



УДК 57.002

## ФАКТОГРАФИЧЕСКИЕ БАЗЫ ДАННЫХ В ОБЛАСТИ БИООРГАНИЧЕСКОЙ ХИМИИ (СПРАВОЧНО-АНАЛИТИЧЕСКИЙ ОБЗОР)\*

© 1998 г. И. Н. Тележинская<sup>#</sup>, Т. В. Овчинникова

Институт биоорганической химии им. М.М.Шемякина и Ю.А.Овчинникова РАН,  
117871, Москва, ГСП-7, ул. Миклухо-Маклая, 16/10

Поступила в редакцию 30.06.97 г. Принята к печати 10.12.97 г.

Приводятся сведения об основных фактографических базах данных в области биоорганической химии. Описаны возможности оперативного доступа и поиска информации, приведены почтовые и электронные адреса, а также ссылки на литературу, в которой рассматриваются возможности этих баз данных.

*Ключевые слова:* фактографические базы данных, доступ к информации, аминокислотная последовательность, нуклеотидная последовательность.

Современный путь получения информации лежит через ее электронные хранилища – базы данных (БД). Большинство журналов, прежде чем опубликовать статью, содержащую нуклеотидные или аминокислотные последовательности, проверяет их наличие в различных БД. Все новые, а также ранее определенные последовательности, включаются в фактографические БД, причем каждая из них имеет свой кодовый номер.

Основным дистрибьютером фактографических БД в области биоорганической химии является Европейский институт биоинформатики в Великобритании (European Bioinformatics Institute – EBI) [3]. Более 50 фактографических БД доступны пользователям компьютерной сети Internet через сетевую службу этого Института (Приложение 1). Доступ к 30 из них может осуществляться путем подписки на лазерные диски, которые обновляются практически каждые два месяца (Приложение 2.1).

### EMBL NUCLEOTIDE SEQUENCE DATABASE

К числу наиболее популярных фактографических БД, содержащих нуклеотидные последовательности, относится библиотека данных EMBL Европейской лаборатории молекулярной биологии в Гейдельберге (European Molecular Biology Laboratory), основанная в 1980 г. БД EMBL содер-

жит нуклеотидные последовательности из опубликованных статей, а также непосредственно предоставленные авторами. Версия 48 (сентябрь 1996 г.) содержит 609 302 252 нт в 931 582 последовательностях, и ожидается, что к концу 1996 г. число последовательностей превысит 1 млн. [53]. БД EMBL доступна пользователям компьютерной сети Internet, а также путем подписки на лазерные диски и магнитные ленты (Приложение 2.2).

### GENBANK

Еще одной широко известной БД по нуклеотидным последовательностям является GenBank Национального центра биотехнологической информации (National Center for Biotechnology Information – NCBI), представляющего собой отдел Национальной медицинской библиотеки при Национальном институте здоровья США (National Library of Medicine, National Institutes of Health – NIH) [54]. GenBank также содержит опубликованные в различных журналах (около 3500 наименований) и непосредственно предоставленные авторами нуклеотидные последовательности, число которых стремительно растет. Так, например, версия GenBank 67,0 (октябрь 1991 г.) содержала 55 169 276 п.о. в 43 903 фрагментах [55], версия 90,0 (август 1995 г.) – 353 713 490 п.о. в 492 483 последовательностях [56], а версия 96,0 (август 1996 г.) – 602 072 354 п.о. в 920 588 последовательностях [57].

БД GenBank доступна пользователям компьютерной сети Internet, а для подписчиков выпускается на магнитных лентах и лазерных дисках, обновляемых каждые два месяца (Приложение 2.3).

\* См. также работы [1,2].

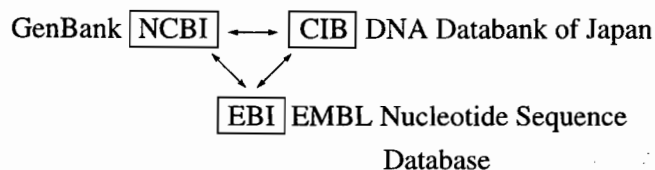
<sup>#</sup> Автор для переписки.

## DNA DATA BANK OF JAPAN (DDBJ)

Банк данных DDBJ, основанный Национальным институтом генетики в Японии (National Institute of Genetics – NIG), начал функционировать в 1986 г. в содружестве с EMBL и GenBank как банк по нуклеотидным последовательностям. В 1995 г. был образован специальный центр DDBJ (Center for Information Biology – CIB) в г. Сидзуока. Версия DDBJ 26 (июль 1996 г.) содержит 835 552 последовательности [58].

Существует три системы предоставления данных DDBJ: Authorin, Sakura и Yamato II, причем последняя создана только в январе 1996 г. Статистика показывает, что в настоящее время 73% данных предоставляется через систему Sakura, которая имеет две версии – японскую и английскую (Приложение 2.4).

Производители и дистрибьютеры трех перечисленных выше БД (EMBL, GenBank и DDBJ) работают в содружестве, обмениваясь информацией ежедневно.



Сегодня примерно 95% всех данных по нуклеотидным последовательностям исследователи могут получить непосредственно из этих БД. Оставшиеся ~5% данных пока получают, как правило, из патентной литературы. Патентные данные предоставляются EBI международной патентной службой (European Patent Office – EPO). В связи с этим с 1993 г. EPO требует от авторов предоставления данных по аминокислотным и нуклеотидным последовательностям в электронной форме [3].

