NCBI News

National Center for Biotechnology Information

National Library of Medicine

National Institutes of Health

Fall 1999

Entrez Search Services Enhanced

NCBI's popular Entrez search service has been overhauled to accommodate additional data resources and facilitate more finely tuned searches demanded by the explosion of data in the sequence, structure, and literature databases. The new Entrez not only offers an enhanced search interface to the five familiar databases—Nucleotide, Protein, Structure, Genome, and PubMed—but also adds a sixth database, PopSet for population studies, to the mix.

Central to the new interface is a standard query box on every page that allows you to select a target database and initiate a new search from anywhere within Entrez, making it easier to move from database to database. The new Entrez also offers several features for composing detailed search queries and managing the results, including Limits, Index, History, and Clipboard. These options can be selected from every Entrez page, just below the query box. Detailed documentation, including search examples, is available from a **Help** link on all Entrez pages.

Limits: Setting Field Restrictions

The Limits feature is used to restrict a search term to a particular database field. Although field specification has always been available in Entrez, the new release has two major changes. First, commonly used field restrictions, such as Molecule Type for DNA sequences and Language for PubMed searches, are listed for each Entrez database and can be selected easily from pull-down menus. Commonly desired data exclusions, such as EST and STS sequences, are easily invoked by checking a box. Second, limits can be applied globally to all searches of a particular database, relieving you of having to incorporate the same restrictions repeatedly in multiple iterations of the searching process.

Just as data fields vary from database to database within Entrez, the parameters available for limiting are also database-dependent. For example, limits unique to PubMed include Publication Type and Language. Limits for the Nucleotide and Protein databases include Gene Location and Modification Date, among others.

Index: Browsing Available Search Terms

The Index option allows you to browse the alphabetical indexes of

Continued on page 2

NCBI Home Page Redesigned

NCBI's recently reorganized home page is designed to better accommodate its growing set of resources. Since 1994, when the Web site was launched, the number of services has grown from 4 to almost 40! The new design not only offers the needed flexibility and organization, but also provides added interest through frequently updated announcements and news items.

The new page retains the familiar top row of buttons for prominent access to the most frequently used resources—and adds a new query box that launches searches of GenBank, PubMed, and other databases directly from the home page.

Continued on page 8

In this issue

NCBI News

NCBI News is distributed four times a year. We welcome communication from users of NCBI databases and software and invite suggestions for articles in future issues. Send correspondence to *NCBI News* at the address below.

NCBI News
National Library of Medicine
Bldg. 38A, Room 8N-803
8600 Rockville Pike
Bethesda, MD 20894
Phone: (301) 496-2475
Fax: (301) 480-9241
E-mail: info@ncbi.nlm.nih.gov

Editors
Dennis Benson
Barbara Rapp

Contributors
Mark Cavanaugh
Paul Kitts
Tom Madden
Margaret McGhee
Ilene Mizrachi
Steve Sherry
Yanli Wang

Writer
David Wheeler

Editing, Graphics, and Production
Marla Fogelman
Veronica Johnson
Jennifer Vyskocil

Design Consultant
Gary Mosteller

In 1988, Congress established the National Center for Biotechnology Information as part of the National Library of Medicine; its charge is to create information systems for molecular biology and genetics data and perform research in computational molecular biology.

NIH Publication No. 99-3272

ISSN 1060-8788
ISSN 1098-8408 (Online Version)

Entrez Search Services

Continued from page 1

search terms in each field of the Entrez databases. This is the same basic functionality that was provided through the List Terms search mode in earlier versions of Entrez, with some improvements in ease of use. As you browse, you can select and search the terms you find. Compound searches can be constructed by choosing terms iteratively from the Index list, and combining them by using one of three logical operator buttons provided: **AND, OR,** or **NOT.**

History: Recording Previous Searches

The new Entrez automatically retains a search history that can be used to build up a very selective search from a combination of previous searches. The History extends back to a maximum of 100 searches per Entrez session and expires after about 1 hour of inactivity. The History is presented as a table giving, for each search, a search number, the search query, the time of the search, and the number of database hits. Clicking on the number of hits recalls the results of the search. The search number can be entered into the query box as a substitute for the literal query and may be considered a variable. To combine searches "#1" and "#2" and retrieve their intersection, simply enter **"#1 AND #2"** into the main query box. This result becomes Search History entry #3 and may be used to construct other searches by using "#3" in place of the actual query. A separate History is maintained for each database.

