

Базы данных в протеомике

Типы баз данных

По характеру заполнения информацией

- **Архивные** (*GeneBank & EMBL, PDB*),
- **Курируемые** (*Swiss-Prot, KEGG, FlyBase, COG*),
- **Производные** (*SCOP, PFAM, GO, ProDom, AsMamDB*),
- **Интегрированные** (*NCBI Entrez, Ecocyc*).

Архивные базы данных представляют собой **неупорядоченный** или **малоупорядоченный** массив информации, который формируется свободным образом, то есть любой пользователь может поместить свои данные в такие базы. Очень часто в архивных базах данных можно встретить **ошибочные данные**. Поэтому к информации, полученной из таких баз данных, нужно относиться с большой осторожностью и обязательно подвергать дополнительной проверке. В качестве примеров архивных баз данных, можно привести следующие: **GeneBank & EMBL** (база данных первичных нуклеотидных последовательностей) и **PDB** (база данных пространственных структур белков).

Protein Data Bank, PDB — банк данных 3-D структур белков и нуклеиновых кислот. Информация, полученная методами рентгеновской кристаллографии или ЯМР-спектроскопии, вносится в базу данных биологами и биохимиками со всего мира, и доступна бесплатно через интернет.

PDB является один из важнейших ресурсов для учёных, работающих в области структурной биологии. Большинство научных журналов и некоторые фонды финансирования исследований, например, NIH в США требуют от авторов статей и получателей грантов, чтобы все структурные данные были размещены в PDB. Protein Data Bank содержит, в основном, первичные данные о структуре биологических молекул, в то время как существуют сотни других банков данных, категоризирующих первичные данные или выявляющие закономерности между строением молекул и эволюционным родством.

Структура PDB



История создания PDB

Dr. Walter Hamilton



Protein Data Bank был создан учёными. В 1971 году, Уолтер Хэмилтон в Национальной лаборатории Брукхавена (*Brookhaven National Laboratory*) создал банк данных для Брукхавена. После смерти Хэмилтона в 1973 году, PDB управлял Том Кэцтл.

В январе 1994 года главой Protein Data Bank стала Джол Суссман.

В октябре 1998 года Protein Data Bank был перенесён в Research Collaboratory for Structural Bioinformatics (RCSB); перенос информации был закончен в июне 1999 года. Новым директором стала Хелен Берман из Университета Рутгерса.

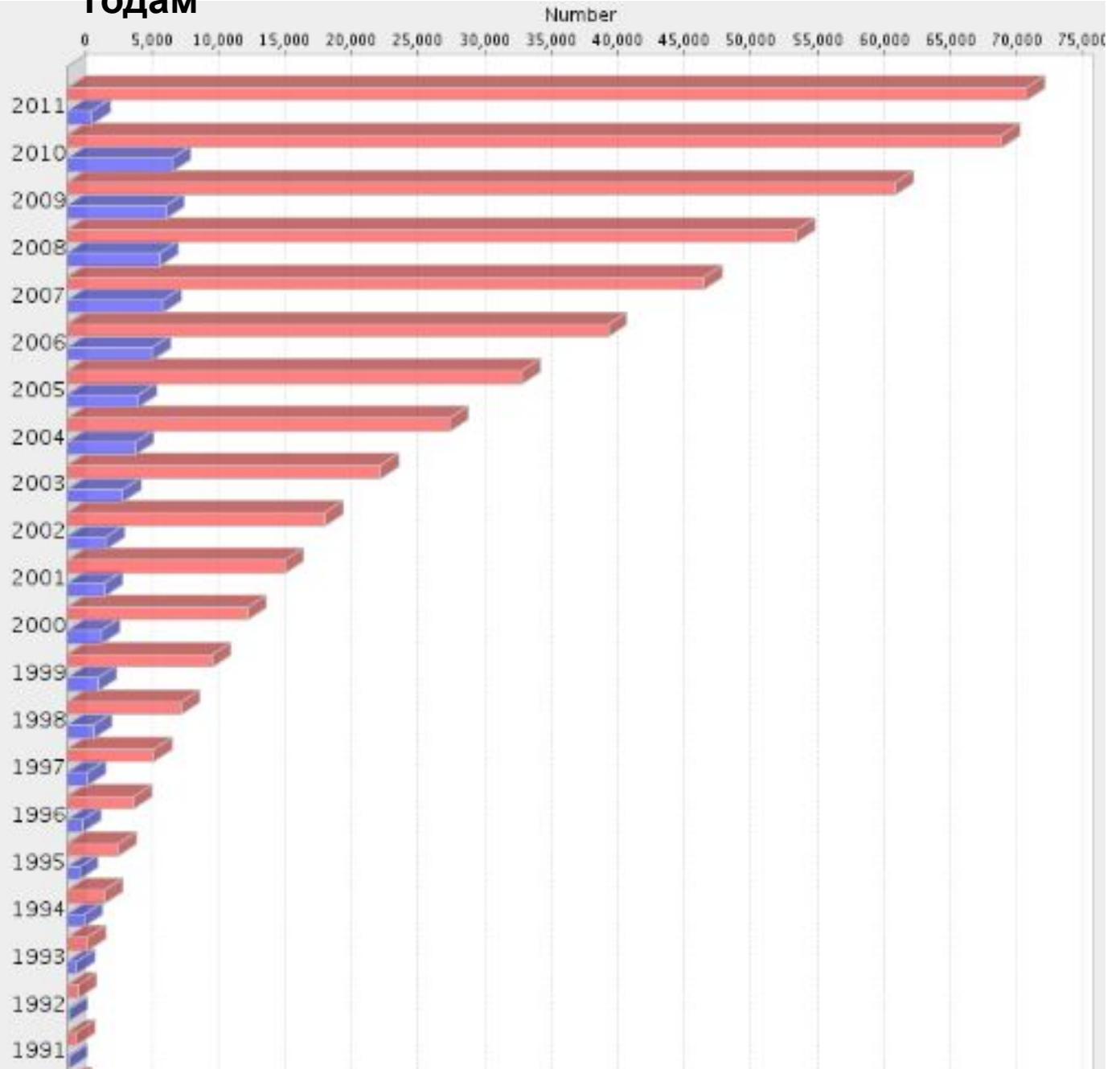
В 2003 году, после образования wwPDB, Protein Data Bank стал международной организацией.

