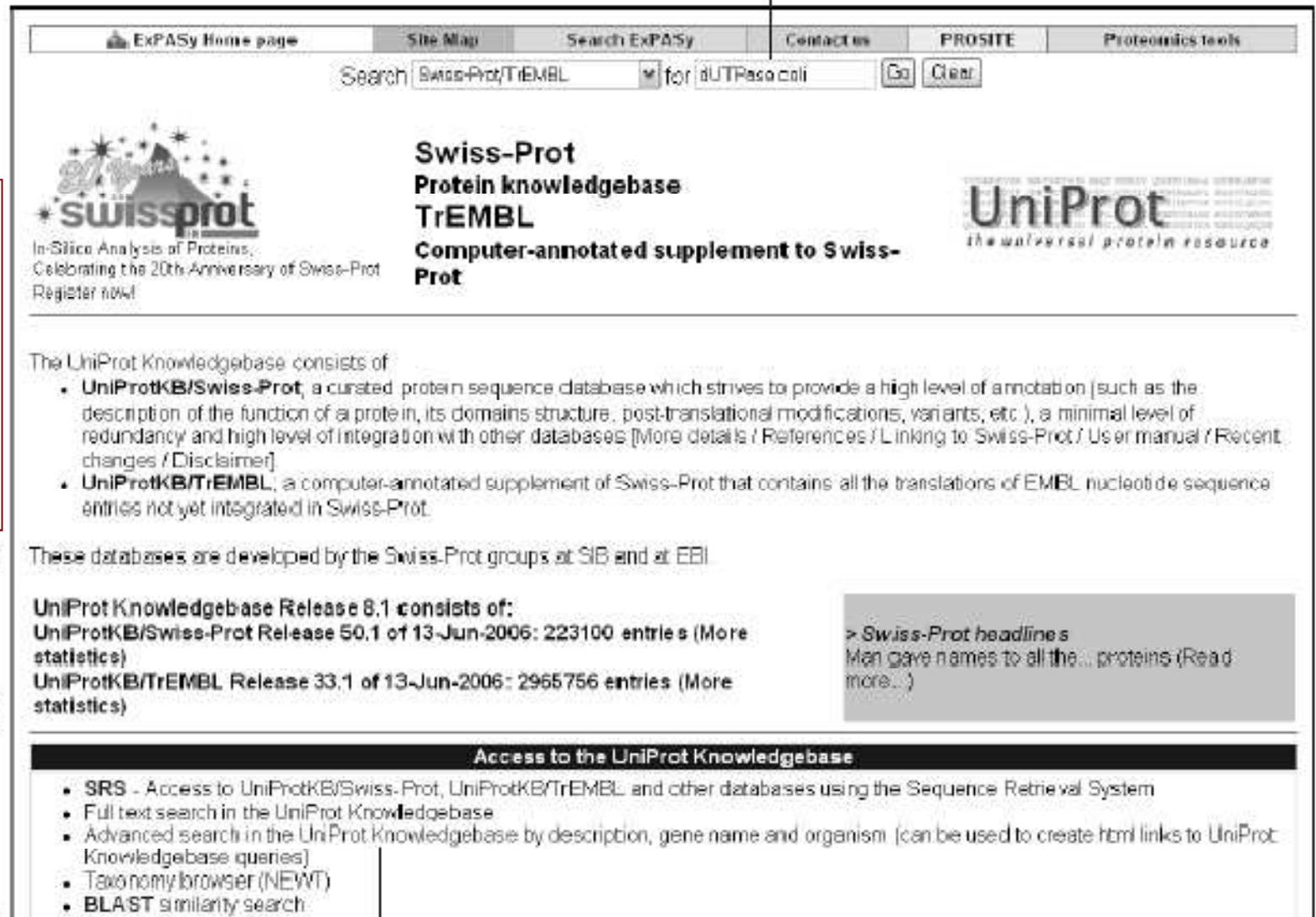
- Using other sites it is possible to retrieve relevant protein sequences from the Web to find out more about the subject at the molecular level.
- Using: **ExPASy**: A prime Internet site for protein information.
- The ExPASy server, is managed by Prof. Amos Bairoch, and is a world-leading resource for protein information.

Using ExPASy

- 1. Point your browser to www.expasy.org/sprot/, the Swiss-Prot database home page.**
- 2. Type dUTPase coli in the Search window, and then click the Search button.**
 - A list of three relevant protein sequences appears.
 - Now, when you click the last DUT_ECOLI (P06968) link, a full page of information about this dUTPase protein of *E. coli* appears on-screen, as shown in Figures 2-12 and 2-13. (No single browser window can hold such a wealth of information, so we've broken up the Results page into two figures.)

Search window

The Search window at the top of the Swiss-Prot home page.