Clipboard: Collecting Search Results

The Clipboard serves as a collection bin for search results. It allows you to save some or all of the documents retrieved from many separate searches as you go along. Items are placed on the Clipboard by checking a selection box next to their item number. Multiple items may be checked in this manner and then added to the Clipboard with a single click. Alternatively, an entire set of results can be saved at once by clicking on **Add to Clipboard.** Separate Clipboards are maintained for each database, and contents are removed after 1 hour of inactivity.

PopSet: Adding a New Entrez Database

The new PopSet database is available only under the new release of Entrez. It can be selected in the same manner as the other databases, that is, from the list box next to the main query box if you want to launch a search directly from another database, or from the top row of database names if you want to go to the PopSet home page.

PopSet contains aligned nucleotide or protein sequences submitted as a set resulting from a population, a phylogenetic, or mutation study. These kinds of populations sets are useful in describing such events as evolution and population variation. —*MM, BR, DW*

Both new and old interfaces are running in parallel during an initial testing phase. To try the new Entrez, click on the prominent green ellipse atop the Entrez home page.

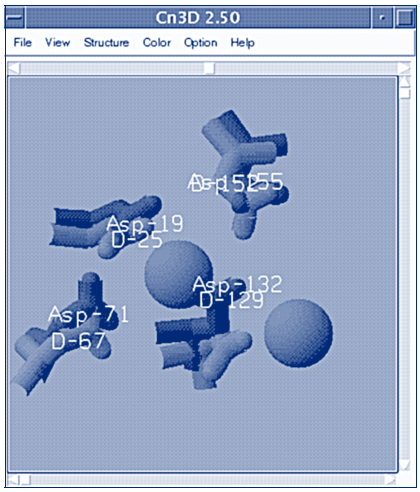
Cn3D 2.5 Offers Feature Editing and Global Save

NCBI has recently released version 2.5 of the Cn3D macromolecular structure viewer. PC users will enjoy a smooth, single-click installation process and all users will appreciate the usability enhancements in this latest release.

Cn3D 2.5 offers a greatly expanded array of annotation tools, including the ability to define molecular features and specify their display characteristics. Greater control over every aspect of the molecular display is facilitated through four control panels, which govern molecular color schemes, visibility, rendering, and labeling. After a molecule, or set of molecules, has been rendered and annotated, the entire ensemble of coordinate data and display specifications can be saved for later viewing or transmittal to colleagues.

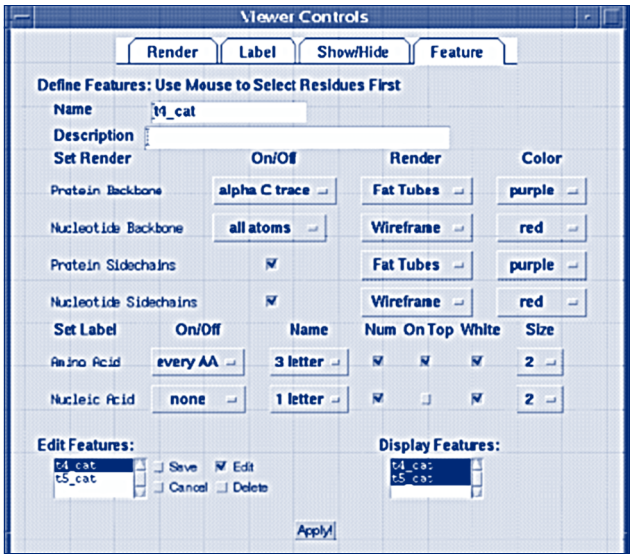
These new annotation features are illustrated below, based on an example from the online Cn3D tutorial. Figure 1 shows the area of the active site of two structurally aligned 5′-3′ exonucleases. The structural alignment was produced by NCBI's Vector Alignment Search Tool (VAST) and downloaded by Cn3D from the Molecular Modeling Database (MMDB). The dark-colored residues are those of the bacteriophage T4 enzyme with PDB code "1TFR,"[1] and the light-colored residues are those of the T5 enzyme with PDB code "1EXN."[2] The two dark spheres represent catalytically important magnesium ions present in the 1TFR structure. The 1TFR and 1EXN exonucleases are responsible for removing RNA primers used during bacteriophage DNA replication.