BROOKHAVEN
NATIONAL LABORATORY

PDB по состоянию на 22.03.11

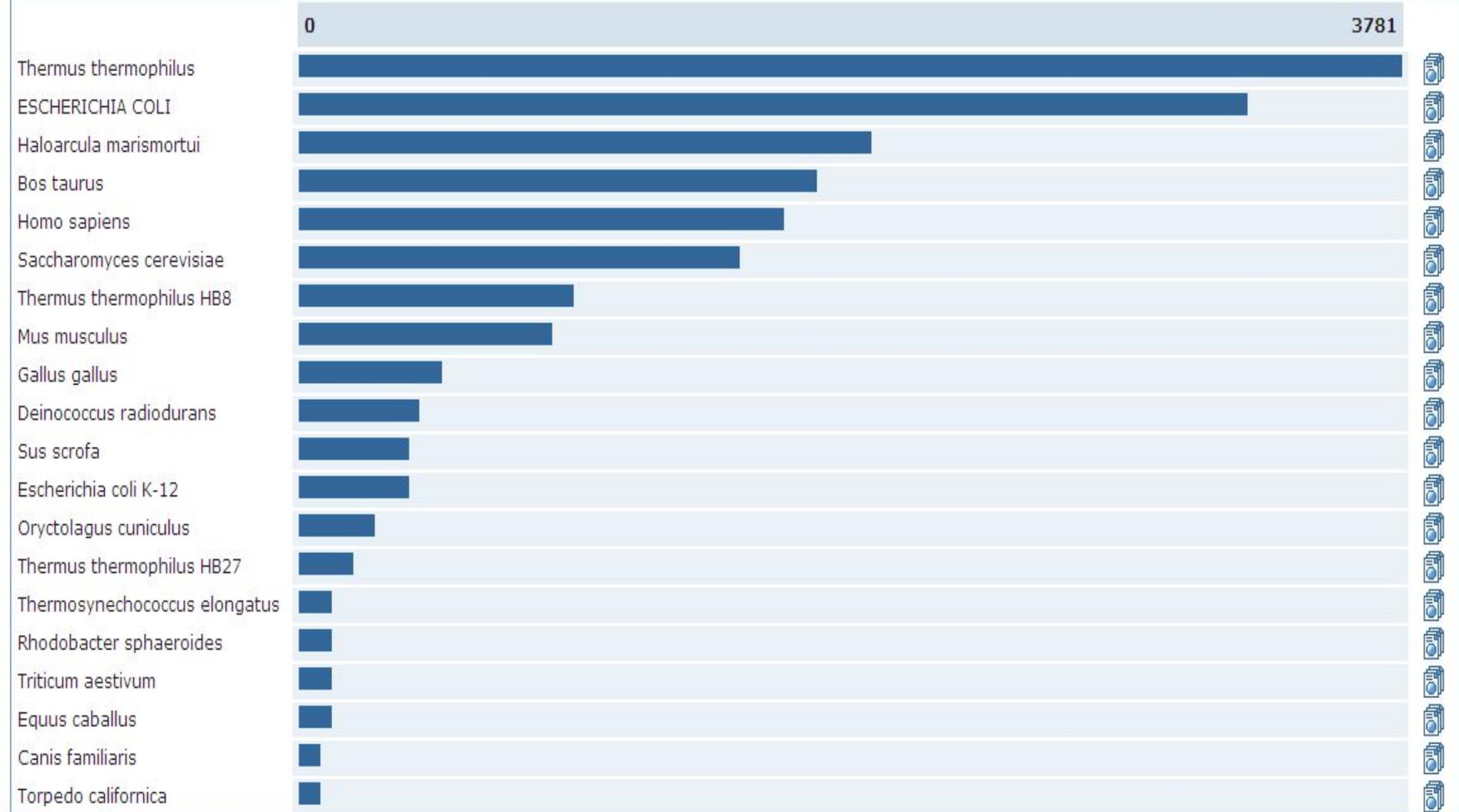
Экспериментальный метод	Белки	Нуклеиновые кислоты	Комплексы белок/НК	Другое	Всего
Дифракция рентгеновских лучей	58478	1262	2836	17	62593
ЯМР	7694	943	169	7	8813
Электронная микроскопия	249	22	91	0	362
Смешанный	28	3	1	1	33
Другое	132	4	5	13	154
Всего	66581	2234	3102	38	71995

Накопление структур в базе данных по годам



Определено структур у различных организмов

Total Count (not null): 18844



PDB ID

Каждая структура, опубликованная в PDB получает **четырёхзначный идентификатор** (комбинация цифр и букв латинского алфавита). Данный шифр не может служить идентификатором биомолекул, так как часто разные структуры одной и той же молекулы, например, в различной среде, могут иметь различные PDB ID.