The screenshot shows the UniProt search interface. At the top, there is a navigation bar with links: [EXPASy Home page](#), [Site Map](#), [Search ExPASy](#), [Contact us](#), [PROSITE](#), and [Proteomics tools](#). Below this is a search bar with a dropdown menu set to 'Swiss-Prot/TrEMBL', a text input field containing 'dUTRaso coli', and 'Go' and 'Clear' buttons. The main content area features the Swiss-Prot logo on the left, the UniProt logo on the right, and a central heading: 'Swiss-Prot Protein knowledgebase TrEMBL Computer-annotated supplement to Swiss-Prot'. Below the heading, there is a paragraph about the UniProt Knowledgebase and a bulleted list of database details. A grey box on the right contains a link to 'Swiss-Prot headlines'. At the bottom, a black bar with white text reads 'Access to the UniProt Knowledgebase', followed by a bulleted list of search tools.

[EXPASy Home page](#) [Site Map](#) [Search ExPASy](#) [Contact us](#) [PROSITE](#) [Proteomics tools](#)

Search for

**Swiss-Prot**
Protein knowledgebase
TrEMBL
Computer-annotated supplement to Swiss-Prot

**UniProt**
the universal protein resource

In-Silico Analysis of Proteins:
Celebrating the 20th Anniversary of Swiss-Prot
[Register now!](#)

The UniProt Knowledgebase consists of

- **UniProtKB/Swiss-Prot**, a curated protein sequence database which strives to provide a high level of annotation (such as the description of the function of a protein, its domains structure, post-translational modifications, variants, etc.), a minimal level of redundancy and high level of integration with other databases ([More details](#) / [References](#) / [Linking to Swiss-Prot](#) / [User manual](#) / [Recent changes](#) / [Disclaimer](#))
- **UniProtKB/TrEMBL**, a computer-annotated supplement of Swiss-Prot that contains all the translations of EMBL nucleotide sequence entries not yet integrated in Swiss-Prot.

These databases are developed by the Swiss-Prot groups at SIB and at EBI

UniProt Knowledgebase Release 8.1 consists of:
UniProtKB/Swiss-Prot Release 50.1 of 13-Jun-2006: 223100 entries ([More statistics](#))
UniProtKB/TrEMBL Release 33.1 of 13-Jun-2006: 2965756 entries ([More statistics](#))

[> Swiss-Prot headlines](#)
Man gave names to all the... proteins ([Read more...](#))

Access to the UniProt Knowledgebase

- **SRS** - Access to UniProtKB/Swiss-Prot, UniProtKB/TrEMBL and other databases using the Sequence Retrieval System
- Full text search in the UniProt Knowledgebase
- Advanced search in the UniProt Knowledgebase by description, gene name and organism (can be used to create html links to UniProt Knowledgebase queries)
- Taxonomy browser (NEWT)
- BLAST similarity search

Click here for Advanced Search.

Three dUTPase sequences from the Escherichia coli bacterium.

ExPASy Home page | Site Map | Search ExPASy | Contact us | Swiss-Prot

Hosted by SIB Switzerland | Mirror sites: |Australia| Brazil| Canada| China| Korea|

Search for

Search in UniProt Knowledgebase (Swiss-Prot and TrEMBL) for: dUTPase coli

UniProtKB/Swiss-Prot Release 50.1 of 13-Jun-2006
UniProtKB/TrEMBL Release 33.1 of 13-Jun-2006

- Number of sequences found in UniProt Knowledgebase (Swiss-Prot₍₀₎ and TrEMBL₍₀₎): 3
- Note that the selected sequences can be saved to a file to be later retrieved; to do so, go to the bottom of this page.
- For more directed searches, you can use the Sequence Retrieval System SRS.

Search in UniProtKB/Swiss-Prot: There are matches to 3 out of 223100 entries

- DUT_EC057 (P64007)
Deoxyuridine 5'-triphosphate nucleotidohydrolase (EC 3.6.1.23) (dUTPase) (dUTP pyrophosphatase) {GENE: Name=dut, Ordered_LocusNames=Z5084, ECs4515} - Escherichia coli O-157:H7
- DUT_EC06 (P64006)
Deoxyuridine 5'-triphosphate nucleotidohydrolase (EC 3.6.1.23) (dUTPase) (dUTP pyrophosphatase) {GENE: Name=dut, Ordered_LocusNames=c_4464} - Escherichia coli O6
- DUT_EC01 (P06968)
Deoxyuridine 5'-triphosphate nucleotidohydrolase (EC 3.6.1.23) (dUTPase) (dUTP pyrophosphatase) {GENE: Name=dut, Synonyms=chaS, sof, Ordered_LocusNames=b3640} - Escherichia coli

3 relevant dUTPase entries found.