Figure 1



Cn3D view of structural alignment of 1EXN and 1TFR, as restricted by user-specified features.

Figure 2



Cn3D Feature panel showing custom rendering settings for the four catalytic acidic residues of the user-defined "t4_cat" feature.

To generate this view, the alpha carbon traces of the default Cn3D Neighbor representation have been turned off, and two groups of amino acids have been defined as "Features" and independently rendered as "fat tubes." This rendering mode reveals the excellent alignment of the corresponding side chains in the two structures. The background color was also changed from its default of black to gray, another new option in Cn3D 2.5, using the **Options** menu.

Features were defined and given independent rendering and visibility settings using the **Viewer Controls** window. To define the feature "t4_cat," for instance, 1TFR residues D19, D71, D132, and D155 were first selected in the **Sequence** window. The **Feature** panel of the **Viewer Controls** window, shown in Figure 2, was then used to give the selected residues a feature name and define their display and visibility properties.

The alignment, rendering settings, and user-defined features, "t4_cat" and "t5_cat," can be saved in a single file for later retrieval by using the new **Global Save** function provided by the **File/Save/All** menu option.

Cn3D is a project of NCBI's Molecular Structure research group, headed by Steve Bryant. The new features in release 2.5 were developed primarily by NCBI scientists Yanli Wang, Lewis Geer, Colombe Chappey, and Jonathan Kans.

To download or read more about Cn3D, visit www.ncbi.nlm.nih.gov/Structure/CN3D. *—DW, YW*

Notes

1. Mueser, TC, et. al. *Cell* 85(7):1101–12, 1996.
2. Ceska, TA, et. al. *Nature* 382(6586): 90–3, 1996.

News Briefs

Drosophila Sequence Gets 40-Megabase Boost from Celera

The complete *Drosophila melanogaster* sequence is near completion. Currently, GenBank contains 40 million base pairs of finished *Drosophila* genomic sequence from the NIH-funded Berkeley *Drosophila* Genome Project and the European *Drosophila* Genome Project. An additional 100 million base pairs of unfinished sequence, including over 600 BACs from the Berkeley project, are in the HTG division of GenBank. Recently, Celera Genomics submitted another 10 million base pairs of unfinished contigs, also in the HTG division. To BLAST against these new sequences, select the HTGS database from the Advanced BLAST page, and restrict the search to *Drosophila melanogaster* using the organism filter box.

More dbSNP Data by FTP

NCBI now provides the full dbSNP database in three formats: flatfile, FASTA, and SQL DDL/table dumps. Complete copies of the database will be refreshed weekly. See the README file for a complete description of the various formats (ftp://ncbi.nlm.nih.gov/snp/00readme).

Human Contig Breaks 10-Megabase Barrier

A segment of DNA sequence from chromosome 22 has become the first human continuous sequence over 10 Mb in length. The Sanger Centre, Washington University, and the University of Oklahoma contributed to this sequence. See the list of contigs for human chromosome 22 on NCBI's **Human Genome Sequencing** page.

UniGene for Zebra Fish

The zebra fish *(Danio rerio)* has been added to the UniGene lineup. Over 5,600 zebra fish clusters are represented. Zebra fish joins the human, mouse, and rat versions of UniGene.

Submitting GSS Sequences

After January 1, 2000, GenBank will no longer process Genome Survey Sequence (GSS) submissions made with BankIt or Sequin. Instead, use the custom GSS submission procedures (www. ncbi.nlm.nih.gov/dbEST/how_to_submit.html. The e-mail address for GSS submissions is: batch- sub@ncbi.nlm.nih.gov.