## PIR-INTERNATIONAL PROTEIN SEQUENCE DATABASE

Среди фактографических БД, содержащих аминокислотные последовательности, наиболее известна база PIR (Protein Information Resource). БД была образована в 1988 г. международным содружеством PIR-International, в которое входят Национальный фонд биомедицинских исследований в США (National Biomedical Research Foundation – NBRF), Институт последовательностей белков и Институт биохимии им. Макса Планка в Германии (Martinsried Institute for Protein Sequences – MIPS, Max Planck Institute for Biochemistry), Международная база данных по белкам в Японии (Japan International Protein Information Database – JIPID)[32].

БД PIR содержит первичные структуры белков, опубликованные в литературе, предоставленные PIR-International и транслированные из нуклеотидных последовательностей, предоставленных GenBank [59], EMBL [13] и DDBJ [60]. БД PIR имеет широкое распространение и взаимодействует с другими БД по белковым последовательностям, предоставляемым EBI [3] и NCBI [61,62]. Версия PIR 27,0 содержит 7 620 688 а.о. в 26 798 последовательностях [63]. БД PIR содержит информацию о названии белка и организма, из которого он был выделен, аминокислотную последовательность, основные характеристики белка и его функций в организме, а также ссылки на литературу. Для выявления гомологичных участков в белках MIPS создал специальную программу PileUp.

БД доступна через Internet, а также на магнитных лентах и лазерных дисках (Приложение 2.5).

## SWISS-PROT PROTEIN SEQUENCE DATA BANK

Швейцарская база данных по аминокислотным последовательностям SWISS-PROT была создана в 1986 г. и поддерживается содружеством библиотек данных EMBL и Женевского университета (University of Geneva) [48]. SWISS-PROT использует различные источники данных, в том числе аминокислотные последовательности из PIR и транслированные нуклеотидные последовательности из EMBL, причем распространяется она в том же формате, что и БД EMBL. Версия 32,0 (октябрь 1995 г.) содержит около 17 млн. а.о. в 48 440 последовательностях. Существует приложение к SWISS-PROT – TrEMBL (Translation from EMBL), которое содержит кодирующие последовательности (Coding Sequences, или CDS), выведенные из нуклеотидных последовательностей EMBL и включенные в SWISS-PROT. Формат документа содержит описание функций белков, структурных доменов, посттрансляционных модификаций и т.д. БД SWISS-PROT отличается значительной степенью интеграции с другими БД. SWISS-PROT обменивается данными с 24 БД по нуклеотидным и аминокислотным последовательностям и третичным структурам белков. Так, версия 32,0 (октябрь 1995 г.) имела в среднем 3.5 обмена на каждую последовательность [48], а в октябре 1996 г. на 60 тыс. сиквенсов, содержащих около 21 млн. а.о., приходилось 50 тыс. обменов [65]. Кроме того, уникальным свойством SWISS-PROT является интеграция со специальными базами данных, в качестве примера которых можно привести БД по ретровирусам человека и ВИЧ, БД REBASE по ферментам рестрик-

ции, БД ENZYMES по классификационным номерам ферментов, присвоенным специальной комиссией Международного биохимического союза и др. [55].

SWISS-PROT распространяется на магнитных лентах и лазерных дисках, которые обновляются каждые три месяца, а также по телекоммуникационным сетям через сервер EBI (Приложение 2.6).

#### DGENE

БД DGENE основана в 1981 г. английской информационной фирмой Derwent (Derwent Information Ltd., UK). БД содержит информацию о нуклеотидных и аминокислотных последовательностях, опубликованных в оригинальных патентных документах 40 международных патентных организаций. БД DGENE включает нуклеотидные последовательности длиной более 9 оснований и аминокислотные последовательности длиной более 3 аминокислот. Большая половина последовательностей БД DGENE не включена в другие фактографические БД. В настоящее время БД содержит 193 053 последовательности и обновляется 2 раза в неделю.

БД доступна в режиме on-line через международную научно-техническую информационную сеть STN International (Scientific & Technical Information Network) [1] (Приложение 2.7). Российские ученые могут осуществлять выход в систему STN International через два центра в Москве (Международный центр научно-технической информации и Институт органической химии им. Н.Д.Зелинского РАН) и Санкт-Петербургский центр научно-технической информации.

#### PROTEIN DATA BANK (PDB)

БД PDB Брукхейвенской национальной лаборатории США (Brookhaven National Laboratory) была основана в 1971 г. и содержит информацию о трехмерных структурах белков и нуклеиновых кислот, полученную непосредственно от авторов и из опубликованных результатов рентгеноструктурного анализа, ЯМР-спектроскопии и молекулярного моделирования. Данные, хранящиеся на магнитных лентах, PDB получает из центров в Брукхейвене (США), Кембридже (Великобритания) и Токио (Япония). В октябре 1996 г. БД PDB содержала 4873 структуры, а в сентябре 1997 г. – 6398, из них 5907 белков, 479 нуклеиновых кислот и 12 углеводов. Доступ к БД PDB может осуществляться по телекоммуникационным сетям (Приложение 2.8).

#### NRL\_3D SEQUENCE-STRUCTURE DATABASE

Для быстрого обмена информацией между PIR и PDB и приведения этих БД к единому формату была создана БД NRL\_3D, которая содержит сведения о пространственных структурах белков из PDB и аминокислотных последовательностях из PIR [55]. Спонсорами проекта являются управления научных исследований военно-морского флота и армии США (Office of Naval Research, US Army Medical Research and Development Command).

БД NRL\_3D доступна не только в режиме on-line, но и на магнитных лентах и лазерных дисках (Приложение 2.9). Первая версия БД NRL\_3D (октябрь 1988 г.) включала в себя 293 из 382 белковых структур БД PDB [26]. Версия 20,0 (сентябрь 1995 г.) содержит 6063 структуры [32].

