

Министерство образования Республики Беларусь
Белорусский государственный университет

А. А. Рагойша

Поиск информации о структуре химического вещества
в онлайн-базах данных

Практикум по курсу "Информационные технологии в химии"
для студентов специальности
1-31 05 01 Химия (по направлениям)

Минск
2013

УДК 54:004.65(076.5)(075.8)

Рекомендовано к изданию
Учебно-методической комиссией химического факультета
Белорусского государственного университета
22 мая 2013 г., протокол № 4

Рецензенты:

кандидат химических наук, доцент кафедры электрохимии
Белорусского государственного университета *Ю. С. Головкин*,

кандидат химических наук, доцент кафедры химии Белорусского
государственного педагогического университета имени Максима Танка
А. С. Тихонов

Рагойша, А. А.

Поиск информации о структуре химического вещества в онлайн-базах данных [Электронный ресурс] : практикум по курсу "Информационные технологии в химии" для студентов спец. 1-31 05 01 Химия (по направлениям) / А. А. Рагойша. — Минск: БГУ, 2013. — 101 с. : ил. — Библиогр.: с. 100.

Методическое руководство к практическим занятиям по разделу "Химическая структура" учебного курса "Информационные технологии в химии". В пособии рассматриваются наиболее важные веб-ресурсы, а также алгоритмы целенаправленного поиска, извлечения и анализа научной химической информации.

Для студентов и магистрантов химического факультета БГУ.

УДК 54:004.65(076.5)(075.8)

© А.А. Рагойша, 2013

СОДЕРЖАНИЕ

ПРЕДИСЛОВИЕ	5
1. ВВЕДЕНИЕ	6
2. КРИСТАЛЛОГРАФИЧЕСКАЯ ИНФОРМАЦИЯ В WWW	7
2.1. Введение	7
2.2. Mercury — специализированная программа визуализации кристаллографической информации	7
2.2.1. Подготовка к работе	7
2.2.2. Mercury : Краткая инструкция.....	8
2.2.3. Тренировочные упражнения	13
2.3. Примеры источников кристаллографической информации	13
2.3.1. Inorganic Crystal Structure Database (ICSD)	14
2.3.2. Crystallography Open Database (COD)	16
2.3.3. Журнал Acta Crystallographica Section E: Structure Reports Online	18
3. СТРУКТУРНЫЙ ПОИСК. ФОРМИРОВАНИЕ ЗАПРОСА С ПОМОЩЬЮ АППЛЕТА	19
3.1. Структурный поиск.....	19
3.2. Контрольное задание 1. Расчет коэффициента Танимото	19
3.3. Апплет	20
3.4. Апплет NIST Molecular Editor	21
3.5. Апплет JME	24
3.6. Апплет MarvinSketch.....	25
3.7. Знакомство с элементами семантического веба.....	28
4. ЛИНЕЙНЫЕ НОТАЦИИ. CAS RN, SMILES, InChI.....	29
4.1. Регистрационные номера химических веществ.	29
4.2. CAS Registry Number и справочник Common Chemistry.....	29
4.3. CAS RN в базах данных.	33
4.4. SMILES	35
4.5. InChI и InChIKey	36
5. КАТАЛОГИ ХИМИЧЕСКИХ РЕАКТИВОВ	39
6. ОРГАНИЧЕСКИЙ СИНТЕЗ: ОНЛАЙНОВЫЕ РЕСУРСЫ.....	40
6.1. Введение	40
6.2. ChemSynthesis	40
6.3. Organic Chemistry Portal	42
6.4. ChemSpider SyntheticPages.....	43
6.4.1. Структура базы данных	43
6.4.2. Апплет ACD/Labs SDA на сайте ChemSpider.....	45
6.5. Organic Syntheses	47
6.6. Послесловие: Spotlight Reagents.	53
7. СПЕКТРАЛЬНЫЕ БАЗЫ ДАННЫХ: ЯМР	54
7.1. Краткое описание метода ЯМР-спектроскопии.	54
7.2. ЯМР-спектры в ChemSpider	58
7.3. Справочный сайт Organic Chemistry Info.....	59
7.4. The Sadtler Spectral Handbooks.....	61
7.5. NMRShiftDB2	62
7.6. SDBS — Spectral Database for Organic Compounds.....	64
8. СПЕКТРАЛЬНЫЕ БАЗЫ ДАННЫХ: МАСС-СПЕКТРОМЕТРИЯ	69
8.1. Схема метода масс-спектрометрии.	69