Standalone BLAST 2.0.10

The latest version of Standalone BLAST is the first to include a standalone version of BLAST 2 Sequences (called bl2seq). Standalone BLAST 2.0.10 also allows searches of multiple databases, which circumvents two current limits on database size—a maximum of 2 gigabytes for any database and a maximum of 4 billion base pairs for nucleotide databases. Databases that exceed these limits can now be formatted as a series of smaller "volumes" using the program formatdb, also included in the BLAST 2.0.10 package. The blastall program performs the multidatabase search.

COGs Includes 21 Genomes

Clusters of Orthologous Groups (COGs) now incorporates 21 complete genomes, tripling the number of organisms initially represented. In addition to phylogenetic patterns, the COGs may now be searched using free-text words or protein and gene names. PSI-BLAST searches may be launched using a set of similar sequences from the COGs database to construct a position-specific scoring matrix (PSSM). This PSSM can then be used to search for remote homologs.

BLAST E-Mail Server

The BLAST e-mail server has been upgraded to use the new QBLAST system. The server therefore no longer supports BLAST 1.4 searches or the RIPEM encryption software.

GenBank Adds CON Division for Assembly Data

GenBank, EMBL, and DDBJ have established a special-purpose division, Contig (CON), for exchanging assembly instructions for data in the international DNA sequence databases. The CON division contains no sequence data, but rather instructions for the assembly of sequence data from multiple GenBank records.

DNA sequence records in GenBank (as well as EMBL and DDBJ) are currently limited to 350 kb for flexibility in data exchange, analysis, searching, and display. Sequences that exceed 350 kb are split into multiple smaller records with separate accession numbers. The CON division contains information on how to reassemble the full-length contig. It also includes instructions for constructing assemblies of non-contiguous sequence shown in Entrez as "segmented sets."

General users of GenBank and Entrez should notice no changes in the services. This data has been present in the ASN.1 format and implemented in Entrez for some time. The CON division simply formalizes procedures in development over the past few years, and makes the assembly information available in a more familiar format to users who download the GenBank database.

The CON division will be available with the December 15, 1999 release of GenBank. It applies only to sequence data as deposited in GenBank, EMBL, or DDBJ. It is not being used for other genome assembly projects. —*BR, DW*

VecScreen: BLAST the Vector Out of Your Sequence

(GenBank Indexers Say "Use It!")

VecScreen is a system for quickly identifying segments of a nucleic acid sequence that may be of vector origin. VecScreen was originally developed at NCBI for the GenBank Annotation Staff, who use it to verify that sequences submitted for inclusion in the database are free from contaminating vector sequence. It is now available to the research public through the VecScreen Web site to help researchers identify and remove any segments of vector origin prior to sequence analysis or submission. Early identification of any foreign segments can avert erroneous conclusions about the biological significance of the sequence, prevent time and effort from being wasted in analysis of contaminated sequence, and speed the release of the sequence in a public database.

The UniVec Database

VecScreen performs an optimized blastn search, with the sequence to be screened as the query, of a specialized non-redundant vector database called UniVec. Screening against UniVec is efficient because a large number of redundant subsequences have been eliminated to create a database with only one copy of every unique sequence segment from a large number of vectors. In addition to vector sequences, UniVec contains sequences for those adapters, linkers, and primers commonly used in the process of cloning cDNA or genomic DNA. This enables detection of contamination with these oligonucleotide sequences during the vector screen. Elimination of redundant sequence segments reduces UniVec to less than 15% of the size of an equivalent database containing the full sequences for the same set of vectors. UniVec also uses a "pseudo-circularization" process, appending the first 49 bases of a circular vector sequence to the end of the vector to avoid missing a match due to end effects. The current version of UniVec represents 971 vector and oligonucleotide sequences. To see the sequences used to build the database, go to the **UniVec Representation List,** accessible from the VecScreen page.