#### NUCLEIC ACID DATABASE (NDB)

Начиная с 1996 г. все данные по пространственным структурам нуклеиновых кислот предоставляются БД NDB Университета в Нью-Джерси, США (The State University of New Jersey) по почтовому адресу или электронной почте (Приложение 2.10). БД также получает информацию из EBI, Центра структурной биологии Национального института биологических наук и изучения человека в Японии (Structural Biology Center, National Institute of Bioscience and Human-Technology – NIBH), Компьютерного центра в Сан-Диего, США (San Diego Supercomputer Center). БД NDB распространяется в PDB-формате.

Одновременно данные отправляются в PDB для внесения в единый центральный архив. Статьи по пространственным структурам белков и нуклеиновых кислот не могут быть опубликованы, пока авторы не получат в PDB или NDB регистрационный номер.

#### THE INTEGRATED DATABASE (ID)

Чтобы обеспечить данными GenBank, Национальный центр биотехнологической информации (NCBI) в США поддерживает международную Интегрированную базу данных ID [56]. Источниками данных для ID являются БД по нуклеотидным последовательностям EMBL, DDBJ, GSDB (Genome Sequence DataBase) [67], dbEST [10], dbSTS [11] и патенты, а также БД по аминокислотным последовательностям PIR, SWISS-PROT, PRF (Protein Research Foundation) и PDB.

#### OWL

Еще одним примером интегрированного банка данных, включающего в себя восемь БД по первичным структурам белков, определенным тра-

диционными методами и выведенным из нуклеотидных последовательностей, является БД OWL. Название БД OWL не является аббревиатурой, а происходит от английского написания слова "сова", изображение которой присутствует на гербе университета в Лидсе (University of Leeds, UK) – одного из основных спонсоров проекта OWL [68]. Проект поддерживают также Совет по науке и технологиям (Science and Engineering Council, UK) и Программа белковой инженерии (Protein Engineering Initiative, UK).

В БД OWL включены данные из PIR, SWISS-PROT, GenBank, PDB, NRL\_3D, NBRF-PIR NEW [69], NEWAT86 [70], PSD-Киото [71]. Все первоисточники конвертированы в общий формат, в основе которого лежит формат PIR [72]. БД OWL доступна пользователям компьютерной сети Internet, а также на магнитных лентах и лазерных дисках (Приложение 2.11).

#### A PROTEIN SEQUENCE/STRUCTURE DATABASE ISIS

OWL в свою очередь является частью интегрированной БД по аминокислотным последовательностям и пространственным структурам белков – ISIS (Integrated Sequence/Integrated Structure), в которую входят БД: PIR, NBRF-PIR NEW, SWISS-PROT, GenBank, NEWAT86, PDB, JIPID [73] (Приложение 2.12).

#### DNASTAR

Еще одним примером комбинированного банка данных, в который входят аминокислотные последовательности из PIR и SWISS-PROT и нуклеотидные последовательности из EMBL и GenBank, является DNASTAR. Так, объединенная версия SWISS-PROT + PIR Translated Release 98,0 (январь 1997 г.) содержит 263 254 аминокислотные последовательности, а версия EMBL+GenBank Release 99,0 (январь 1997 г.) – 374 914 нуклеотидных последовательностей [74]. DNASTAR распространяется на лазерных дисках с обновлением каждые два месяца (Приложение 2.13).

Система для анализа последовательностей состоит из семи модулей: EDITSEQ, GENEMAN, MAPDRAW, MEGALIGN, PRIMERSELECT, PROTEAN, SEQMAN. Каждый из них может использоваться независимо от других. Программа позволяет конвертировать аминокислотную последовательность в нуклеотидную и наоборот, а также сравнивать последовательности между собой, моделировать вторичную структуру, проводить различные теоретические исследования на основе сведений о специфичности протеолитических ферментов и рестриктаз. Формат документа модуля GeneMan представлен в Приложении 3.

#### COMPLEX CARBOHYDRATE STRUCTURAL DATABASE (CCSD)

Наряду с фактографическими БД по аминокислотным и нуклеотидным последовательностям существует ряд других специализированных банков, среди которых можно упомянуть CCSD – БД университета в Джорджии, США (Complex Carbohydrate Research Center at the University of Georgia) по структурам углеводов [55]. Спонсорами проекта также являются Министерство энергетики США (US Department of Energy), Национальный институт здоровья США (НИИ), Европейское экономическое сообщество (European Economic Community). БД CCSD содержит информацию об углеводсодержащих биополимерах, в том числе гликопротеинах и гликолипидах. CCSD выпускается на флоппи-дисках для IBM-совместимых персональных компьютеров (Приложение 2.14).

#### CAMBRIDGE STRUCTURAL DATABASE SYSTEM (CSD SYSTEM)

Кембриджский кристаллографический центр CCDC (Cambridge Crystallographic Data Centre) был создан при химических лабораториях Кембриджского университета (University Chemical Laboratories) в 1965 г., а в 1989 г. он становится независимым институтом. CCDC изучает пространственные структуры небольших молекул органических и металлоорганических соединений, таких, как нуклеотиды и нуклеозиды, аминокислоты и пептиды, порфирины и их комплексы, стероиды и их производные, терпены, углеводы и другие соединения и молекулярные комплексы. БД CSD System Кристаллографического центра CCDC в апреле 1997 г. содержала информацию о пространственной структуре 167 797 соединений, из них 83 632 – органических и 84 165 – металлоорганических. БД CSD System распространяется на лазерных дисках (Приложение 2.15). Пользователи компьютерной сети могут получить информацию о CCDC через Internet (<http://expasy.hcuge.ch>).