3CZH **Crystal structure of CYP2R1 in complex with vitamin D2**

Authors: Strushkevich, N.V. , Tempel, W. , Gilep, A.A. , Loppnau, P. , Arrowsmith, C. , Bochkarev, A. , Park, H. , Structural Genomics Consortium (SGC) 

Release Date: 2008-08-05 **Classification:**

Experiment: X-RAY DIFFRACTION with resolution of 2.30 Å

Compound: 1 Polymer [[Display Full Polymer Details](#) | [Display for All Results](#)]
4 Ligands [[Display Full Ligand Details](#) | [Display for All Results](#)]

Citation: Not Available.

3DL9 **Crystal structure of CYP2R1 in complex with 1-alpha-hydroxy-vitamin D2**

Authors: Strushkevich, N.V. , Tempel, W. , Gilep, A.A. , Loppnau, P. , Arrowsmith, C. , Bochkarev, A. , Park, H. , Structural Genomics Consortium (SGC) 

Release Date: 2008-08-05 **Classification:**

Experiment: X-RAY DIFFRACTION with resolution of 2.72 Å

Compound: 1 Polymer [[Display Full Polymer Details](#) | [Display for All Results](#)]
3 Ligands [[Display Full Ligand Details](#) | [Display for All Results](#)]

Citation: Not Available.

Максимально возможное число молекулярных структур, описываемых четырьмя символами по определенным на сегодняшний день правилам:

$$10 \cdot (26 + 10) \cdot (26 + 10) \cdot (26 + 10) = 466560$$

Форма хранения информации – PDB-файл

```
HEADER          OXIDOREDUCTASE                      04-FEB-08   3C6G
TITLE          CRYSTAL STRUCTURE OF CYP2R1 IN COMPLEX WITH VITAMIN D3
COMPND        MOL_ID: 1;
COMPND        2 MOLECULE: CYTOCHROME P450 2R1;
COMPND        3 CHAIN: A, B;
COMPND        4 FRAGMENT: RESIDUES 32-501;
COMPND        5 SYNONYM: VITAMIN D 25-HYDROXYLASE;
COMPND        6 EC: 1.14.14.-;
COMPND        7 ENGINEERED: YES
SOURCE        MOL_ID: 1;
SOURCE        2 ORGANISM_SCIENTIFIC: HOMO SAPIENS;
SOURCE        3 ORGANISM_COMMON: HUMAN;
SOURCE        4 ORGANISM_TAXID: 9606;
SOURCE        5 GENE: CYP2R1;
SOURCE        6 EXPRESSION_SYSTEM: ESCHERICHIA COLI;
SOURCE        7 EXPRESSION_SYSTEM_TAXID: 562;
SOURCE        8 EXPRESSION_SYSTEM_STRAIN: JM109;
SOURCE        9 EXPRESSION_SYSTEM_VECTOR_TYPE: PLASMID;
SOURCE        10 EXPRESSION_SYSTEM_PLASMID: PCW
KEYWDS        CYTOCHROME P450, VITAMIN D3, VITAMIN D 25-HYDROXYLASE, DRU
KEYWDS        2 METABOLISM, STRUCTURAL GENOMICS, STRUCTURAL GENOMICS
KEYWDS        3 CONSORTIUM, SGC, DISEASE MUTATION, ENDOPLASMIC RETICULUM,
KEYWDS        4 HEME, IRON, MEMBRANE, METAL-BINDING, MICROSOME,
KEYWDS        5 MONOOXYGENASE, OXIDOREDUCTASE
EXPDTA        X-RAY DIFFRACTION
AUTHOR        N.V.STRUSHKEVICH,J.MIN,P.LOPPNAU,W.TEMPEL,C.H.ARROWSMITH,
AUTHOR        2 A.M.EDWARDS,H.SUNDSTROM,J.WEIGELT,A.BOCHKAREV,A.N.PLOTNIK
AUTHOR        3 H.PARK,STRUCTURAL GENOMICS CONSORTIUM (SGC)
REVDAT        3 24-FEB-09 3C6G 1 UERSN
REVDAT        2 01-JUL-08 3C6G 1 JRNL
REVDAT        1 19-FEB-08 3C6G 0
SPRSDE        19-FEB-08 3C6G 20JD
JRNL          AUTH  N.STRUSHKEVICH,S.A.USANOV,A.N.PLOTNIKOV,G.JONES,
JRNL          AUTH 2 H.W.PARK
JRNL          TITL  STRUCTURAL ANALYSIS OF CYP2R1 IN COMPLEX WITH
JRNL          TITL 2 VITAMIN D3.
JRNL          REF   J.MOL.BIOL.                      U. 380   95 2008
JRNL          REFN                      ISSN 0022-2836
JRNL          PMID  18511070
JRNL          DOI   10.1016/J.JMB.2008.03.065
REMARK        1
REMARK        2
REMARK        2 RESOLUTION.      2.80 ANGSTROMS.
```

Molecule of the Month Archive

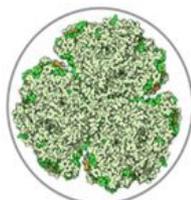
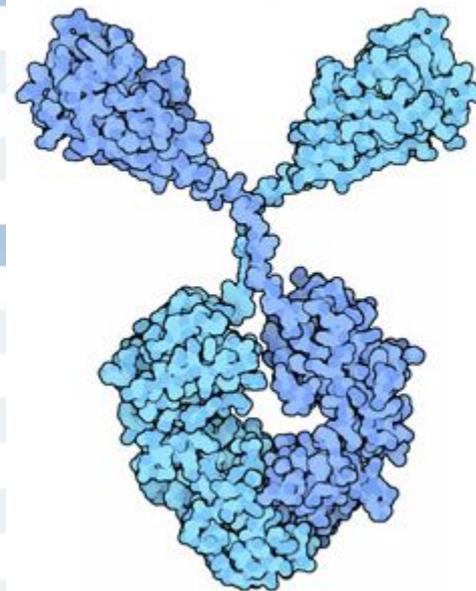
2011 | 2010 | 2009 | 2008 | 2007 | 2006