The VecScreen Graphic

The blastn output from VecScreen is summarized using a graphical representation of the query sequence, which is color-coded to show the location of segments that match vector sequences. The matches are color-coded at four levels of significance: strong, moderate, weak, and suspicious.

Give VecScreen a try at www.ncbi.nlm.nih.gov/VecScreen/VecScreen.html.
—PK, DW

Selected Recent Publications by NCBI Staff

Boeke, JD, and **OK Pickeral.** Retroshuffling the genomic deck. *Nature* 398(6723):108–9, 111, 1999.

Copley, RR, J Schultz, **CP Ponting,** and P Bork. Protein families in multicellular organisms. *Curr Opin Struct Biol* 9(3):408–15, 1999.

Gladyshev, VN, M Krause, XM Xu, KV Korotkov, GV Kryukov, QA Sun, BJ Lee, **JC Wootton,** and DL Hatfield. Selenocysteine-containing thioredoxin reductase in *C. elegans*. *Biochem Biophys Res Commun* 259(2):244–9, 1999.

Hahn, T, E Matala, **C Chappey,** and N Ahmad. Characterization of mother-infant HIV type 1 *gag* p17 sequences associated with perinatal transmission. *AIDS Res Hum Retroviruses* 15(10):875–88, 1999.

Kondrashov, AS, and FA Kondrashov. Interactions among quantitative traits in the course of sympatric speciation. *Nature* 400(6742):351–4, 1999.

Krizman, DB, **L Wagner, A Lash**, R Strausberg, and MR Emmert-Buck. Cancer Genome Anatomy Project: EST sequencing and the genetics of cancer progression. *Neoplasia* 1(2):101–6, 1999.

Leipe, DD, L Aravind, and **EV Koonin.** Did DNA replication evolve twice independently? *Nucleic Acids Res* 27(17):3389–401, 1999.

Matsuo, Y, and **SH Bryant.** Identification of homologous core structures. *Proteins* 35(1):70–9, 1999.

Wolf, YI, L Aravind, and **EV Koonin**. *Rickettsiae* and *Chlamydiae*: evidence of horizontal gene transfer and gene exchange. *Trends Genet* 15(5):173–5, 1999.

Zaman, V, M Zaki, J Howe, M Ng, **DD Leipe,** ML Sogin, and JD Silberman. *Hyperamoeba* isolated from human feces: description and phylogenetic affinity. *Eur J Protistol* 35(2):197c–207, 1999.

Using the New QBLAST Reformatting Feature to Detect Conserved Amino Acids

One of the strengths of the recently implemented QBLAST system is its ability to generate output quickly in any of several formats after a search has been completed. Choosing one of the Master-Slave alignment formats, rather than the default Pairwise format, allows one to easily visualize patterns of residue conservation in sequence families.

As an example, consider the CxC family of small cytokines. This protein family is characterized by a conservation of four cysteine residues involved in disulfide bond formation. One can compare the human sequence of Interleukin-8 (P10145), a member of this group, with the sequences of related cytokines from the sheep, pig, and cow by running a blastp search of the nr database using a 49–amino acid stretch from P10145 containing all four cysteines as the query. It is possible to exclude other very closely related human sequences and focus on the sheep, pig, and cow by entering "Cetartiodactyla" into the organism filter box on the BLAST search page.

