



УДК 57.002

**НАУЧНО-БИБЛИОГРАФИЧЕСКАЯ ИНФОРМАЦИЯ  
ДЛЯ ОБЕСПЕЧЕНИЯ ИССЛЕДОВАНИЙ В ОБЛАСТИ  
ФИЗИКО-ХИМИЧЕСКОЙ БИОЛОГИИ  
(СПРАВОЧНО-АНАЛИТИЧЕСКИЙ ОБЗОР)**© 1997 г. Л. С. Шкаренкова<sup>#</sup>, Т. Т. Орловская, Т. В. ОвчинниковаИнститут биоорганической химии им. М.М. Шемякина и Ю.А. Овчинникова РАН,  
117871, Москва, ГСП-7, ул. Миклухо-Маклая, 16/10

Поступила в редакцию 30.03.96 г.

Представлены основные научно-библиографические базы данных, интересные для исследователей, работающих в области физико-химической биологии. Описаны возможности оперативного доступа и эффективного поиска необходимых источников информации.

*Ключевые слова:* информационное обеспечение, научно-библиографические базы данных, доступ, поиск.

Научная информация как фундаментальная основа исследовательской работы должна быть всеохватывающей и легко доступной. Современный путь получения информации лежит через базы данных (БД) – электронные хранилища научных и практических знаний. В этой статье даны характеристики и описаны основные принципы работы с библиографическими БД, интересными для исследователей в области физико-химической биологии, приведены их адреса и ссылки на литературу.

Любая научно-библиографическая БД представляет собой электронную версию определенного печатного издания. Научно-библиографические БД обычно строятся на основе множества периодических журналов, книг, обзоров, отчетов, относящихся к определенной области знаний. Для того чтобы в этом море информации найти нужные источники, необходимо оптимизировать стратегию поиска. Такая оптимизация подразумевает, во-первых, выбор необходимых БД с учетом их тематической направленности и, во-вторых, правильное составление поискового предложения (или запроса). Поисковое предложение состоит из набора поисковых терминов, управляемых логическими и контекстными операторами. Работа в любой БД осуществляется с помощью программного обеспечения, которое разрабатывается либо фирмой-производителем БД, либо коммерческой организацией, предоставляющей базы пользователю, – хостом.

Хост (от английского *host* – множество, хозяин) – это коммерческий автоматизированный банк

<sup>#</sup> Автор для переписки.

данных, объединяющий большое количество БД, производимых различными информационными фирмами, с целью их распространения. Каждый хост имеет свой командный язык, особенности которого необходимо знать для работы с БД. Наиболее известны следующие хосты: Dialog (США), Data-Star (Швейцария), Dimdi (Германия), Questel/Orbit (Франция, США). В отличие от перечисленных выше хост STN International (Scientific & Technical Information Network) представляет собой международную распределенную коммерческую сеть с центрами в трех странах: Германии (г. Карлсруэ), США (г. Колумбус, где расположено Американское химическое общество) и Японии (г. Токио, на основе Японского информационного центра науки и технологии). Эти три центра связаны между собой спутниковой связью.

В связи с тем что в последние годы российские пользователи имеют существенные льготы при работе с STN International, мы будем рассматривать те БД, которые доступны через этот хост.

В настоящее время STN International содержит около 190 БД, в том числе учебные. Пользователь, обращающийся в эту сеть, получает быстрый ответ независимо от того, в каком центре находится запрошенная база. Наряду с научными и техническими БД в этом хосте представлены также БД с коммерческой, экономической и юридической информацией, а также патентные БД.

STN International и другие хосты издают фирменные каталоги представляемых ими пользователям БД [1, 2]. Однако такие каталоги обычно малодоступны, достаточно громоздки и рассчитаны

на узкий круг специалистов по организации и поиску информации современными инженерно-техническими средствами. В обзоре предпринята попытка выделить и охарактеризовать те БД данного хоста, которые наиболее интересны для научных сотрудников, ведущих активную научно-исследовательскую работу в физико-химической биологии и смежных областях и публикующих свои результаты в периодической научной печати соответствующего профиля.