8.2. Масс-спектры в NIST Chemistry WebBook.....	71
8.3. Масс-спектры в Spectral Database for Organic Compounds (SDBS).....	74
9. СПЕКТРАЛЬНЫЕ БАЗЫ ДАННЫХ: ИНФРАКРАСНАЯ (ИК) СПЕКТРОСКОПИЯ.....	77
9.1. Схема метода ИК-спектроскопии (Infra-Red (IR) Spectroscopy).....	77
9.2. Онлайн-базы данных.....	78
10. SPECTRAL GAME.....	79
11. ПАКЕТ ПРОГРАММ ACD/Labs FREEWARE	80
11.1. ACD/ChemSketch как молекулярный редактор двумерных химических структур.	80
11.1.1. Введение	80
11.1.2. Работа в режиме Structure	81
11.1.3. Режим Structure. Контрольные задания	87
11.4. ACD/ChemSketch как графический редактор.....	89
11.1.5. Режим Draw. Контрольные задания	91
11.2. ACD/3D Viewer	92
11.3. Файлы в формате MOL	94
12. ПРИЛОЖЕНИЕ. ИССЛЕДОВАТЕЛЬСКАЯ КОНТРОЛЬНАЯ РАБОТА	95
ЛИТЕРАТУРА	100
BIBLIOGRAPHIC RECORD.....	101

ПРЕДИСЛОВИЕ

Данное пособие содержит материалы второго раздела практикума учебного курса "Информационные технологии в химии".

Цель этого этапа работы — познакомиться с некоторыми важнейшими структурными базами данных, попытаться оценить их достоинства и недостатки, определить области целесообразного их использования в научной практике, изучить приемы ведения структурного информационного поиска и анализа извлекаемых результатов.

Основная часть структурной информации находится в области "скрытого" Интернета, поэтому для обнаружения необходимых сведений следует применять не универсальные поисковые системы, а специализированные поисковые программы. Как правило, каждая база данных имеет свой поисковый инструмент, который, кроме общепринятых элементов, может содержать в себе весьма специфические детали. В каждой главе пособия приводятся методические руководства по работе с соответствующими типами программ; в тех случаях, когда такая инструкция отсутствует, студент обязан самостоятельно разобраться с алгоритмом функционирования программы, основываясь на своем предыдущем опыте.

В ходе практикума обсуждаются простейшие задачи, с которыми особенно часто встречается химик в своей повседневной работе — поиск информации о структуре и свойствах вещества, идентификация индивидуального вещества по спектральным данным, письменное оформление химически значимого текста. Предполагается, что навыки освоения таких операций должны быть доведены до автоматизма, требующегося при решении комплексных проблем на старших курсах обучения.

Главы, посвященные спектральным базам данных, содержат краткие описания соответствующих методов исследования.

Терминология хемоинформатики рассмотрена в лекционном курсе (презентации лекций размещены на сайте [ABC Chemistry — Азбука веб-поиска для химиков](http://www.abc.chemistry.bsu.by/), <http://www.abc.chemistry.bsu.by/>).

Интернет — среда чрезвычайно изменчивая, поэтому материалы практикума будут постоянно обновляться. Самую последнюю версию данного учебного пособия (в HTML-формате) можно найти на сайте **ABC Chemistry** по адресу: <http://www.abc.chemistry.bsu.by/structure/>.

1. ВВЕДЕНИЕ

С точки зрения классификации, любая база данных, содержащая сведения о топологии и/или топографии структуры химического вещества, является *структурной*.

