

# Биологични бази данни – принципи на работа Изабел Ангова F 57273



# Нов Български Университет

Изготвил: Изабел Ангова F57273

Специалност: Приложна биология II курс

Курс: BIOV007 История на биомедицинската  
кибернетика

Преподавател: гл. ас. д-р Полина Михова



# Съдържание

- **Какво са биологични бази данни?**
- **Какво съдържат?**
- **За какво се използват?**
- **За какво се използват?**
- **Една от първите бази данни**
- **Как работят базите данни?**
- **Какво прави всяка една функция на “ИСКА”?**
- **Идентификатор**
- **Идентификационен код**
- **Каква е разликата между идентификатора и идентификационния код?**



# Какво са биологични бази данни?

Биологичните бази данни са множество библиотеки на информация за науките за живота, които са събрани от научни експерименти, публикувана литература, високо пропускателна екпериментна технология и изчислителни анализи.



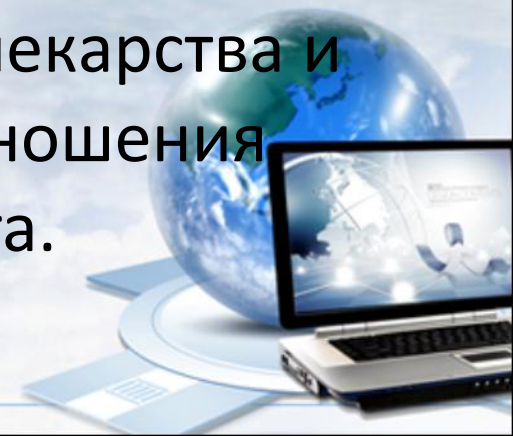
## Какво съдържат?

- Те съдържат информация от изследователски области, включително геномика, протеомика, метаболомика, микрорлъчава генна експресия и филогенетика.
- Информацията, съдържаща се в биологичните бази данни включва:  
генфукция, структура, локализация (клетъчна и хромозмна), клинични ефекти на мутации, а също така и прилики на биологичните секвенции и структури.



## За какво се използват?

Биологичните бази данни са важен инструмент, който помага на учените да разберат и да обяснят множеството биологични явления, от структурата на биомолекулите и тяхното взаимодействие до целият метаболизъм на организмите и разбирането на еволюцията на видовете. Този метод помага да се улесни борбата срещу болестите, разработването на лекарства и отикриването на основните взаимоотношения между видовете в историята на живота.



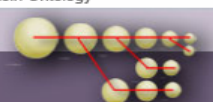
## Една от първите бази данни


- Една от първите последователни биологични бази данни е книгата "Атлас на протеиновите секвенции и структури" от Маргарет Дайхоф. Тя е публикувана за първи път през 1965 година. Съдържа протеинови последователности, които са определени по това време. Книгата е обновявана до 1970 година, след това тя се превръща в основа за базата данни на PIR.



**INTEGRATED PROTEIN INFORMATICS RESOURCE FOR  
GENOMIC, PROTEOMIC AND SYSTEMS BIOLOGY RESEARCH**

**UniProt**  
The Universal Protein Resource (UniProt) provides the scientific community with a single, centralized, authoritative resource for protein sequences and functional information.  
UniProtKB | UniRef | UniParc Current release: 2012\_03

**PRO**  
Protein Ontology  
  
▪ Representation of protein objects with descriptions and relationships  
▪ [Browse PRO](#)  
▪ Annotate with [RACE-PRO](#)  
  
\*[Sample PRO report](#)\*

**iProClass**  
Integrated Protein Knowledgebase  
  
▪ Value-added reports for [UniProtKB](#) and unique [UniParc](#) proteins  
▪ Functional analysis and [protein ID mapping](#)  
  
\*[Sample protein report](#)\*

**iProLINK**  
Literature Information & Knowledge  
  
▪ Source for text mining and ontology development  
▪ [RLIMS-P](#) text mining tool, [BioThesaurus](#)  
▪ [Bibliography mapping](#)  
  