2011

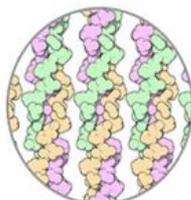
- April ▶ Nanobodies
- March ▶ Integrase
- February ▶ Integrin
- January ▶ Nitric Oxide Synthase

2010

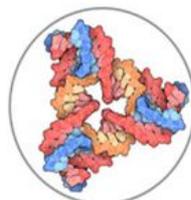
- December ▶ Adenovirus
- November ▶ Inteins
- October ▶ Riboswitches
- September ▶ Isocitrate Dehydrogenase (PDF 2.8 MB)
- August ▶ Interferons (PDF 2.3 MB)
- July ▶ Crystallins (PDF 1.8 MB)
- June ▶ Epidermal Growth Factor (PDF 2.1 MB)
- May ▶ Parvoviruses (PDF 3.6 MB)
- April ▶ Concanavalin A and Circular Permutation (PDF 2.2 MB)
- March ▶ P-glycoprotein (PDF 2.8 MB)
- February ▶ Enhanceosome (PDF 2.1 MB)
- January ▶ 70S Ribosomes (PDF 2.9 MB)



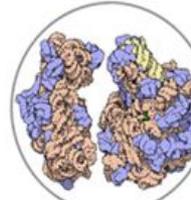
Biological Energy



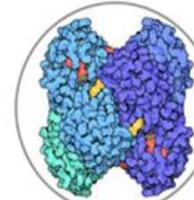
Infrastructure and Communication



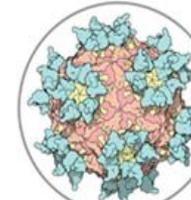
Biotechnology and Nanotechnology



Protein Synthesis



Enzymes



Health and Disease

Курируемые базы данных обязательно **контролируются** организаторами или владельцами сайта, постоянно подвергаются проверке со стороны кураторов или аннотаторов. Информация в таких базах отбирается специальными экспертами из архивных баз данных, предварительно проверяя ее достоверность, экспериментальные доказательства и т.д. Попадание случайной информации в такие базы данных почти исключено. К наиболее известным курируемым базам данных относятся: **Swiss-Prot** (наиболее качественная база данных, содержащая аминокислотные последовательности белков), **KEGG** (база данных различных метаболических путей), **FlyBase** (посвящена *Drosophila*), **COG** (база данных

Банк данных Swiss-Prot

1986



Swiss-Prot – база знаний о
белковых последовательностях

- Курируемая база данных
- “**Золотой стандарт**” аннотации

Банк данных Swiss-Prot



С 1987 поддерживается в сотрудничестве между

Swiss Institute of Bioinformatics (SIB)

European Bioinformatics Institute (EBI)