Все упомянутые БД регулярно обновляются и отличаются точным архивированием поступающих данных, стабильными и стандартизированными источниками информации, динамичной адаптацией к новым концепциям в соответствующей области знаний. Лазерные диски, подписка на которые является наиболее распространенным способом предоставления фактографических БД, обновляются практически каждые два месяца. Кроме того, БД доступны пользователям компьютерной сети Internet. Более подробную информацию можно получить в Национальном центре биотехнологической информации США по электронной почте ([info@ncbi.nlm.nih.gov](mailto:info@ncbi.nlm.nih.gov)) или через Internet (<http://expasy.hcuge.ch>).

## Приложение 1. Базы данных, поддерживаемые EBI

Название	Содержание	Ссылка
3D ali	Structure-based sequence alignments	[4]
Alu	ALU sequences and alignments	[5]
*Berlin RNA	5S rRNA sequences	[6]
*Bio-Catalog	Directory of molecular biology and genetics software	[3]
*Blocks	Protein Blocks Database	[7]
*CpGisle	CpG islands database	[8]
Cutg	Codon usage tabulated from GenBank	[9]
dbEST	Expressed sequence tags	[10]
dbSTS	Sequence tagged sites	[11]
DSSP	Secondary structure assignments of pdb files	[3]
*ECDC	<i>Escherichia coli</i> database collection	[12]
*EMBL	Nucleotide sequence database	[13]
*Enzyme	Database of EC nomenclature	[14]
*EPD	Eukaryotic promoter database	[15]
FlyBase	Drosophila genetic map database	[16]
FSSP	Families of structurally similar proteins	[17]
HaemA	Haemophilia A database	[18]
*HaemB	Haemophilia B database	[19]
*HLA	HLA class I and II sequence database	[20]
*HSSP	Protein structure-sequence alignments	[21]
*Kabat	Proteins of immunological interest	[3]
*LiMB	List of molecular biology databases	[22]
*Lista	Yeast protein coding sequences	[23]
Methyl	Site-specific methylation	[24]
Misfolded	Deliberately misfolded protein models	[25]
NRL_3D	Sequence-structure database	[26]
NRSub	Non-redundant <i>Bacillus subtilis</i> genome database	[27]
*Nucleosomal DNA	Nucleosomal DNA sequences	[28]
*P53	P53 mutations database	[29]
PDB	Brookhaven protein structures database	[30]
PDB Select	Representative list of PDB chain identifiers	[31]
PIR	Protein sequence database	[32]
*PKCDD	Protein kinase catalytic domain sequence database	[33]
*Prints	Protein motif fingerprint database	[34]
Prodom	Protein sequence modules (recurring domains)	[35]
*Prosite	Prosite pattern database	[36]
PUU	Database of structural domains	[37]
RDP	Ribosomal database project	[38]
*REBASE	Restriction enzyme database	[39]
*RELibrary	Comprehensive restriction enzyme lists	[40]
*RepBase	Prototypic human repetitive DNA sequences	[41]
*RHdb	Radiation hybrid database	[42]
*RLDB	Reference library database	[43]
*rRNA	Small subunit rRNA sequences	[44]
*SBASE	Protein domain database	[45]
*SeqAnalRef	Sequence analysis bibliography	[3]
SmallRNA	Compilation of small RNA sequences	[46]
*SRP	Signal recognition particle database	[47]
*SWISS-PROT	Protein sequence database	[48]
TFD	Transcription factor database	[49]
*TransFac	Eukaryotic <i>cis</i> -acting regulatory DNA elements and <i>trans</i> -acting factors	[50]
*Transterm	Translational termination signal database	[51]
*tRNA	Database of tRNA sequences	[52]
Yeast	Yeast chromosome database	[3]

\* Базы данных, предоставляемые пользователям на лазерных дисках CD-ROM.

## ПРИЛОЖЕНИЕ 2.

## 1. Как контактировать с EBI [3]

- European Bioinformatics Institute (EBI), Hinxton Hall, Hinxton, Cambridge CB10 1RQ, UK
- ftp.ebi.ac.uk (FTP server)
- gopher.ebi.ac.uk (Gopher server)
- http://www.ebi.ac.uk (World Wide Web)
- http://www.ebi.ac.uk/dbases/topdata.html
- Datalib@ebi.ac.uk (для справок)
- Datasubs@ebi.ac.uk (для предоставления данных БД EMBL и SwissProt)
- Update@ebi.ac.uk (для коррекции нуклеотидных последовательностей)
- RNdb@ebi.ac.uk (для предоставления данных БД RNdb)
- Netserv@ebi.ac.uk (e-mail file server)
- NetHelp@ebi.ac.uk (для справок по Network server)
- blitz@ebi.ac.uk (Mpsrch protein sequence search server)
- fasta@ebi.ac.uk (FastA sequence search server)
- Tel.: +44 1223 494 400
- Fax: +44 1223 494 468

## 2. Как контактировать с EMBL [54,55]

- EMBL Nucleotide Sequence Data Library, European Molecular Biology Laboratory, Postfach 102209, Meyerhofstrasse 1, 6900 Heidelberg, Germany.
- EMBL Nucleotide Sequence Submissions, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK
- http://www.ebi.ac.uk/subs/emblsubs.html (для предоставления данных)
- Datasubs@ebi.ac.uk (для предоставления данных)
- Datalib@ebi.ac.uk (для справок)
- Update@ebi.ac.uk (для уведомления о публикациях)
- Datalib@embl.bitnet
- Datalib@embl-heidelberg.de
- Tel.: +44(0)1223 494411
- Fax: +44(0)1223 494472 (для предоставления данных) +44(0)1223 494468 (основной)