В литературе термин **структурная база данных** (*Structural database*) часто применяют в более узком смысле — так называют только те массивы, где хранится информация о *молекулярной структуре*. Возможность формулирования *структурного запроса* считается обязательным элементом таких ресурсов.

В **кристаллографических базах данных** (*Crystallographic database*) хранится информация о *трехмерной структуре кристаллических* веществ молекулярного и немолекулярного строения.

В **спектральных базах данных** (*Spectral database*) основным информационным компонентом является спектр химического вещества. Наиболее солидные базы данных этого типа позволяют пользователю проводить и прямое, и обратное соотнесение *вещество ↔ спектр*.

Базы данных химических реакций (*Reaction database*) предназначены для решения задач *реагент → продукт* и *продукт → реагент*.

Границы между разными типами баз данных весьма условны. База данных может содержать одновременно сведения о молекулярной и кристаллической структурах, спектры, численные значения физико-химических величин, текстовые описания химических свойств. Эту неоднозначность следует учитывать при работе с предметными каталогами, где нужна пользователю структурная база данных может оказаться в одной из многочисленных категорий (*Structural databases, Spectral, Factual, Data, Property, General, Specific Classes* и т. д.).

Хорошими стартовыми точками при поиске структурных баз данных являются метасайты, например:

[Chemical Information Sources/SIRCh](http://en.wikibooks.org/wiki/Chemical_Information_Sources/SIRCh)

http://en.wikibooks.org/wiki/Chemical_Information_Sources/SIRCh

[Organic Chemistry Resources Worldwide](http://www.organicworldwide.net/)

<http://www.organicworldwide.net/>

[Sixty-Four Free Chemistry Databases](http://depth-first.com/articles/2011/10/12/sixty-four-free-chemistry-databases/)

<http://depth-first.com/articles/2011/10/12/sixty-four-free-chemistry-databases/>

[Open Directory](http://www.dmoz.org/Science/Chemistry/Chemical_Databases/)

http://www.dmoz.org/Science/Chemistry/Chemical_Databases/

[Spectra & Spectral Data](http://libweb1.lib.buffalo.edu/guide/guide.asp?id=101)

<http://libweb1.lib.buffalo.edu/guide/guide.asp?id=101>

2. КРИСТАЛЛОГРАФИЧЕСКАЯ ИНФОРМАЦИЯ В WWW

2.1. Введение

Экспериментальные сведения из кристаллографических баз данных мы будем использовать для исследования структуры вещества.

В тех базах данных, с которыми познакомимся, кристаллографическая информация хранится в стандартных *CIF*-файлах.

CIF-файл можно прочитать, открыв его в обычном текстовом редакторе "Блокноте" (*Notepad*).

Более удобны для пользователя специализированные программы — они позволяют отобразить на экране трехмерную структуру вещества, а также рассчитать некоторые молекулярные параметры.

2.2. Mercury — специализированная программа визуализации кристаллографической информации

2.2.1. Подготовка к работе

а) Загрузка учебных файлов.

Создайте папку для временного хранения учебных файлов.

Существенная деталь: ни в названии этой папки, ни в описании пути к этой папке **не должно быть кириллических букв и пробелов**.

Учебные *CIF*-файлы находятся по адресу:

<http://abc.chemistry.bsu.by/2/1-cif.zip>.

Скопируйте архив на свой компьютер и разархивируйте его. Полученные файлы (*1.cif*, *2.cif*, *3.cif*) вы будете использовать при изучении функций программы *Mercury*.

По окончании занятия папку с временными файлами удалите.

б) Загрузка программы:

Ярлык *Mercury* находится на Рабочем столе вашего компьютера.

Самостоятельная работа.

Загрузите *Mercury* и изучите основные возможности программы.