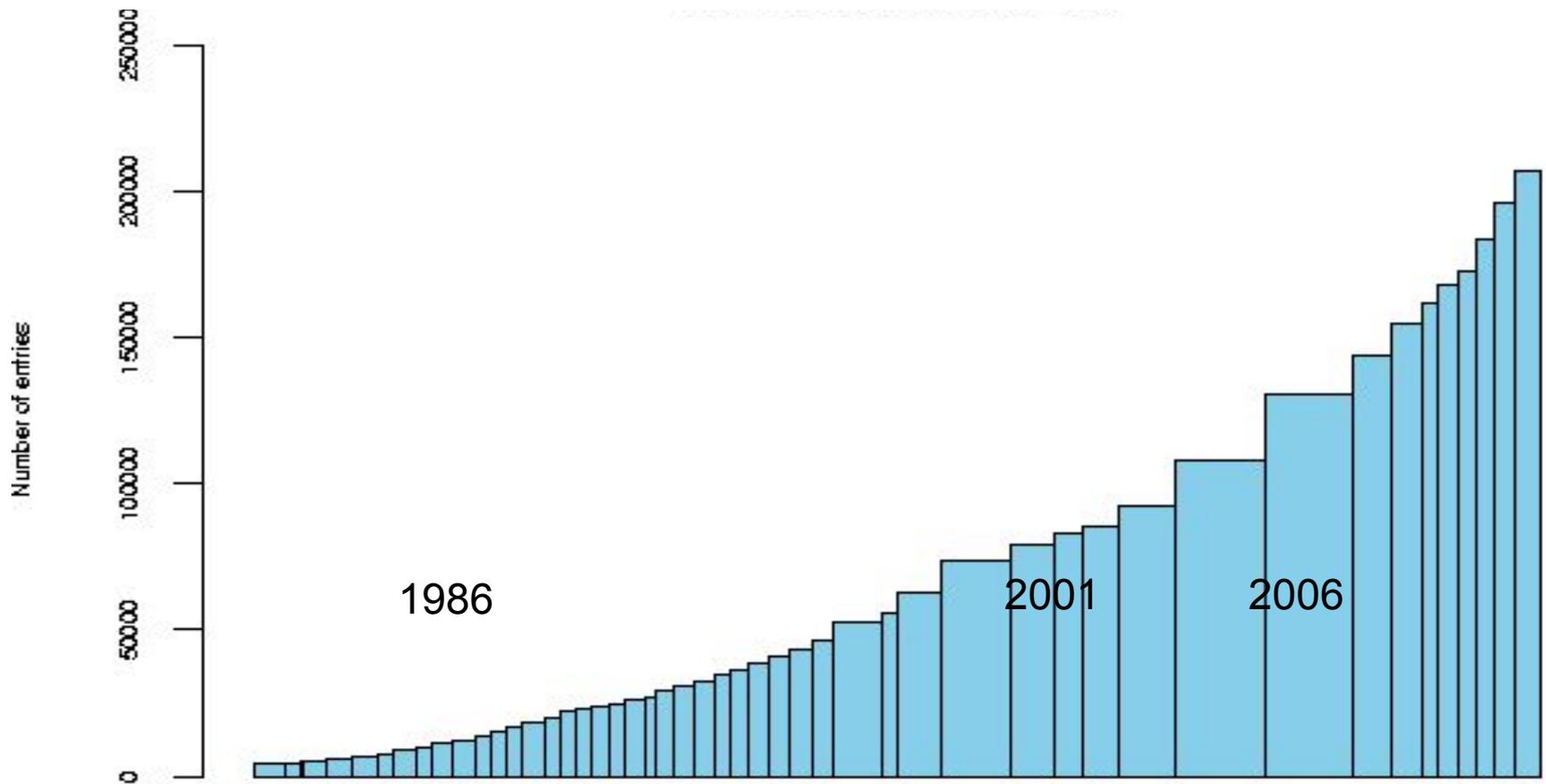


Амос Байрох

Руководитель группы Swiss-Prot в
Швейцарском Институте Биоинформатики

Банк данных Swiss-Prot

Статистика роста количества документов



Банк данных TrEMBL



TrEMBL (Tranlated **EMBL**)

Формальная трансляция всех кодирующих
нуклеотидных последовательностей из банка EMBL

Автоматическая классификация и аннотация

Релиз **31.9** (24 января 2006) содержит 2 586 884 документа

Тенденция объединения

2002



PIR Protein Information Resource



Банк данных UniProt



UniProt (Universal Protein Resource)

- UniProt Knowledgebase – **SwissProt+TrEMBL**
- UniProt Archive – **UniParc**
- UniProt Reference – **UniRef**

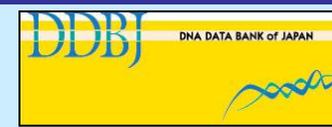
ttttacctcttttagtgatattgtgatagagcaaaaatcccgcacattgtgctcgggattgttttaaaccttgttgatttaattttcaatcgcttctttataaagaagtagtggtgccc
acaacactcacattgcatatcaatacggcctttatgttcggctaataatttcgcaatttcttcatcagagatgagcagtagatgcagaactagaacgctcagcagagcagccaca
gaaaaattgtacatcttgtgctggataaagattaacggtttctcgtgatataaacgataggagtaacttctcagggagaccaataattcttcatctttactgttgctgagc
gtagttaaatgctcaaaatcttctggtgtaccagaaccatcaggcataattgtaataacatacctgctgccactggctgccttcatattctccagtacgaataattaattgagttg



GenBank



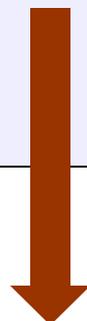
EMBL



DDBJ



компьютерный поиск гена, трансляция и компьютерная аннотация



Базы данных научной литературы



~2 500 000 последовательностей

Экспертиза



~200 000 последовательностей

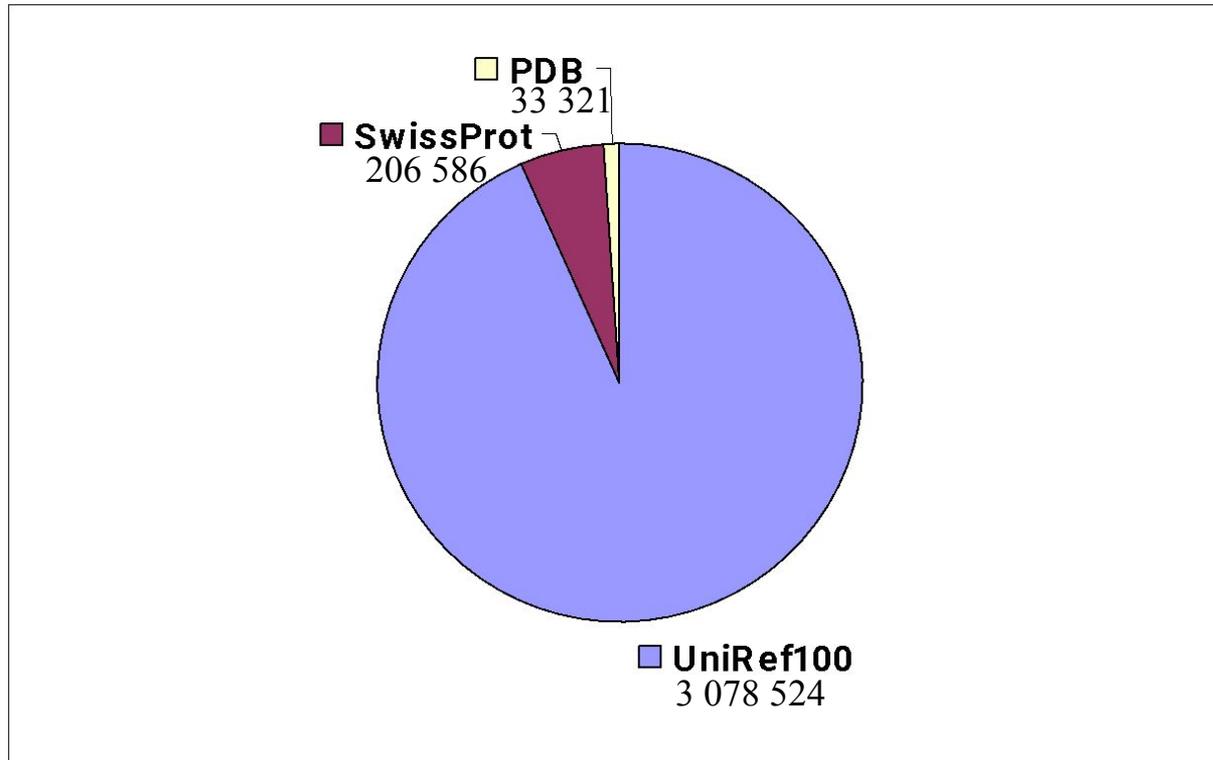


UniParc
(UniProt Archive)

UniRef
(UniProt non-redundant Reference databases)



Соотношение числа белков, представленных в разных банках



Последовательностей во много раз больше, чем структур!