## 3. Как контактировать с GenBank [54-56]

- GenBank(R) Submissions, National Center for Biotechnology Information, Building 38A, Room 8N-803, 8600 Rockville Pike, Bethesda, MD 20894, USA
- http://www.ncbi.nlm.nih.gov/
- Gb-sub@ncbi.nlm.nih.gov (для предоставления данных БД GenBank)
- Info@ncbi.nlm.nih.gov (для справок)
- Update@ncbi.nlm.nih.gov (для проверки последовательностей в БД и уведомления о публикациях)
- Gos@genbank.bio.net
- Gb-sub%life@lanl.gov (для предоставления данных)
- Tel.: +1 301 496 2475
- Fax: +1 301 480 9241
- E-mail: Authorin@ncbi.nlm.nih.gov
- Tel.: +1 301 402 1301

## 4. Как контактировать с DDBJ [54]

- DNA Data Bank of Japan, Center for Information Biology, National Institute of Genetics, Mishima, Shizuoka 411, Japan
- http://sakura.ddbj.nig.ac.jp
- http://www.ddbj.nig.ac.jp [WWW Uniform Resource Locator (URL)]
- E-mail: Ddbjsub@ddbj.nig.ac.jp (для предоставления данных)
- Ddbj@ddbj.nig.ac.jp (для справок)
- Ddbjupdt@ddbj.nig.ac.jp
- Tel.: +81 559 81 6853
- Fax: +81 559 81 8849

## 5. Как контактировать с PIR-International [32,64]

USA (NBRF)

- National Biomedical Research Foundation (NBRF), 3900 Reservoir Road N.W. Washington, D.C. 20007, USA
- Tel.: +1 202 687 2121
- Fax: +1 202 687 1662
- E-mail: Pirmail@nbrf.georgetown.edu (PIR Technical Development Bulletin)
- Pirmail@gunbrf.bitnet

Europe (MIPS)

- Martinsried Institute for Protein Sequences (MIPS), Max Planck Institute for Biochemistry, D-82152 Martinsried, Germany
- http://www.mips.biochem.mpg.de
- http://www.gcg.com (PileUp program)
- E-mail: Mewes@mips.embnet.org
- Tel.: +49 89 8578 2657
- Fax: +49 89 8578 2655

Asia and Australia (JIPID)

- Japan International Protein Information Database (JIPID), Science University of Tokyo, 2669 Yamazaki, Noda 278, Japan
- E-mail: TSUGITA@JPNSUT31.BITNET
- Tel.: +81 471 239778
- Fax: +81 471 221544
- ftp.bchs.uh.edu (The University of Houston Gene-Server)
- http://www.gdb.org/Dan/proteins/pir.html (The Johns Hopkins University)
- http://www.ncbi.nlm.nih.gov/ (NCBI)

## 6. Как контактировать с SWISS-PROT [48,55,65]

- Department of Medical Biochemistry, University of Geneva, 1 rue Michel Servet, 1211 Geneva 4, Switzerland
- The EMBL Outstation-European Bioinformatics Institute, Hinxton Hall, Hinxton, Cambridge CB10 1RQ, UK
- EBI ftp server: ftp.ebi.ac.uk (или 192.54.41.33)
- EBI server: http://www.ebi.ac.uk/
- http://expasy.hcuge.ch/ (WWW URL)
- ftp://ftp.ebi.ac.uk/pub/databases/trembl/

- NCBI Repository, National Library of Medicine, NIH, Washington, DC:

ncbi.nlm.nih.gov (или 130.14.20.1)

- ExPASy (Expert Protein Analysis System) server, University of Geneva, Switzerland:

expasy.hcuge.ch (или 129.195.254.61)

- National Institute of Genetics (Japan)

ftp server: ftp2.ddbj.nig.ac.jp (или 133.39.3.6)

- Tel.: +44 1223 494 400

- Fax: +44 1223 494 468

- E-mail: Datasubs@ebi.ac.uk (для предоставления последовательностей)

Junker@ebi.ac.uk (для справок)

Bairoch@cmu.unige.ch

Datalib@embl-heidelberg.de

Datalib@ebi.ac.uk

### 7. Как контактировать с DGENE

- STN-International, c/o FIZ Karlsruhe D-76344 Eggenstein-Leopoldshafen, Germany

- <http://www.fiz-karlsruhe.de>

- E-mail: hlpdeskk@fiz-karlsruhe.de

- Fax: (+49) 72 47/ 808-131

- STN-International, c/o Chemical Abstract Service, 2540 Olentangy River Road P.O. Box 3012 Columbus, Ohio 43210-0012 USA

- <http://www.cas.org>

- E-mail: help@cas.org

- Fax: (+1) 614-447-3798

- STN-International, c/o The Japan Science and Technology Corporation, Information Center for Science and Technology (JICST) 5-3 Yonbancho Chiyoda-ku, Tokyo 102, Japan

- <http://www.jicst.go.jp>

- Fax: (+81) 3/5214-8410

- Москва: avas@icsti.su

khutor@ioc.ac.ru

### 8. Как контактировать с PDB [54,55,66]

- Protein Data Bank, Chemistry Department, Building 555, Brookhaven National Laboratory, PO Box 5000, Upton, NY 11973-5000, USA

- <http://www.pdb.bnl.gov>

- <ftp://pdb.pdb.bnl.gov>

- <gopher://pdb.pdb.bnl.gov>

- E-mail: pdb@bnl.gov

pdb@bnlchm.bitnet

- Tel.: +1 516 282 3629

- Fax: +1 516 282 5751

- University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, England

- University of Tokyo, Hongo, Tokyo, Japan

### 9. Как контактировать с NRL\_3D [55]

- Office of Naval Research, US Army Medical Research and Development Command

- E-mail: pirmail@gunbrf.bitnet

### 10. Как контактировать с NDB [54]

- Anke Gelbin, NDB, Department of Chemistry, Rutgers, The State University of New Jersey, PO Box 939, Piscataway, NY 08855-0939, USA.