Выход в систему STN International осуществляется в России 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. Для контакта с московскими центрами можно использовать их электронные адреса: avas@icsti. su и khutor@ioc.ac.ru.

Работу с информационными источниками в режиме удаленного доступа в дальнейшем мы будем обозначать термином on-line. К другим способам хранения библиографической информации относятся флоппи, а также лазерные (или компактные) диски, на которых фирмы, производители БД, распространяют свою продукцию по подписке.

Что касается сети Internet, то фирмы, производители по сути своей коммерческих научно-библиографических БД, открывают пока доступ лишь к рекламным фрагментам своих баз. Так, по мнению специалиста одной из таких фирм, доступ к БД через Internet возможен как альтернатива подписке на эти базы на лазерных дисках. Проблему здесь составляет малая скорость поиска, а также юридическая и финансовая основа, на которой такой доступ для подписчиков мог бы быть предоставлен.

Приступая к работе с информационным массивом, необходимо провести его анализ и классифицировать его по отражению тематик [3, 4]. Критериями выбора БД могут служить источники формирования, оперативность отражения материала, частота поступления новой информации, глубина ретроспективы, стоимость работы в БД.

Часто для достижения требуемой полноты информации приходится работать в нескольких БД. Однако различные БД формируются в определенной мере на основе одних и тех же источников, поэтому возможно получение дублей документов. Таким образом, эффективность выбора БД зависит от того, насколько исследователь знаком с их общей характеристикой, а также от технических возможностей доступа к ним и финансовых затрат. В общем виде в режиме on-line стоимость работы в БД складывается из оплаты времени подсоединения к базе хоста и цены каждого полученного документа. Однако в частности каждая БД имеет

свою гибкую систему льгот и скидок для различных категорий пользователей, описание которых не входит в задачу нашей статьи.

### *Характеристика научно-библиографических БД*

Научно-библиографические базы состоят из документов, каждый из которых представляет собой формализованную запись о статье, книге, обзоре и т.д., имеющую определенный формат. Под термином "формат" понимают набор полей, по которым разбита информация в документе. Обычно библиографическая информация содержит такие поля, как фамилии авторов (поле AU), название статьи (поле TI), название источника (поле SO или JL); ключевые слова или индексные термины находятся в поле KW (DE, IW или другое название), поле реферата обозначается как AB, а адрес исполнителей работы – AD. Каждый документ в любой библиографической БД имеет указанный набор полей, а также другие поля, которые отражают специфику БД. Наличие тех или иных полей – важная характеристика базы, помогающая эффективному поиску в ней.

**БД SciSEARCH.** Создателем этой БД является Институт научной информации ISI (Institute for Scientific Information, Филадельфия, США). Источниками служат библиографическая информация и процитированные ссылки из 5000 ведущих научных, технических и медицинских журналов, а также книг, материалов конференций и патентов в области наук о жизни, клинической практики, сельскохозяйственной и пищевой технологии, химии, физики и математики, метеорологии, фармакологии и экологии. SciSEARCH выпускается с 1974 г., пополняется еженедельно и содержит более 13 млн. записей. В настоящее время она доступна как на компактных дисках, так и в режиме on-line через хосты Data-Star, Dialog. С 1995 г. она предоставляется пользователям сети STN International и, следовательно, стала доступной в режиме on-line для российских ученых. Отличительной чертой SciSEARCH является то, что в ее документах имеются ссылки на все работы, процитированные автором оригинальной статьи. Следует отметить, что в цитируемых ссылках приводится фамилия только первого автора, и это надо помнить при определении индекса цитируемости.