Для справки:

Инсталлятор программы *Mercury* находится по адресу:

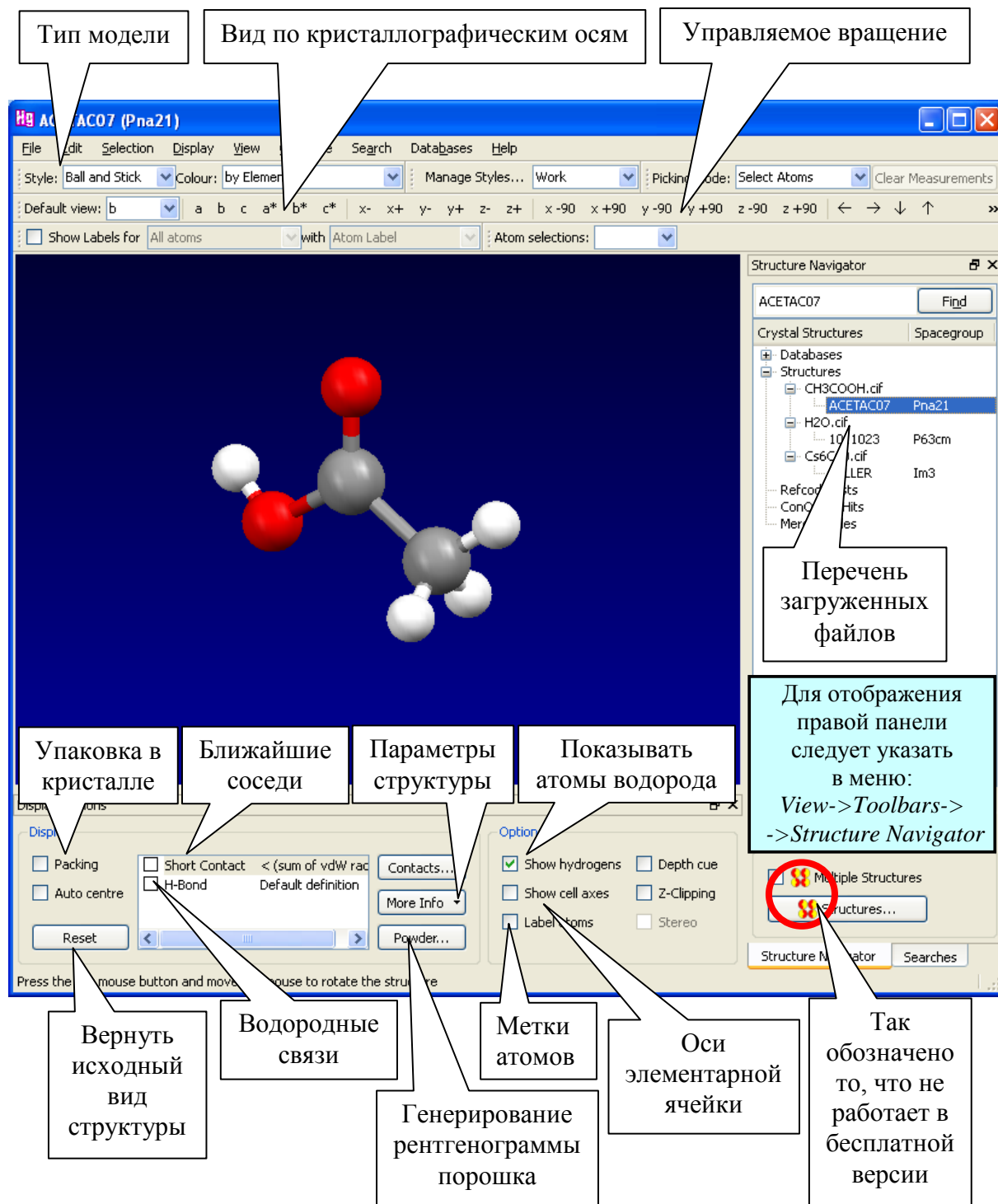
http://www.ccdc.cam.ac.uk/free_services/mercury/downloads/

2.2.2. Mercury: Краткая инструкция

Откройте файлы *1.cif*, *2.cif*, *3.cif* (*File -> Open...*).

Поэкспериментируйте с программой, чтобы изучить ее возможности.