- NDB home: <http://ndbserver.rutgers.edu/>

- E-mail: deposit@ndbserver.rutgers.edu

ndbadmin@ndbserver.rutgers.edu

- Tel.: +1 908 445 0103

- Fax: +1 908 445 5958

### 11. Как контактировать с OWL [55,68,72]

- Departments of Genetics and Biophysics, University of Leeds, Leeds LS2 9JT, UK

- S-ing2.dl.ac.uk (FTP server Europe/Africa/Asia)

- Ncbi.nlm.nih.gov (FTP server USA/Pacific rim)

- URL: <http://www.gbd.org/Dan.proteins/owl.html> (WWW server)

- E-mail: Uig@daresbury.ac.uk

### 12. Как контактировать с ISIS [68]

- Software Distribution Office, Room C12, SERC Daresbury Laboratory, Daresbury, Warrington WA4 4AD, UK

- E-mail: Uig@daresbury.ac.uk

- Tel.: (0925) 603351

### 13. Как контактировать с DNASTAR [74]

- DNASTAR Inc 1228 South Park Street, Madison, WI 53715, USA

- Internet: support@dnastar.com

- Website: [www.dnastar.com](http://www.dnastar.com)

- E-mail: sales@dnastar.com

- Tel.: (608) 258-7420

- Fax: (608) 258-7439

### 14. Как контактировать с CCSD [55]

- Complex Carbohydrate Structural Database, Complex Carbohydrate Research Center at the University of Georgia, USA.

- E-mail: CarBank@uga.bitnet

### 15. Как контактировать с CDS System [54,55]

- Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, UK

- WWW: <http://www.ccdc/cam.ac.uk>

- E-mail: fileserv@chemcrys.cam.ac.uk

admin@ccdc.cam.ac.uk

software@ccdc.cam.ac.uk

teched@ccdc.cam.ac.uk

- Tel.: +44(0)1223 336 408

+44(0)1223 336 031/2 (технический редактор)

- Fax: +44(0)1223 336 033

**Приложение 3.** Формат документа базы данных DNASTAR [модуль GeneMap, объединенная версия SWISS-PROT + PIR Translated Release 98,0 (июль 1997 г.)]

ENTRY SW: 12KD\_MYCLE STANDARD  
 PRT; 156 AA  
 ACCESSION: P15878;  
 PIR: S08427 #type complete  
 ACCESSION: S08427  
 GB: MLMLA12  
 ACCESSION: X51328

DESC 12 KD PROTEIN

DATE 01-APR-1990 (REL. 14, CREATED)  
 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)  
 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)

GENE MLA12A

SOURCE MYCOBACTERIUM LEPRAE.

REF [1]

COM SEQUENCE FROM N.A.

MED MEDLINE; 90206800

AUTH HARTSKEERL R.A., STABEL L.F.E.M., HERMANS C.R., KLATSER P.R., THOLE J.E.R.

CIT NUCLEIC ACIDS RES. 18:1294-1294 (1990)

TAXONOMY PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; MYCOBACTERIACEAE

COMMENT—CAUTION: IT IS UNCERTAIN WHETHER MET-1, MET-24, VAL-25 OR VAL-49 IS THE INITIATOR.  
 Nucleic Acid Features translated to generate this entry:  
 CDS 70...540  
 /note="open reading frame (AA 1-156)"  
 /codon\_start=1  
 /db\_xref="PID:g44413"  
 /db\_xref="SWISS-PROT:P15878"  
 /transl\_table=1

SUMMARY AA 156; MW 17149; CN 1

SEQUENCE

1	MNDIILKLFH	ISLNATTWIG	RIGMVILPLL
	VYFITYRWCI	GLQRSDRAVL	EHGIETGI IK
61	RLPHGAYIEL	HQPLGPVDDH	GHIPILE YQG
	TAVPKRMNKL	GSAGSPSSGS	FLFADPVSED
121	AALREATHVA	EQRALTALRE	HQDSIASSPN
	GERGKH		

## СПИСОК ЛИТЕРАТУРЫ

1. Шкаренкова Л.С., Орловская Т.Т., Овчинникова Т.В. // Биоорган. химия. 1997. Т. 23. С. 152–158.
2. Шкаренкова Л.С., Орловская Т.Т., Овчинникова Т.В. // Биоорган. химия. 1998. Т. 24. С. 68–73.
3. Rodriguez-Tome P., Stoehr P.J., Cameron G.N., Flores T.P. // Nucleic Acids Res. 1996. V. 24. P. 6–12.
4. Pascarella S., Argos P. // Protein Engng. 1992. V. 5. P. 121–137.
5. Jurka J., Smith T. // Proc. Natl. Acad. Sci. USA. 1988. V. 85. P. 4775–4778.
6. Specht T., Wolters J., Erdmann V.A. // Nucleic Acids Res. 1991. V. 19. P. 2189–2191.
7. Pietrovski S., Henikoff J.G., Henikoff S. // Nucleic Acids Res. 1996. V. 24. P. 197–200.
8. Larsen F., Gundersen G., Lopez R., Prydz H. // Genomics. 1992. V. 13. P. 1095–1107.
9. Nakamura Y., Wada K., Wada Y., Doi H., Kanaya S., Gojobori T., Ikemura T. // Nucleic Acids Res. 1996. V. 24. P. 214–215.
10. Boguski M.S., Lowe T.M.J., Tolstoshev C.M. // Nature Genet. 1993. V. 4. P. 332–333.