Большинство последовательностей не аннотированы!

Документ банка данных Swiss-Prot

```
ID YSEA_STACA STANDARD; PRT; 165 AA.
AC P47995;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Hypothetical protein in secA 5' region (ORF1) (Fragment).
OS Staphylococcus carnosus.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1281;
RN [1]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RC STRAIN=TM300;
RA Freudl R.;
RL Submitted (JUN-1994) to the EMBL/GenBank/DDBJ databases.
CC -!- SIMILARITY: Belongs to the ribosomal protein S30Ae family.
CC -!- CAUTION: This is a conceptual translation.
CC -!- CAUTION: Ref.1 sequence differs from that shown due to frameshifts
CC in positions 25 and 46.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
DR EMBL; X79725; CAA56161.1; ALT_FRAME; Genomic_DNA.
DR PIR; S47148; S47148.
DR InterPro; IPRO03489; Ribosomal_S30S54.
DR Pfam; PFO2482; Ribosomal_S30AE; 1.
KW Hypothetical protein.
FT NON_TER 1 1
SQ SEQUENCE 165 AA; 19138 MW; BF8CB91ADE194DDO CRC64;
LERYFTNVPN VNAHVKVKTY ANSSKIEVTI PLNDVTLRAE ERNDDIYAGI DKITNKLECG
VRKYKTRVNR KKRKESHEP FPATPETPPE TAVDHDKDDE IEIIRSKQFS LKPMDSSEAV
LQMDLLGTFD FIFNDRETDG TSIVYRRKDG KYGLIETVEK LICDI
```

Описание документа: идентификатор,
имя, дата создания и модификации

Аннотация
последовательности

Последовательность

Основные поля записи SwissProt

- ID
- AC
- DE
- OS
- OC

И сама последовательность, конечно.

UniProt

← → ↻ www.uniprot.org

UniProt

Search Blast Align Retrieve ID Mapping

Search in Protein Knowledgebase (UniProtKB) Query

Search Clear Advanced Search »

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	Protein knowledgebase, consists of two sections: <ul style="list-style-type: none">★ Swiss-Prot, which is manually annotated and reviewed.★ TrEMBL, which is automatically annotated and is not reviewed. Includes Complete Proteome Sets.
UniRef	Sequence clusters, used to speed up sequence similarity searches.
UniParc	Sequence archive, used to keep track of sequences and their identifiers.
Supporting data	Literature citations, taxonomy, keywords and more.

Getting started

- Text search
- Sequence similarity searches (BLAST)
- Sequence alignments
- Batch retrieval
- Database identifier mapping (ID Mapping)



NEWS

UniProt release 2011_05 - May 3, 2011

Complete proteome sets for Homo sapiens and Mus musculus

- > Statistics for UniProtKB:
 - Swiss-Prot · TrEMBL
- > Forthcoming changes
- > News archives

SITE TOUR



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

PROTEIN SPOTLIGHT

on the spur of a whim March 2011

There are a number of biological molecules which are involved in a bewildering amount of activities. Serotonin is one...

Производные базы данных образуются в результате обработки данных из архивных и курируемых баз данных. Например, **SCOP** (база данных структурной классификации белков), **PFAM** (семейства белков), **GO** (Gene Ontology) (классификация генов, одной из основных задач этой базы является упорядочивание терминологии названий генов), **ProDom** (база данных посвящена белковым доменам), **AsMamDB** (база по альтернативному сплайсингу млекопитающих).

SCOP

Structural Classification Of Proteins

← → ↻ scop.mrc-lmb.cam.ac.uk/scop/



Structural Classification of Proteins



Welcome to **SCOP**: Structural Classification of Proteins.

1.75 release (June 2009)

38221 PDB Entries. 1 Literature Reference. 110800 Domains. (excluding nucleic acids and theoretical models).

Folds, superfamilies, and families [statistics here](#).

[New folds](#) [superfamilies](#) [families](#).

[List of obsolete entries and their replacements](#).

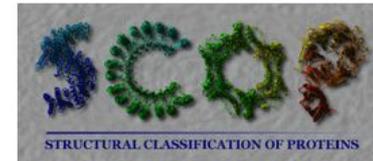
Authors: Alexey G. Murzin, John-Marc Chandonia, Antonina Andreeva, Dave Howorth, Loredana Lo Conte, Bartlett G. Ailey, Steven E. Brenner, Tim J. P. Hubbard, and Cyrus Chothia. scop@mrc-lmb.cam.ac.uk

Reference: Murzin A. G., Brenner S. E., Hubbard T., Chothia C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* 247, 536-540. [\[PDF\]](#)

Recent changes are described in: Lo Conte L., Brenner S. E., Hubbard T.J.P., Chothia C., Murzin A. (2002). SCOP database in 2002: refinements accommodate structural genomics. *Nucl. Acid Res.* 30(1), 264-267. [\[PDF\]](#).