\*[Sample Biblio. report](#)\*

**O OTHER RESOURCE**  
▪ [Representative Proteomes](#)  
▪ [PIR Grid-Enablement](#): Data node on NCI's [caBIG](#)

**P PEPTIDE SEARCH** ?  
DATABASE: UniProtKB

**T TEXT SEARCH** ?  
DATABASE: iProClass

**NEW** \* [PIR is hosting the Fifth International Biocuration Conference April 2-4, 2012](#)

\* **Bioinformatics & Computational Biology Graduate Programs:**  
[MS program at Georgetown University](#)  
[MS, PSM and Graduate Certificate programs at University of Delaware](#)  
▪ *Now recruiting new students for Fall of 2012*  
▪ *Scholarships available*



## Как работят базите данни?

Биологичните бази данни са създадени на принципа “ИСКА”, или Импортиране, Сливане, Класифициране и Анотация.”ИСКА” обикновено се прилага за всички бази данни изключение на анотацията, която се прилага в биологичните бази данни. Това е термин чрез който се прикачва биологична функция на микро молекули.



## Какво прави всяка една функция на “ИСКА”?

Импортирането е колекция от информация от ресурсите. Сливането е правилна подредба на бази данните в файл на Excell или с помощта на друг вид компютърни средства. Класификацията се използва за задаване на критерии, с помощта на които след това се търси в бази данните. Анотацията, както споменахме преди малко, се използва само в биологичните бази данни за прикачване на биологични функции.



Биологичните данни се предлагат в множество формати. Тези формати включват текст, последователни данни, протеинови структури и връзки. Всяка от тях може да се намери от определени източници, например :

- Текстови формати са осигурени от PubMed и OMIM
- Последователни данни се предоставят от GenBank по отношение на ДНК, и UniProt, по отношение на белтъчини
- Протеинови структури са осигурени от PDB, SCOP, CATH





## PubMed

PubMed comprises more than 21 million citations for biomedical literature from MEDLINE, life science journals, and online books. Citations may include links to full-text content from PubMed Central and publisher web sites.

### Using PubMed

[PubMed Quick Start Guide](#)

[Full Text Articles](#)

[PubMed FAQs](#)

[PubMed Tutorials](#)

[New and Noteworthy](#)

### PubMed Tools

[PubMed Mobile](#)

[Single Citation Matcher](#)

[Batch Citation Matcher](#)

[Clinical Queries](#)

[Topic-Specific Queries](#)

### More Resources

[MeSH Database](#)

[Journals in NCBI Databases](#)

[Clinical Trials](#)

[E-Utilities](#)

[LinkOut](#)



You are here: NCBI > Literature > PubMed

[Write to the Help Desk](#)

#### GETTING STARTED

- [NCBI Education](#)
- [NCBI Help Manual](#)
- [NCBI Handbook](#)
- [Training & Tutorials](#)

#### RESOURCES

- [Chemicals & Bioassays](#)
- [Data & Software](#)
- [DNA & RNA](#)
- [Domains & Structures](#)
- [Genes & Expression](#)
- [Genetics & Medicine](#)
- [Genomes & Maps](#)
- [Homology](#)
- [Literature](#)
- [Proteins](#)
- [Sequence Analysis](#)
- [Taxonomy](#)
- [Training & Tutorials](#)
- [Variation](#)

#### POPULAR

- [PubMed](#)
- [Nucleotide](#)
- [BLAST](#)
- [PubMed Central](#)
- [Gene](#)
- [Bookshelf](#)
- [Protein](#)
- [OMIM](#)
- [Genome](#)
- [SNP](#)
- [Structure](#)

#### FEATURED

- [GenBank](#)
- [Reference Sequences](#)
- [Map Viewer](#)
- [Genome Projects](#)
- [Human Genome](#)
- [Mouse Genome](#)
- [Influenza Virus](#)
- [Primer-BLAST](#)
- [Sequence Read Archive](#)