Вид основного рабочего окна программы *Mercury*:



Порядок выполнения отдельных операций

Управление видом структурной модели

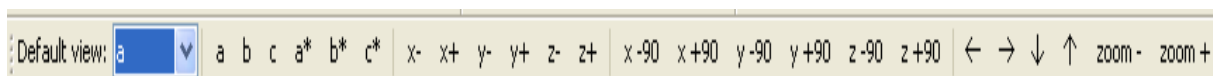
а) с помощью мышки.

Установите курсор в поле, где находится модель.

- **Вращение в трехмерном пространстве:** удерживая **левую** клавишу мышки в нажатом состоянии, передвигайте курсор по горизонтали и по вертикали.
- **Вращение в плоскости экрана:** при нажатой клавише **Shift** и удерживая **левую** клавишу мышки в нажатом состоянии, передвигайте курсор по горизонтали.
- **Увеличение-уменьшение:** удерживая **правую** клавишу мышки в нажатом состоянии, передвигайте курсор по вертикали.
- **Перемещение по полю:** при нажатой клавише **Ctrl** и удерживая **левую** клавишу мышки в нажатом состоянии, передвигайте курсор в нужном направлении.

б) с помощью кнопок

Нажимайте соответствующие кнопки на панели:

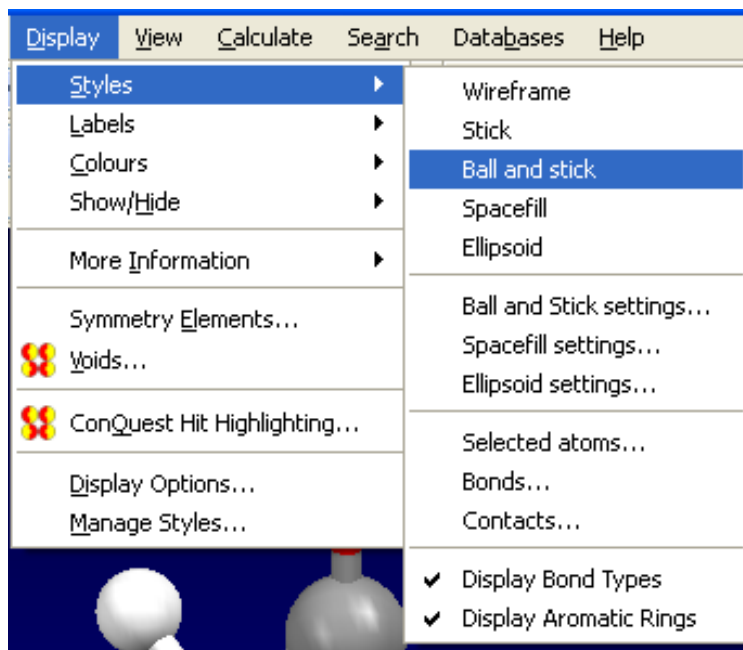
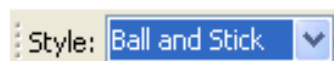


Выбор типа модели (стиль):

а) через основное меню

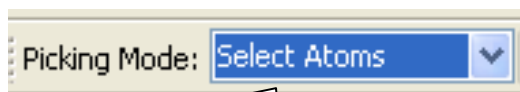


б) из списка Style (находится на верхней панели)

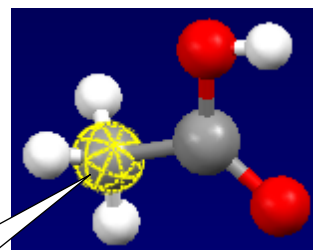


Выделить атом: щелкнуть по атому.

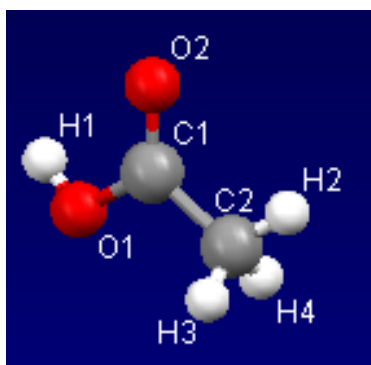
Снять выделение: повторно щелкнуть по выделенному атому.