11. Olson M., Hood L., Cantor C., Botstein D. // *Science*. 1989. V. 254. P. 1434–1435.
12. Kroger M., Wahl R. // *Nucleic Acids Res.* 1996. V. 24. P. 29–31.
13. Emmert D.B., Stoehr P.J., Stoesser G., Cameron G.N. // *Nucleic Acids Res.* 1994. V. 22. P. 3445–3449.
14. Bairoch A. // *Nucleic Acids Res.* 1996. V. 24. P.221–222.
15. Bucher P., Trifonov E.N. // *Nucleic Acids Res.* 1986. V. 14. P.10009–10026.
16. Gelbart W.M., Rindone W.P., Chillemi J., Russo S., Crosby M., Matthews B., Ashburner M., Drysdale R.A., de Grey A., Whitfield E.J., Kaufman T., Matthews K., Gilbert D., Tolstoshev C. // *Nucleic Acids Res.* 1996. V. 24. P.53–56.
17. Holm L., Sander C. // *Nucleic Acids Res.* 1996. V. 24. P. 206–209.
18. Tuddenham E.G., Schwaab T., Seehafer J., Millar D.S., Gitschier F., Higuchi M., Bidichandani S., Connor J.M., Hoyer L.W., Yoshioka A. // *Nucleic Acids Res.* 1994. V. 22. P. 4851–4868.
19. Giannelli F., Green P.M., Sommer S.S., Lillcrap D.P., Ludwig M., Schwaab R., Reitsma P.H., Goossens M., Yoshioka A., Brownlee G.G. // *Nucleic Acids Res.* 1994. V. 22. P. 3534–3546.
20. Bodmer J.G., Marsh S.G., Albert E.D., Bodmer W.F., Dupont B., Erlich H.A., Mach B., Mayr W.R., Parham P., Sasazuki T. // *Tissue Antigens.* 1994. V. 44. P. 1–18.
21. Schneider R., Sander C. // *Nucleic Acids Res.* 1996. V.24. P.201–205.
22. Keen G., Redgrave G., Lawton J., Cinkosky M., Mishra S., Fickett J., Burks C. // *Mathl. Comput. Modelling.* 1992. V. 16. P. 93–101.
23. Dolz R., Mosse M.-O., Slonimski P.P., Bairoch A., Linder P. // *Nucleic Acids Res.* 1996. V. 24. P. 50–52.
24. McClelland M., Nelson M., Raschke E. // *Nucleic Acids Res.* 1994. V. 22. P. 3640–3659.
25. Holm L., Sander C. // *J. Mol. Biol.* 1992. V. 225. P. 93–105.
26. Pattabiraman N., Namboodiri K., Lowrey A., Gaber B.P. // *Protein Seq. Data Anal.* 1990. V. 3. P. 387–405.
27. Perriere G., Moszer I., Gojobori T. // *Nucleic Acids Res.* 1996. V. 24. P. 41–45.
28. Isohikhes I., Trifonov E.N. // *Nucleic Acids Res.* 1993. V.21. P.4857–4859.
29. Hollstein M., Rice K., Greenblatt M.S., Soussi T., Fuchs R., Sorlie T., Hovig E., Smith-Sorenson B., Montesano R., Harris C.C. // *Nucleic Acids Res.* 1994. V. 22. P. 3551–3555.
30. Bernstein F.C., Koetzle T.F., Williams G.J.B., Meyer E.F., Brice M.D., Rodgers J.R., Kennard O., Shimanouchi T., Tasumi M. // *J. Mol. Biol.* 1977. V. 112. P. 535–542.
31. Hobohm U., Scharf M., Schneider R., Sander C. // *Protein Sci.* 1992. V. 1. P. 409–417.
32. George D.G., Barker W.C., Mewes H.W., Pfeiffer F., Tsugita A. // *Nucleic Acids Res.* 1996. V. 24. P. 17–20.
33. Hanks K.S., Quinn A.M. // *Methods Enzymol.* 1991. V. 200. P. 38–62.
34. Attwood T.K., Beck M.E., Bleasby A.J., Degtyarenko K., Parry Smith D.J. // *Nucleic Acids Res.* 1996. V. 24. P. 182–188.
35. Sonnhammer E.L.L., Kahn D. // *Protein Sci.* 1994. V. 3. P. 482–492.
36. Bairoch A., Bucher P., Hofmann K. // *Nucleic Acids Res.* 1996. V.24. P.189–196.
37. Holm L., Sander C. // *Proteins.* 1994. V.19. P.256–268.
38. Maidak B.L., Olsen G.J., Larsen N., Overbeek R., McCaughey M.J., Woese C.R. // *Nucleic Acids Res.* 1996. V. 24. P. 82–85.
39. Roberts R.J., Macelis D. // *Nucleic Acids Res.* 1996. V. 24. P. 223–235.
40. Raschke E. // *Genetic Analysis. Techniques and Applications.* 1993. V. 10. P. 49–60.
41. Jurka J., Walichiewicz J., Milosavljevic A. // *J. Mol. Evol.* 1992. V. 35. P. 286–291.
42. Rodriguez-Tome P., Lijnzaad P. // *Nucleic Acids Res.* 1997. V. 25. P. 81–84.
43. Lehrach H. // *Genome Analysis.* 1990. V.1. P.39–81.
44. Van de Peer Y., Nicolai S., De Rijk P., De Wachter R. // *Nucleic Acids Res.* 1996. V. 24. P. 86–91.
45. Murvai J., Gabrielian A., Fabian P., Hatsagi Z., Degtyarenko K., Hegyi H., Pongor S. // *Nucleic Acids Res.* 1996. V. 24. P. 210–213.
46. Gu J., Reddy R. // *Nucleic Acids Res.* 1996. V. 24. P. 73–75.
47. Larsen N., Zwieb C. // *Nucleic Acids Res.* 1996. V. 24. P. 80–81.
48. Bairoch A., Apweiler R. // *Nucleic Acids Res.* 1996. V. 24. P. 21–25.
49. Ghosh D. // *Nucleic Acids Res.* 1992. V. 20. P. 2091–2093.
50. Wingender E., Dietze P., Karas H., Knuppel R. // *Nucleic Acids Res.* 1996. V. 24. P. 238–241.
51. Dalphin M.E., Brown C.M., Stockwell P.A., Tate W.P. // *Nucleic Acids Res.* 1996. V. 24. P. 216–218.
52. Sprinzl M., Steegborn C., Hubel F., Steinberg S. // *Nucleic Acids Res.* 1996. V. 24. P. 68–72.
53. Stoesser G., Sterk P., Tuli M.A., Stoehr P.J., Cameron G.N. // *Nucleic Acids Res.* 1997. V. 25. P. 7–13.
54. Instructions to Authors // *Nucleic Acids Res.* 1997. V. 25. № 1.
55. Boguski M.S., Ostell J., States D.J. // *Protein Engineering: a Practical Approach/ Eds A.R. Rees, M.J.E. Sternberg, R. Wetzel. Oxford: IRL Press, 1992. P. 57–88.*
56. Benson D.A., Boguski M., Lipman D.J., Ostell J. // *Nucleic Acids Res.* 1996. V. 24. P. 1–5.
57. Benson D.A., Boguski M.S., Lipman D.J., Ostell J. // *Nucleic Acids Res.* 1997. V. 25. P. 1–6.
58. Tateno Y., Gojobori T. // *Nucleic Acids Res.* 1997. V. 25. P. 14–17.
59. Benson D., Boguski M., Lipman D.J., Ostell J. // *Nucleic Acids Res.* 1994. V. 22. P. 3441–3444.
60. Tateno Y., Ugawa Y., Yamazaki Y., Hayashida H., Saitou N., Gojobori T. // *CODATA Bull.* 1991. V. 23. P. 74–75.
61. Benson D., Boguski M., Lipman D.J., Ostell J. // *Genomics.* 1990. V. 6. P. 389–391.