#### NCBI INFORMATION

- [About NCBI](#)
- [Research at NCBI](#)
- [NCBI Newsletter](#)
- [NCBI FTP Site](#)
- [NCBI on Facebook](#)
- [NCBI on Twitter](#)
- [NCBI on YouTube](#)



All Databases

Search

- NCBI Home
- Resource List (A-Z)
- All Resources
- Chemicals & Bioassays
- Data & Software
- DNA & RNA
- Domains & Structures
- Genes & Expression
- Genetics & Medicine
- Genomes & Maps
- Homology
- Literature
- Proteins
- Sequence Analysis
- Taxonomy
- Training & Tutorials
- Variation

## Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

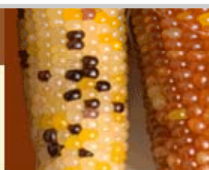
[About the NCBI](#) | [Mission](#) | [Organization](#) | [Research](#) | [RSS Feeds](#)

### Get Started

- [Tools](#): Analyze data using NCBI software
- [Downloads](#): Get NCBI data or software
- [How-To's](#): Learn how to accomplish specific tasks at NCBI
- [Submissions](#): Submit data to GenBank or other NCBI databases

### Genomic Structural Variation

dbVar archives large scale genomic variation data and associates defined variants with phenotypic information.



1 2 3 4 5 6 7 8

### Popular Resources

- PubMed
- Bookshelf
- PubMed Central
- PubMed Health
- BLAST
- Nucleotide
- Genome
- SNP
- Gene
- Protein
- PubChem

### NCBI Announcements

- BLAST+ v.2.2.26 is now available**  
08 Mar 2012  
The new BLAST+ release contains a number of important changes and improvements, including the addition of
- NIH/NCBI launches the Genetic Testing Registry (GTR)**  
29 Feb 2012  
The GTR is a new resource for finding information about genetic tests and test
- NCBI Discovery Workshop: A Practical Hands-On Course**  
24 Jan 2012  
February 21-22, 2012 @ the NIH: Space is still available in the 2-day Discovery

[More...](#)

You are here: NCBI > National Center for Biotechnology Information

[Write to the Help Desk](#)

GETTING STARTED  
NCBI Education

RESOURCES  
Chemicals & Bioassays

POPULAR  
PubMed  
Nucleotide

FEATURED  
GenBank  
Reference Sequences

NCBI INFORMATION  
About NCBI  
Research at NCBI

## THE FIRST 10 YEARS OF UNIPROT

### Saturday 8 September 2012, Basel, Switzerland

Celebrate UniProt's 10th anniversary with us! Renowned speakers from the fields of interactions & protein modelling, proteomics, protein structure & function, and genome analysis & annotation will highlight how protein databases are underpinning life sciences.

Submit abstracts for talks and posters to [abstractsUP12@isb-sib.ch](http://abstractsUP12@isb-sib.ch)

## WELCOME

The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

### What we provide

UniProtKB	<p>Protein knowledgebase, consists of two sections:</p> <ul style="list-style-type: none"> <li>★ Swiss-Prot, which is manually annotated and reviewed.</li> <li>★ TrEMBL, which is automatically annotated and is <b>not</b> reviewed.</li> </ul> <p>Includes <a href="#">complete and reference proteome sets</a>.</p>
UniRef	Sequence clusters, used to speed up sequence similarity searches.
UniParc	Sequence archive, used to keep track of sequences and their identifiers.
Supporting data	<a href="#">Literature citations</a> , <a href="#">taxonomy</a> , <a href="#">keywords</a> , <a href="#">subcellular locations</a> , <a href="#">cross-referenced databases</a> and more.