Что именно выделяем — указываем в меню, расположенном в верхней панели



Выделенный атом

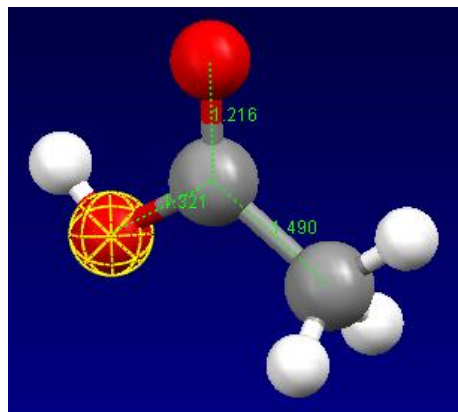
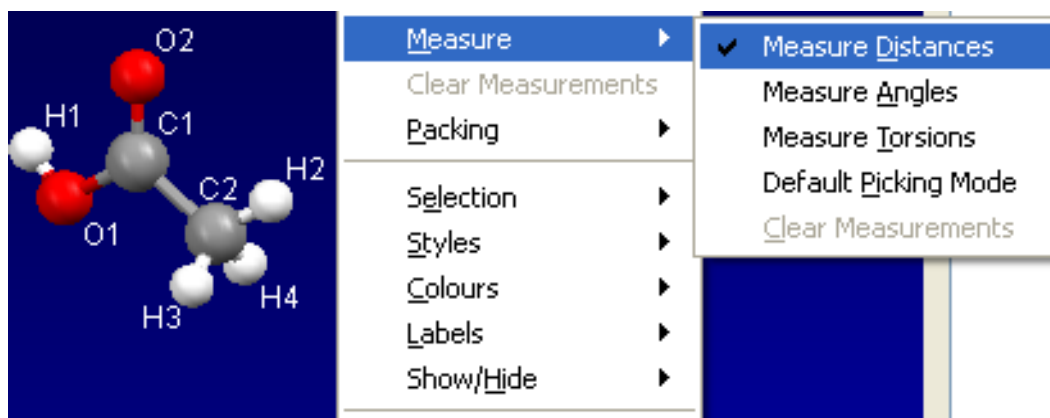


Указать химические символы и номера атомов —

использовать выключатель *Label atoms* (находится на нижней панели)

Измерение отдельных геометрических параметров

1. Откройте контекстное меню и выберите требуемый параметр.

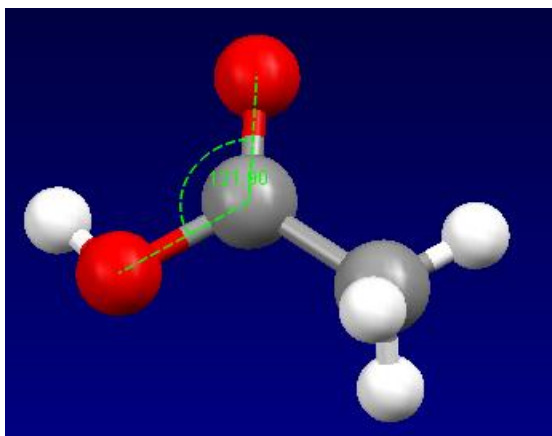


2.

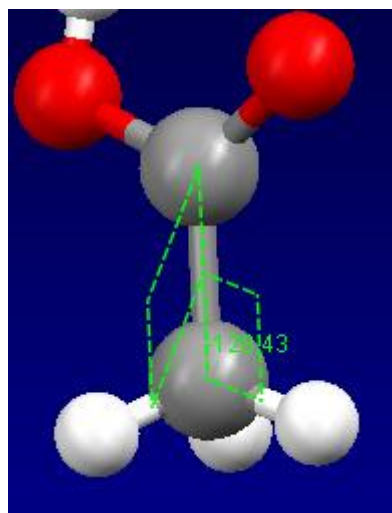
Для измерения **длины связи** (*Measure Distances*) — щелкните по двум связанным атомам.