Поиск в SciSEARCH можно проводить по цитируемым ссылкам и элементам библиографического описания (автор, год, название журнала, адрес, номер патента, имя патентовладельца, страна), по терминам из заглавия и дополнительным ключевым словам. Благодаря уникальной возможности поиска информации по цитируемости можно запросить все документы, содержащие ссылку на интересующую статью, и таким образом выделить

кластер документов, в определенной степени тематически связанных между собой. При этом для каждой работы на компактном диске будет указана степень связи в виде числа общих библиографических ссылок, причем любая их них может быть использована в качестве исходной для начала нового поиска [5]. Простейший кластер образуется уже в том случае, если в двух документах имеется общая ссылка на одну и ту же работу. Основатель Института научной информации ISI (Филадельфия, США) и идеолог кластерного анализа Е. Гарфилд в своих работах [6, 7] показывает динамику создания новых кластеров. По его мнению, возникновение нового кластера – процесс объективный, он соответствует появлению новой тенденции в данной области знания и позволяет выделить перспективные направления.

После очередной обработки массива информации специальной программой, необходимой для проведения кластерного анализа, каждому новому кластеру присваивается название. Это название представлено в БД только в режиме on-line. В формате документа оно находится в поле RF (Research Fronts). Таким образом, в поле RF представлены названия кластеров, объединяющих статьи, выделенные на основе цитирования. Термины из этих названий также являются поисковыми, что открывает дополнительные возможности для поиска. Каждый кластер имеет четырехзначный номер, перед которым стоит двузначное число, обозначающее год создания информационного массива, обработанного при помощи программы кластерного анализа. По этому номеру можно найти документы, сопряженные в кластер. Сравнивая текст в поле RF по данной тематике за несколько лет, можно видеть, как идет ее развитие. Следует также отметить, что каждая статья при обработке программой кластерного анализа попадает по цитируемости в несколько кластеров, поэтому в БД в режиме on-line каждый документ в поле RF имеет соответственно столько же названий кластеров. БД SciSEARCH на компактных дисках есть в Библиотеке по естественным наукам РАН (выпуски с 1986 г. по настоящее время) и в Институте биохимической химии РАН (выпуски 1987–1990 гг.). Контактные телефоны в библиотеке и институте следующие: 291-11-33 и 330-64-47.

**БД CURRENT CONTENTS.** Создается также Институтом научной информации ISI (США). Также, как и ее печатный аналог, БД CURRENT CONTENTS состоит из семи мультидисциплинарных серий, формат документов которых к настоящему времени значительно усовершенствован по сравнению с первыми выпусками, а именно около 90% документов снабжены рефератами, двумя подуровнями ключевых слов и адресами авторов, что создает возможность непосредственного обращения к автору за репринтом статьи. Каждая се-

рия CURRENT CONTENTS – самостоятельная БД и может быть предметом отдельной подписки. К интересующей нас области знания относятся следующие серии: “Науки о жизни” (1358 журналов), “Сельскохозяйственные науки и науки об окружающей среде” (~900 журналов), “Физика, химия и науки о земле” (~900 журналов), “Клиническая медицина” (~1000 журналов) и т.д. В этих БД можно найти информацию о любой статье в области физико-химической биологии, опубликованной в научном периодическом издании, поскольку все имеющиеся в мире научные журналы этого профиля отнесены хотя бы к одной (или к нескольким одновременно) из перечисленных выше серий. Следует отметить, что в CURRENT CONTENTS поиск по словам из реферата невозможен. Как и в других БД, поиск проводится по всем элементам библиографического описания, а также по словам заглавия и ключевым словам (как авторским, так и дополнительным). База доступна в виде еженедельно обновляемой версии на флоппи-дисках (для DOS и Windows), а перечисленные выше серии CURRENT CONTENTS, начиная с 1993 г., предоставляются подписчикам также в виде версии на лазерных дисках.

Все сведения о возможности подписаться на БД SciSEARCH или любую серию CURRENT CONTENTS можно запросить в европейском отделении института ISI (ISI European Branch, UK) по факсу +44-1895-256710.

Через систему STN International может быть получен доступ к нескольким БД, создаваемым реферативной службой по химии CAS (CHEMICAL ABSTRACTS SERVICE). CAS присваивает всем известным соединениям регистрационные номера, по которым очень удобен поиск в тех базах, где они имеются. Обычно в формате любых БД регистрационные номера CAS находятся в поле RN.