The search shown here was run using an Expect value of ".01," and the Flat-Master Slave with Identities alignment format. This alignment format uses periods in place of residues in the "slave" sequences (database hits) that are identical to the aligned residues in the "master" sequence (the query sequence). The output here clearly shows the four conserved cysteine residues as uninterrupted columns of periods (highlighted).
—DW

```
                                                                    Score    E
Sequences producing significant alignments:                         (bits)  Value

sp|P36925|IL8_SHEEP  INTERLEUKIN-8 PRECURSOR (IL-8) >gi|5431...       96   7e-22
sp|P79255|IL8_BOVIN  INTERLEUKIN-8 PRECURSOR (IL-8) >gi|1699...       93   4e-21
emb|CAA43461|  (X61151) interleukin-8 [Sus scrofa]                    92   6e-21
pir||A44253  alveolar macrophage chemotactic factor-I (AMCF-...       92   6e-21
sp|P26894|IL8_PIG  INTERLEUKIN-8 PRECURSOR (IL-8) (ALVEOLAR ...      92   6e-21
gb|AAD02808|  (AF061521) interleukin-8 [Bos taurus]                   65   1e-12
sp|P22952|AMC2_PIG  ALVEOLAR MACROPHAGE CHEMOTACTIC FACTOR I...       58   2e-10
sp|P80221|GCP2_BOVIN  GRANULOCYTE CHEMOTACTIC PROTEIN 2 (GCP...       57   3e-10
sp|P30034|PLF4_PIG  PLATELET FACTOR 4 (PF-4)                          53   4e-09
sp|O46677|GROB_BOVIN  GROWTH REGULATED PROTEIN HOMOLOG BETA ...       53   5e-09
sp|O46676|GROA_BOVIN  GROWTH REGULATED PROTEIN HOMOLOG ALPHA...       53   5e-09
sp|O46675|GROG_BOVIN  GROWTH REGULATED PROTEIN HOMOLOG GAMMA...       52   7e-09
sp|P43030|PF4L_PIG  PLATELET BASIC PROTEIN PRECURSOR (PBP) >...       52   7e-09
gi|2735499  (U95814) GRO [Ovis aries]                                 52   9e-09
bbs|140204  neutrophil-activating peptide 2, pNAP-2-S {short...       51   2e-08
bbs|140202  neutrophil-activating peptide 2, pNAP-2-L {long...        51   2e-08
sp|P30035|PLF4_SHEEP  PLATELET FACTOR 4 (PF-4)                        49   8e-08
pdb|1PLF|A  Bos taurus >gi|494447|pdb|1PLF|B Bos taurus >gi|...       46   5e-07
sp|P02777|PLF4_BOVIN  PLATELET FACTOR 4 (PF-4) >gi|72110|pir...       44   3e-06
bbs|161256  epithelial-derived neutrophil-activating peptide...       43   6e-06

tmpseq_1 1  LRCQCIKTYSKPFHPKFIKELRVIESGPHCANTEIIVKLSDGRELCLDP      49
463254   32 ........H.T..................E.S......TN.K.V....      80
1699354  32 ........H.T..................E.S......TN.N.V..N.      80
516197   32 ......N.H.T..................E.S......VN.K.V....      80
346456   32 ......N.H.T..................E.S......VN.K.V....      80
1235612  32 ......N.H.T..................E.S......VN.K.V....      80
4106336  1                ...............E.S......TN.N.V..N.      35
399030   50 ...M.LT.-TPGI...M.SD.Q..PA..Q.SKA.V.AT.KN.K.V....      97
462169   10 ...V.LT.-TPGI...TVSD.Q..AA..Q.SKV.V.AT.KN...V....      57
585702   23 ...V.V..I.G-VS..H.SS.E..GA....PSPQL.AT.KK.HKI....      70
3913773  38 .....LQ.L-QGI.L.N.QSVK.TTP....DQ..V.AS.KT.Q.V..N.      85
3913772  38 .....LQ.L-QGI.L.N.QSVK.TTP....DQ..V.AS.KT.Q.V..N.      85
3913771  37 .....LQ.L-QGI.L.N.QSVK.TTP....DQ..V.AT.KT.Q.V..N.      84
457754   52 ...L.LN.V.G-I..SN.QS.E..RA.A...KV.V.AT.KNDKKI....      99
2735499  37 .....LQ.V-QGI.L.NMQSVK.TPP....GQ..V.AT.KT.Q.V..N.      84
452976   6  ...L.LN.V.G-I..SN.QS.E..RA.A...KV.V.AT.KNDDKI....      53
739009   13 ...L.LN.V.G-I..SN.QS.E..RA.A...KV.V.AT.KNDDKI....      60
266800   23 .Q.V.L..T.G-I..RH.SS.E..GA.L..PSPQL.AT.KT..KI...      69
494448   10 .Q.V.L..T.G-IN.RH.SS.E..GA.L..PSPQL.AT.KT..KI...      56
225004   23 .Q.V.L..T.G-IN.RH.SS.E..GA.T..PSPQLLATKKT..KI...      69
861479    5  ...V.LT.-TPGI...TVSD.Q..AA..Q.SKV.V.AT.KN.          45
```

The BLAST Lab feature is intended to provide detailed technical information on some of the more specialized uses of the BLAST family of programs. Topics are selected from the range of questions received by the BLAST Help Group.