Andreeva A., Howorth D., Brenner S.E., Hubbard T.J.P., Chothia C., Murzin A.G. (2004). SCOP database in 2004: refinements integrate structure and sequence family data. *Nucl. Acid Res.* 32:D226-D229. [\[PDF\]](#), and

Andreeva A., Howorth D., Chandonia J.-M., Brenner S.E., Hubbard T.J.P., Chothia C., Murzin A.G. (2007). Data growth and its impact on the SCOP database: new developments. *Nucl. Acids Res.* 2008 36: D419-D425; doi:10.1093/nar/gkm993 [\[PDF\]](#).



Access methods

- Enter **SCOP** at the [top of the hierarchy](#)
- [Keyword search of SCOP entries](#)
- [SCOP parseable files](#)
- [All SCOP releases and reclassified entry history](#)
- [pre-SCOP - preview of the next release](#)
- SCOP domain sequences and pdb-style coordinate files ([ASTRAL](#))
- Hidden Markov Model library for SCOP superfamilies ([SUPERFAMILY](#))
- Structural alignments for proteins with non-trivial relationships ([SISYPHUS](#))

- [Online resources](#) of potential interest to SCOP users

SCOP [mirrors](#) around the world may speed your access.

News

- SCOP has been updated to include many PDB entries released before 23 February 2009. See [folds, superfamilies, and families statistics](#).
- This release no longer classifies all PDB structures released before a certain date. The process of classification of new entries has been changed. For more information please visit [pre-SCOP](#) - a preview of the next release.
- This release is similar in appearance to the previous release, so the generic [release notes](#) from that release still apply. Please read the notes; they contain more detailed explanations and examples of SCOP features.

- [Previous releases' news](#).

Synopsis

Nearly all proteins have structural similarities with other proteins and, in some of these cases, share a common evolutionary origin. The **SCOP** database, created by manual inspection and abetted by a battery of automated methods, aims to provide a detailed and comprehensive description of the structural and evolutionary relationships between all proteins whose structure is known. As such, it provides a broad survey of all known protein folds, detailed information about the close relatives of any particular protein, and a framework for future research and classification.

A more detailed [description](#) of the database is available. [Help](#) on using the database may be obtained on any screen by pressing the question mark button.

Class	Number of folds	Number of superfamilies	Number of families
All alpha proteins	284	507	871
All beta proteins	174	354	742
Alpha and beta proteins (a/b)	147	244	803
Alpha and beta proteins (a+b)	376	552	1055
Multi-domain proteins	66	66	89
Membrane and cell surface proteins	58	110	123
Small proteins	90	129	219
Total	1195	1962	3902

Интегрированные базы данных, представляют собой базы, в которых информация получена **из самых различных источников** (курируемые, архивные базы данных и другие). В таких базах данных в строку поиска можно ввести название белка, получить в максимальном объеме связанную с ним информацию (в каких организмах, тканях и клетках встречается, структура, аминокислотный состав и т.д.) в виде различных ссылок на те или иные базы данных, сайты и прочее. К таким базам данных относятся: **NCBI Entrez** (информация о нуклеотидных и аминокислотных последовательностях), **Ecosys** (база данных о *E. coli*).

NCBI Entrez



HOME SEARCH SITE MAP PubMed All Databases Human Genome GenBank Map Viewer BLAST

Search across databases [Help](#)

Welcome to the Entrez cross-database search page

- | | |
|---|---|
| PubMed: biomedical literature citations and abstracts | Books: online books |
| PubMed Central: free, full text journal articles | Images: images from full text resources at NCBI |
| Site Search: NCBI web and FTP sites | OMIM: online Mendelian Inheritance in Man |
| Nucleotide: Core subset of nucleotide sequence records | dbGaP: genotype and phenotype |
| EST: Expressed Sequence Tag records | UniGene: gene-oriented clusters of transcript sequences |
| GSS: Genome Survey Sequence records | CDD: conserved protein domain database |
| Protein: sequence database | UniSTS: markers and mapping data |
| Genome: whole genome sequences | PopSet: population study data sets |
| Structure: three-dimensional macromolecular structures | GEO Profiles: expression and molecular abundance profiles |
| Taxonomy: organisms in GenBank | GEO DataSets: experimental sets of GEO data |
| SNP: single nucleotide polymorphism | Epigenomics: Epigenetic maps and data sets |
| dbVar: Genomic structural variation | Cancer Chromosomes: cytogenetic databases |
| Gene: gene-centered information | PubChem BioAssay: bioactivity screens of chemical substances |
| SRA: Sequence Read Archive | PubChem Compound: unique small molecule chemical structures |
| BioSystems: Pathways and systems of interacting molecules | PubChem Substance: deposited chemical substance records |
| HomoloGene: eukaryotic homology groups | Protein Clusters: a collection of related protein sequences |
| GENSAT: gene expression atlas of mouse central nervous system | OMIA: online Mendelian Inheritance in Animals |
| Probe: sequence-specific reagents | BioSample: biological material descriptions |
| Genome Project: genome project information | |
| NLM Catalog: catalog of books, journals, and audiovisuals in the NLM collections | MeSH: detailed information about NLM's controlled vocabulary |

Entrez – глобальная поисковая система, позволяющая производить поиск одновременно по многим отдельным базам данных в NCBI путем ввода единственного запроса, работать с последовательностями, структурами и ссылками, визуализировать последовательности генов, белков, хромосомные карты.