**БД CA (CHEMICAL ABSTRACTS)** является электронным аналогом одноименного печатного издания. БД ведется с 1967 г. и содержит более 12 млн. документов, пополнение базы происходит два раза в месяц. Она создается на основе 9000 журналов, патентов из 26 стран и двух международных патентных организаций, материалов конференций, отчетов, книг и диссертаций. База охватывает все области химии, включая биохимию. Только в STN около 87% документов этой базы снабжено рефератами. В базе CA также возможен поиск по всем элементам библиографического описания, индексным терминам и словам из реферата. Имеется учебный файл, который называется ICA.

**БД CA OLD (CHEMICAL ABSTRACTS OLD)** охватывает период с 1957 по 1966 г. и представлена только в STN. Вместо реферата каждый документ CA OLD содержит набор номеров ссылок, которые

можно найти в печатной версии CHEMICAL ABSTRACTS, а также регистрационные номера CAS упоминаемых в нем химических соединений и тип документа (для патентов). Поиск в базе осуществляется по регистрационным номерам CAS.

**БД CAPREVIEWS** содержит новейшие документы, еще не успевшие попасть в печатное издание CHEMICAL ABSTRACTS. Она обновляется ежедневно, и только через 8–10 недель документы из нее передаются в СА. По своему тематическому содержанию эта база идентична СА и содержит около 100 тыс. документов в формате СА. Эта БД не имеет индексных терминов. В остальном она подобна файлу СА, и поиск в ней осуществляется аналогично.

**БД CAPlus** появилась совсем недавно и охватывает базы СА и CAPREVIEWS, обновляется она ежедневно. Поиск в ней проводится аналогично поиску в базе СА.

**БД CASREACT** содержит с 1985 г. сведения о реакциях органических веществ, в том числе информацию о химических реагентах, растворителях, катализаторах, продуктах одно- и многостадийных синтезов. Источником для нее служат разделы по органической химии печатного издания Chemical Abstracts.

**БД CHEMLIST** содержит с 1979 г. информацию о токсичных химических продуктах. Эту БД, как и описываемую ниже базу REGISTRY, строго говоря, нельзя отнести к библиографическим.

**БД REGISTRY**, создаваемая CAS с 1957 г. и представленная только в сети STN International, содержит около 13,5 млн. записей. Она представляет собой справочный файл по структурам и свойствам органических и неорганических веществ. Каждый документ этой базы имеет регистрационный номер CAS, индексные названия веществ по номенклатуре CAS, наиболее распространенные синонимы, структурную и молекулярную формулы и обозначения класса веществ, причем поиск может быть произведен по любому из перечисленных выше элементов. В этой базе имеются классификационные коды для поиска, например, белков и нуклеиновых кислот. БД позволяет найти и вывести на экран общее число ссылок на данное вещество в базах СА, CAOLD, CAPREVIEWS и CAPlus. В REGISTRY возможен показ 10 самых новых библиографических документов по искомому соединению из базы СА. Кроме того, в документах REGISTRY имеется поле LC, где перечислены другие БД, доступные через систему STN, содержащие данные по искомому соединению (см. рисунок). Поисковое предложение было сформулировано так, как оно записано в основном поле CN. Тот же вид документа мы получили бы, если бы вели поиск по регистрационному номеру CAS (поле RN) или по синониму этого соединения

(дополнительное поле CN). Уникальность REGISTRY состоит в том, что она позволяет определить наличие соединения с заданной структурой, если оно описано ранее. Эта база имеет учебный файл LREGISTRY, созданный для тренировок, поскольку поиск по структурам в базе REGISTRY требует опыта и значительных финансовых затрат.