### Getting started

- [Text search](#)
- [Sequence similarity searches \(BLAST\)](#)
- [Sequence alignments](#)

## NEWS

### UniProt release 2012\_03 - Mar 21, 2012

The importance of being manual | Cross-references to DNASU

- › [Statistics for UniProtKB:](#)  
[Swiss-Prot](#) · [TrEMBL](#)
- › [Forthcoming changes](#)
- › [News archives](#)

Follow @uniprot 268 followers

## SITE TOUR



Learn how to make best use of the tools and data on this site.

## PROTEIN SPOTLIGHT

### the ends of our fingers March 2012

Fingertips are hugely sensitive. And, besides the fairly recent mobile phones that rely on them entirely, we put their sensitivity to use constantly...



- Access the PDB FTP:
  - RCSB PDB | PDBe | PDBj
  - Archive Download
  - Chemical Component Dictionary
- Deposit Data to the PDB:
  - RCSB PDB | PDBe
  - PDBj | BMRB
- Search for Structures:
  - RCSB PDB | PDBe
  - PDBj | BMRB
- PDB Archive Snapshots:
  - RCSB PDB | PDBj
- Instructions to Journals
- Documentation
  - Format
  - Annotation and Policies
  - Remediation
- Validation Reports
- Workshops and Task Forces
  - X-ray Validation
  - NMR Validation
  - EM Validation
- PDB40
- wwPDBAC
- EM

The Worldwide Protein Data Bank (wwPDB) consists of organizations that act as deposition, data processing and distribution centers for PDB data.<sup>1</sup> Members are: RCSB PDB (USA), PDBe (Europe) and PDBj (Japan), and BMRB (USA). The wwPDB's mission is to maintain a single PDB archive of macromolecular structural data that is freely and publicly available to the global community.

**wwPDB Statement on Retraction of PDB Entries**

**16-March-2012**  
**wwPDB Milestones: 80,000 Entries Released, PDB40 Meeting Report Published**  
*The Archive Grows*  
 With the March 13, 2012 update, 80,041 PDB entries became available in the archive. The PDB hit the 70,000 mark in December 2010.  
 In 2011, 8865 entries were deposited and annotated, while 8107 were released.  
 wwPDB deposition and download statistics are updated regularly.  
*Meeting Report*  
 A review of the PDB40 meeting has been published:  
 The Protein Data Bank at 40: Reflecting on the Past to Prepare for the Future Helen M. Berman, Gerard J. Kleywegt, Haruki Nakamura, John L. Markley *Structure (2012) 20: 391 - 396*  
 Selected presentations and related details from the meeting are also available from <http://www.wwpdb.org/PDB40.html>.

**6-March-2012**  
**EM Data Bank joins the PDB archive**  
 The EM Data Bank (EMDB), the primary archive for experimentally-determined maps obtained using three-dimensional electron microscopy methods, has joined the PDB archive (<ftp://ftp.wwpdb.org>), as announced previously. **MORE**



**8-February-2012**  
**A vision for validation of 3DEM maps and map-derived models**

[Home](#)

## Welcome to CATH

CATH is a manually curated classification of protein domain structures. Each protein has been chopped into structural domains and assigned into homologous superfamilies (groups of domains that are related by evolution). This classification procedure uses a combination of automated and manual techniques which include computational algorithms, empirical and statistical evidence, literature review and expert analysis.

[Search the CATH database >>](#)

[Find out more about CATH >>](#)

## New in CATH v3.4

CATH v3.4 is built from 104,238 PDB chains. We have added the following data since v3.3:

- 49 folds (total 1,282)
- 163 superfamilies (total 2,549)
- 1,311 sequence families (total 11,330)
- 24,232 domains (total 152,920)

[Download CATH data >>](#)

## New in Gene3D 10.2

Gene3D 10.2 (released Sep 2011) uses CATH domains to provide 16,118,154 structural annotations for 14,963,305 protein sequences. The latest release also offers a number of new features:

- Interaction Network Data
- Interactive Graphical Representations
- Genome Comparisons

[Goto Gene3D >>](#)