Классификация баз данных по специализации

- Базы данных протеомов
- Базы данных белковых семейств
- Базы данных структур белковых молекул
- Базы данных белковых взаимодействий

Базы данных протеомов. Содержат информацию о **протеоме** или какого-нибудь **организма** (человек, мышь, дрозофила, дрожжи и т.д.), или о протеоме каких-то определенных типов **органелл, клеток, тканей** (протеом ядра, протеом мышечных клеток, нервной ткани и т.д.).

Базы данных **белковых семейств**, в основу которых положена классификация белков в семейства, суперсемейства, кланы и т.д. Есть универсальные базы данных, которые собирают информацию по всем семействам, а есть специализированные, которые посвящены конкретному семейству белков.

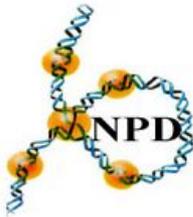
Базы данных, посвященные **структуре белковых молекул**. В основе этих баз данных лежит определенный **уровень структурной организации белков**. Это могут быть базы, представляющие информацию об аминокислотной последовательности белка, о вторичной структуре, третичной. Существуют много баз данных, посвященных различным белковым доменам.

Базы данных, посвященные **белковым взаимодействиям**. К ним относятся базы о белок-белковых взаимодействиях, например базы данных **рецепторов и их лигандов**, базы **белковых коопераций** определенной локализации (мембрана, митохондрии, ядро), белков, контактирующих при выполнении определенной функции (транспортная сеть, сигнальные каскады). Кроме того, в эту группу входят базы данных, посвященные взаимодействиям белков с другими молекулами, например с РНК, ДНК. Есть базы данных по антителам, которые тоже являются белками, и антигенам.

База данных **NPD (Nuclear Protein Database)** посвящена ядерному протеому Позвоночных. Интернет-адрес этой базы данных следующий: <http://npd.hgu.mrc.ac.uk>. Интерфейс базы данных представлен на рисунке 1. К настоящему времени в этой базе находится информация о 1300 белках, локализованных в клеточном ядре. Необходимые данные можно получить, введя в строку поиска определенный белок или какой-либо ядерный компартмент. NPD содержит информацию о размере, массе белковой молекулы, изоэлектрической точке, структуре (аминокислотная последовательность, специфические 17 мотивы, домены), белковых сплайсинговых изоформах, субядерной локализации, биологической и молекулярной функции. Также можно получить сведения о гене, кодирующем этот белок, его нуклеотидной последовательности, хромосомной локализации.

NPD (Nuclear Protein Database)

npd.hgu.mrc.ac.uk/user/



or browse by [compartment](#) or [motif](#)

[About](#)
[Links](#)
[Statistics](#)
[Web Services](#)
[FAQs](#)

The Nuclear Protein Database

The Nuclear Protein Database (NPD) is a searchable database of information on proteins that are localised to the nucleus of vertebrate cells.



How to use the NPD

You can search the whole database using a protein name, protein motif, nuclear compartment name, or keyword term. Alternatively you can view all of the proteins that have been reported as having a localisation to a particular compartment or that are associated with a particular protein domain, using the [nuclear compartment](#) and [domain](#) browsers. An introduction to each of the principal [sub-nuclear compartments](#) is also provided.

What is in the NPD?

The NPD contains information on over 1000 vertebrate proteins (mainly those from mouse and human) that are thought to, or known to, be localised to the cell nucleus. Where known, the [sub-nuclear compartment](#) where the proteins have been found are reported. Also stored is information on the amino acid sequence, predicted protein size and isoelectric point, as well as any repeats, motifs or domains within the protein sequence. Biological and molecular functions of the proteins are described using [GO terms](#). Where appropriate, links to other databases are provided (e.g. [Entrez](#), [SWISS-PROT](#), [OMIM](#), [PubMed](#), [PubMed Central](#)).

What is not in the NPD?

In general only one isoform of the protein is given (usually the largest). The database contains no information on protein isoforms generated by alternative splicing.

Version 2.0

Correspondence: [Professor Wendy Bickmore](#)

© [Medical Research Council Human Genetics Unit](#)

[About](#)
[Links](#)
[Statistics](#)
[Web Services](#)

База данных **SynDB** посвящена нейросинапсам между клетками нервной системы, нейромышечным синапсам и синапсам с железистыми клетками. Интернет-адрес этой базы данных следующий: *http://syndb.cbi.pku.edu.cn*. В базе данных представлена информация об основных структурах синапсов, функциях, опосредованных заболеваниями, терминах, связанных с этой темой. Данные представлены в виде каталогов и гиперссылок на термины. В этой базе данных можно найти информацию о более 14000 синапс-ассоциированных белках, 3000 из которых встречаются у человека.

Где искать описание функции

- Краткое описание функций одного белка и ссылки на другие ресурсы см.



- Краткое описание функций семейств белков и доменов

см. в  и 

- Подробное описание функций генов и их продуктов см в энциклопедиях, таких как





- Подробное описание отдельных классов функций и соответствующих белков см. в специализированных БД, таких как **ENZYME, Rebase.**

Как узнать функцию интересующего нас и

б.м. известного науке белка или гена?

- + /+++ Читайте оригинальные статьи хороших авторов в хороших журналах!
Ищите в PubMed
- /++ Читайте аннотации записи Uniprot
Используйте SRS (Sequence Retrieval System)
- /+ Читайте аннотации записей Pfam и InterPro, содержащие описания семейств доменов, к которым принадлежат домены белка
 - /+ Ищите ваш белок в специализированных БД (БД и энциклопедии, в которых подробно описаны функции генов и их продуктов : KEGG, BIOCYC, ENZYME, TC-DB, REACTOME....)
- Ищите, предположительно, гомологичные белки самостоятельно. Используйте BLAST, psiBLAST, профили и паттерны