62. Benson D. // CODATA Bull. 1991. V.23. P.76–78.
63. Argos P. // Methods Enzymol. 1990. V.182. P.751–778.
64. George D.G., Dodson R.J., Garavelli J.S., Haft D.H., Hunt L.T., Marzec C.R., Orcutt B.C., Sidman K.E., Srinivasarao G.Y., Yeh L.-S.L., Arminski L.M., Ledley R.S., Tsugita A., Barker W.C. // Nucleic Acids Res. 1997. V. 25. P. 24–27.
65. Bairoch A., Apweiler R. // Nucleic Acids Res. 1997. V. 25. P. 31–36.
66. Bernstein F.C., Koetzle T.F., Williams G.J.B., Meyer E.F., Brice M.D., Rodgers J.R., Kennard O., Shimanouchi T., Tasumi M. // J. Mol. Biol. 1977. V. 112. P. 535–542.
67. Harger C., Skupski M., Allen E., Clark C., Crowley D., Dickinson E., Easley D., Espinosa-Lujan A., Farmer A., Fields C., Flores L., Harris L., Keen G., Manning M., McLeod M., O'Neill J., Pumilia M., Reinert R., Rider D., Rohrllich J., Romero Y., Schwertfeger J., Seluja G., Siepel A., Singh G., Smyth L., Stamper D., Stein J., Suggs R., Takkallapalli R., Thayer N., Thompson G., Walsh C., Wedgeworth III F., Schad P.A. // Nucleic Acids Res. 1997. V. 25. P. 18–23.
68. Bleasby A.J., Wootton J.C. // Prot. Engng. 1990. V. 3. P. 153–159.
69. George D.G., Barker W.C., Hunt L. // Nucleic Acids Res. 1986. V. 14. P. 11–15.
70. Doolittle R.F. // Science. 1981. V. 214. P. 149–159.
71. Claverie J.-M., Bricault L. // Prot. Struct. Funct. Genet. 1986. V. 1. P. 60–70.
72. Bleasby A.J., Akrigg D., Attwood T.K. // Nucleic Acids Res. 1994. V. 22. P. 3574–3577.
73. Akrigg D., Bleasby A.J., Dix N.I.M., Findlay J.B.C., North A.C.T., Parry-Smith D., Wootton J.C., Blundell T.L., Gardner S.P., Hayes F., Islam S., Sternberg M.J.E., Thornton J.M., Tickle I.J. // Nature. 1988. V. 335. P. 745–746.
74. Compact Disk DNASTAR. January 1997.

## Molecular Sequence Databases in the Field of Bioorganic Chemistry (Analytical Review)

I. N. Telezhinskaya and T. V. Ovchinnikova

*Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences,  
ul. Miklukho-Maklaya 16/10, Moscow, 117871, Russia*

The main scientific sequence databases of interest for researchers working in the field of bioorganic chemistry are reviewed. Information is given concerning possibilities for rapid access and efficient search for needed information, postal and e-mail addresses, and literature sources in which these databases are comprehensively described.

*Keywords: sequence databases, rapid informational access, amino acid sequence, nucleic acid sequence*

Сдано в набор 04.02.98 г.

Подписано к печати 13.04.98 г.

Формат бумаги 60 × 88<sup>1</sup>/<sub>8</sub>

Офсетная печать

Усл. печ. л. 10.0

Усл. кр.-отт. 3.1 тыс.

Уч.-изд. л. 10.5

Бум. л. 5.0

Тираж 291 экз.

Зак. 3537