Q & A Frequently Asked Questions

Q.	A.
Is it possible to constrain a BLAST search to a particular taxonomic group, such as mammalia?	This is easily accomplished by using the organism filter box located in the middle of the Advanced BLAST page. Typing "mammalia" into this box will limit your BLAST search to the taxon Mammalia. To browse the NCBI taxonomic tree in order to determine which taxon to use, follow the link below the organism filter box entitled **Explore the taxonomy database at NCBI.**
RefSeq NM_xxxxxx and GenBank AFxxxxxx appear to be duplicates. Will one be removed?	No, both records will continue to be available. RefSeq and GenBank are separate databases, and both are included in the Entrez Nucleotides data set. RefSeq is a curated database in which each record encapsulates our current understanding of a gene, similar to a review article. In contrast, GenBank is an archival database where each record represents the work of a specific research group.
What is a taxonomy ID number (tax ID)?	The tax ID is a stable unique identifier for each taxon (for a species, a family, an order, or any other group in the taxonomy database). The tax ID apppears in the GenBank records as a "source" feature, e.g., /db_xref="taxon: 9606" for *Homo sapiens*. Tax IDs can be used to retrieve sequences in Entrez with a syntax such as: txid9606[organism]. This will retrieve only *Homo sapiens* sequences. BLAST searches can also be limited to particular taxa by entering the tax ID into the organism filtering box.
How do I add a feature to my sequence submission using Sequin?	Once you have imported your sequence, use Sequin's **Annotate** menu to see the list of standard features. Select the feature of interest from the submenus, for example, **Coding Regions and Transcripts/CDS.** This will invoke a feature-specific form that allows you to enter essential information about the feature and add it to your GenBank submission. To edit an existing feature, just double click on the feature. This will bring up a feature-specific form showing the current data, which can then be changed.
I have pasted a long sequence into ORF-Finder, but only want ORFs within a small block. Should I remove the surrounding sequence?	No, you can specify a nucleotide range using the "From" and "To" boxes below the query box. ORF-Finder will find ORFs only within the range specified.

New NCBI Home Page

Continued from page 1

A left-hand panel of links serves as a table of contents for the entire site, grouping services into nine functional categories: About NCBI, GenBank, Molecular Databases, Literature Databases, Genomic Biology, Tools, Research, Education, and FTP Site. The central portion of the page offers illustrated and frequently updated information about new services and topics in the news. A right-hand column of Hot Spots highlights additional selected resources.

For a comprehensive, at-a-glance view of NCBI resources, follow the **Site Map** link. The Site Map contains an alphabetical listing of all services in tabular form, brief descriptions of each, and an expanded table of contents.

Take some time to explore. Under **About NCBI,** you will find our conference exhibit schedule and *NCBI News* online. The **GenBank** page includes a sample record in the database overview. The **Molecular Databases** page groups the sequence, mapping, structure, and taxonomy databases built from direct submissions. **Genomic Biology** pulls together resources for compiling and analyzing genomic-scale data, such as UniGene, GeneMap '99, RefSeq, and COGs. Under **Tools**, you will find BLAST, Electronic PCR, ORF Finder, and other analysis software. The **Education** link leads to online tutorials, sample exercises, overview articles, and other teaching resources.

The general organization of the new home page has also been extended to most of the other NCBI pages in the site hierarchy, allowing for fluid navigation among related resources.
—BR, DW

DEPARTMENT OF HEALTH AND HUMAN SERVICES
Public Health Service, National Institutes of Health
National Library of Medicine
National Center for Biotechnology Information
Bldg. 38A, Room 8N-803
8600 Rockville Pike
Bethesda, Maryland 20894