## Using CATH

- [Search](#)
- [Browse](#)
- [Download](#)
- [Tutorials](#)
- [Introduction to CATH](#)
- [FAQ](#)

## CATH Tools

- [Find My Sequence](#)
- [Find My Structure](#)
- [CATHEDRAL Structure Scan](#)
- [Structure Comparison Server](#)
- [Linking to CATH](#)

## About CATH

- [Release Statistics](#)
- [Glossary](#)
- [CATH Team](#)
- [References](#)



Много бази данни в биоинформатиката (SWISS-Prot, EMBL, GenBank, Pfam), се използва система, при която един запис може да бъде намерен по два различни начина:

- Идентификатор
- Идентификационен код



# Идентификатор

Идентификатора (“locus” в GenBank и “entry name” в SWISS-PROT) е наниз от думи и цифри, които като цяло са интерпретирани по някакъв смислен начин за човека, например като познато съкращение на думата протеин или ген.



SWISS-PROT използва система, при която името на записа (entry name) се състои от две части:

първата обозначава протеина и втората част обозначава вида, в който е намерен

Например:

KRAF\_HUMAN е името на записа за Raf-1 oncogene от Homo sapiens.



Идентификаторите понякога може да се променят. Например:

Администраторите на базата данни могат да решат, че идентификаторът за вписване вече не е подходящ. И тогава се задава нов и става по трудно за намирането на даден запис. Това е и минуса на идентификатора.



# Идентификационен код

Идентификационният код е число, което по уникален начин съответства с точно определен запис в бази данните.

Например:

Идентификационния номер за KRAUF\_HUMAN в SWISS-PROT е P04049



В случай, че два записа се съединят в един, той получава двата идентификационни кода на предишните записи като единия става основен, а другия второстепенен. Има случаи, в които един запис се разделя на два. В този случай и двата записа получават по един нов код, но и запазват старите си идентификационни кода като второстепенни.



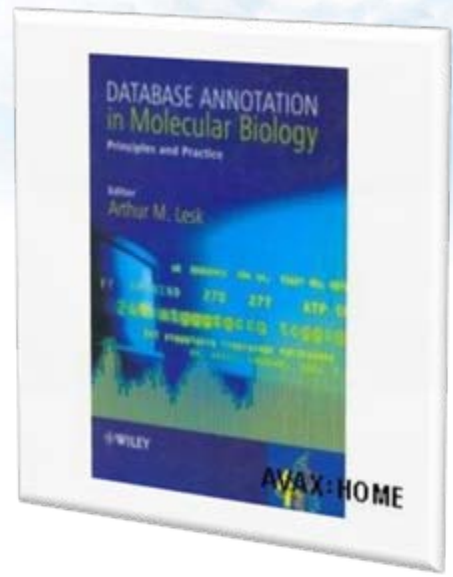
## Каква е разликата между идентификатора и идентификационния код?

Разликата между идентификатора и идентификационния код е ,че кодът се задава веднъж и си остава завинаги дори и след голяма промяна на информацията на даден запис в бази данните, а както споменахме преди малко идентификатора бива сменян след промяна на информацията на даден запис.





BIOLOGICVF DV1VBV2E2  
BIOBASE





## BioLOD.org

The BioLOD.org project is supported by the National Bioscience Database Center (NBDC) of Japan Science and Technology Agency (JST), and RIKEN.

Submitted and archived to BioLOD.org



Download and apply to various research projects



Research



# ИЗПОЛЗВАНИ ИЗТОЧНИЦИ

1. [http://en.wikipedia.org/wiki/Biological\\_database#Working\\_of\\_Database](http://en.wikipedia.org/wiki/Biological_database#Working_of_Database)
2. <http://www.avatar.se/strbio2001/databases/acc.html>
3. <http://www.nbu.bg/PUBLIC/IMAGES/File/departamenti/medikobiologichni/Bio.Aktivni.pdf>



Благодаря за вниманието!!! 😊