**БД MEDLINE** создается Национальной медицинской библиотекой США (U.S. National Library of Medicine). В режиме on-line эта БД имеет ретроспективу с 1966 г. и содержит более 8 млн. документов. В настоящее время база комплектуется на основе более чем 3300 журналов, издаваемых в 70 странах, при этом перечень журналов все время пополняется. MEDLINE содержит библиографические описания и расширенные рефераты публикаций в области биологии, в частности микробиологии, экологии, проблемам создания лекарственных препаратов, гигиены, питания, образования в области медицины. Поиск в БД можно осуществлять по элементам библиографического описания, индексным терминам (поле MESH), словам реферата и регистрационным номерам CAS. В режиме on-line имеется тезаурус для поиска по словам из полей MESH (термины медицинских предметных разделов), СТ (контролируемые термины) и CN (химические названия). Кроме того, для учебной работы с информацией в режиме on-line имеется небольшой фрагмент этой БД, который называется LMEDLINE. В Москве БД MEDLINE доступна как на лазерных дисках (например, в Медицинской библиотеке и некоторых исследовательских институтах), так и в режиме on-line, в том числе по системе STN International. В режиме on-line эта БД пополняется еженедельно, на лазерных дисках – ежемесячно. Кроме того, это одна из самых дешевых БД в режиме on-line, чем отчасти и объясняется ее высокая популярность.

**БД EMBASE** (или EXCERPTA MEDICA) также содержит сведения медико-биологического и фармацевтического профилей, но в отличие от БД MEDLINE состоит из источников, две трети объема которых составляют работы, выполненные не в США. При этом треть объема БД – это работы, выполненные не в англоязычных странах [8]. Эта БД создается в Нидерландах научным издательством Elsevier с 1974 г. В настоящее время в ней содержится около 6 млн. документов, пополняется БД еженедельно, формируется на основе 4500 журналов из 110 стран. Кроме того, она содержит сведения о трудах конференций, диссертациях, монографиях и научных отчетах. Получив информацию по определенному вопросу из БД MEDLINE, можно найти еще 35–40% новой информации, если обратиться с этим же запросом в EMBASE. Она доступна как на лазерных дисках, так и в режиме on-line, где для поиска по полю СТ имеется словарь контролируемых терминов.



поле BC, которое позволяет получить информацию по таксономическим группам организмов. Особенности поиска в этой БД представлены в работе [9]. Следует отметить, что, работая с одним и тем же запросом в базах BIOSIS Previews и MEDLINE, можно обнаружить около 60–70% общих документов. BIOSIS Previews выпускается также и на компактных дисках, однако в России пока ни одна организация не получала ее систематически.

**БД NTIS** (National Technical Information Service) может оказаться полезной в ряде случаев. Она содержит рефераты научно-технических и аналитических отчетов об исследованиях и разработках, финансируемых правительствами США и других стран с 1964 г. NTIS создается Национальной службой технической информации при министерстве торговли США (National Technical Information Service, U.S. Department of Commerce), поставляется потребителю на оптических дисках и доступна через многие хосты, в том числе по системе STN International. В NTIS вводится информация о научно-исследовательских работах, проводимых правительственными агентствами, академическими институтами и фирмами разных стран. Эта БД охватывает все области науки и техники, сельского хозяйства, медицины и экономики. Она соответствует печатному изданию Government Reports Announcements & Index и частично – еженедельному бюллетеню Abstracts Newsletters. Первоисточники, ссылки на которые встречаются в базе, практически недоступны для наших пользователей. К настоящему времени БД имеет около 1.8 млн. документов, пополнение происходит каждые две недели. Поиск в NTIS можно проводить по библиографической информации, индексным терминам и словам из реферата. В Международном центре научно-технической информации есть БД NTIS за 1985–1989 гг. на лазерных дисках.

**БД LIFESCI** доступна через STN International, она создается информационной фирмой Cambridge Scientific Abstracts (США) с 1978 г. Она содержит литературу в области медицины и биологии, включая микробиологию, биохимию, генетику, иммунологию, биотехнологию, а также сведения по некоторым аспектам сельскохозяйственной и ветеринарной тематик. Источниками этой базы являются публикации научных журналов, труды конференций, научные отчеты, монографии и патенты. Поиск в LIFESCI можно проводить по элементам библиографического описания, индексным терминам и словам из заглавия и реферата.

**БД САВА** предназначена для исследователей, работающих над проблемами, связанными с растениеводством, лесным хозяйством и ветеринарией, она охватывает периодическую литературу с 1979 г. и содержит около 3 млн. документов. Ее создает фирма SAV International (Великобрита-

ния). Источниками служат журналы, книги, научные отчеты, диссертации, труды конференций и патенты. Поиск в ней можно проводить по элементам библиографического описания, ключевым словам и словам из заглавия и реферата.

В 80-х годах появляется новый тип БД, содержащих полный текст статей из журналов. Обычно в этих базах для поиска доступна вся информация, кроме рисунков, химических и математических уравнений. В этой связи следует упомянуть о двух БД, создаваемых Американским химическим обществом.

**БД CEN** содержит с 1991 г. полные тексты статей из журнала Chemical and Engineering News.

**БД CJACS** с 1982 г. содержит полные тексты 16 журналов, причем к интересующей нас научной области относится большая их часть, а именно: Analytical Chemistry, Biochemistry, Bioconjugate Chemistry, Biotechnology Progress, Chemical Research in Toxicology, Environmental Science and Technology, Journal of Agriculture and Food Chemistry, Journal of the American Chemical Society, Journal of Medicinal Chemistry, Journal of Pharmaceutical Chemistry, Journal of Organic Chemistry и др.

В дальнейшем полнотекстовые БД по некоторым издаваемым ими журналам стало создавать Королевское химическое общество Великобритании (с 1987 г.) и такие знаменитые научные издательства, как John Wiley & Sons (с 1987 г.) и Elsevier (с 1990 г.).

В следующей части этого обзора будут представлены фактографические и научно-библиографические БД, интересные для исследователей, работающих в области физико-химической биологии и биотехнологии, в том числе патентные и содержащие информацию, необходимую для принятия маркетинговых решений.

## СПИСОК ЛИТЕРАТУРЫ

1. STN Database Catalog, FIZ Karlsruhe, October 1995. P. 43.
2. Международная сеть научно-технической информации STN International. Базы данных STN (краткое описание и кластеры). М.: МПП "Биоинформсервис", 1995. С. 61.
3. Шкаренкова Л.С. // Сб. докл. 3-й конференции "Автоматизация информационного обслуживания". София, 1987. С. 89–94.
4. Шкаренкова Л.С. // Матер. VII Всесоюз. школы патентоведов, изобретателей и актива ВОИР учреждений АН СССР. Пушкино, 1990. С. 50–56.
5. Глушановский А.В., Каленов Н.Е., Лексикова Е.Е. // Институт научной информации (Филадельфия, США). База данных "Science Citation

- Index" на CD-ROM. М.: МПП "Биоинформсервис", 1993. С. 37.
6. ISI Atlas of Science: Biochemistry and Molecular Biology 1978/80. Philadelphia: ISI, 1981. P. 540.
  7. ISI Atlas of Science: Biotechnology and Molecular Genetics 1981/82. Covering 127 Research Front Specialities Including 1983/84. Suppl. Philadelphia: ISI, 1981. P. 540.
  8. Старикова Л.Н. Поиск биомедицинской информации в зарубежных базах данных. М.: ВНИИПАС, 1987. С. 54.
  9. Палева Н.С. // Международная сеть научно-технической информации STN International. Поиск информации по биологии и биомедицине в базе данных BIOSIS Preview/RN. М.: МПП "Биоинформсервис", 1994. С. 42.

## Scientific Informational Support for Research in Physicochemical Biology (Analytical Review)

L. S. Shkarenkova, T. T. Orlovskaya, and T. V. Ovchinnikova

*Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences,  
ul. Miklukho-Maklaya 16/10, GSP-7 Moscow, 117871, Russia*

**Abstract**—A description is provided of the main scientific reference and bibliographical information databases that hold interest for scientists working in the field of physicochemical biology and can provide rapid access and an efficient search for the necessary information.

*Key words: informational support, scientific information databases, access, search.*