Совет: структуру стоит развернуть в наиболее удобное положение и приблизить ее к зрителю.

Примечание: Так же можно измерить расстояние и между несвязанными атомами.



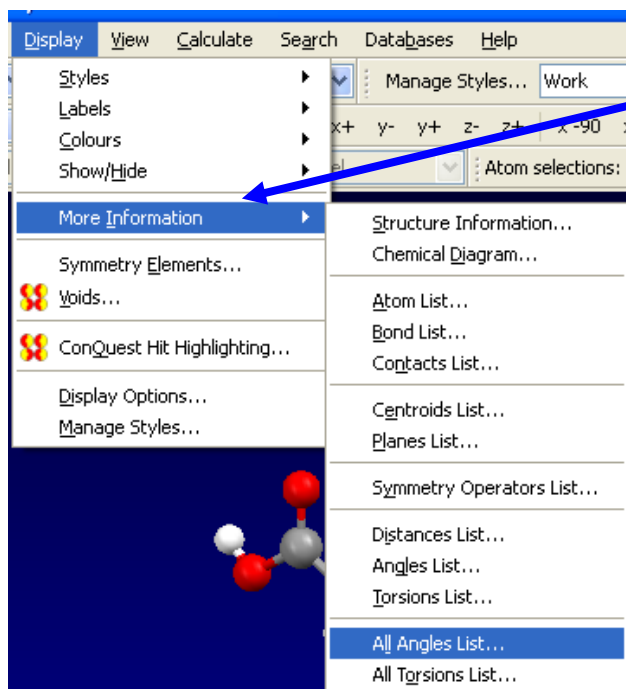
Для измерения **валентного угла между связями** (*Measure Angles*) — щелкните поочередно по трем атомам.



Для измерения **двугранного (торсионного) угла** (*Measure Torsions*) — щелкните поочередно по четырём атомам. Будет измерен угол между плоскостью, в которой лежат атомы 1-2-3, и плоскостью, в которой лежат атомы 2-3-4.

Удаление измеренных геометрических параметров с экрана:
Clear Measurements — команда в контекстном меню и кнопка на верхней панели.

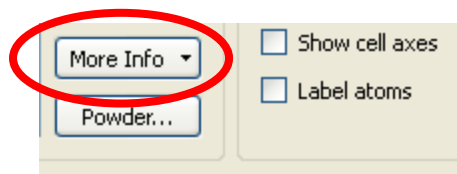
Получение таблиц с числовыми значениями структурных параметров



а) Через меню:

Display -> More Information -> ...

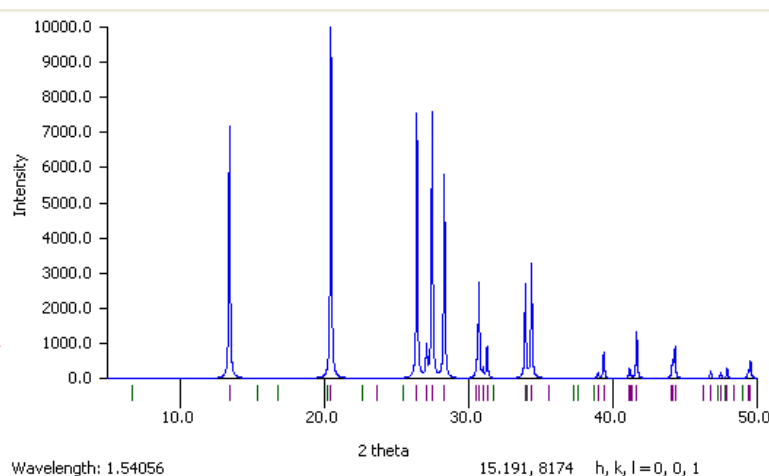
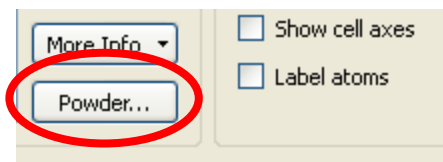
б) Кнопкой *More Info* (расположена на нижней панели)



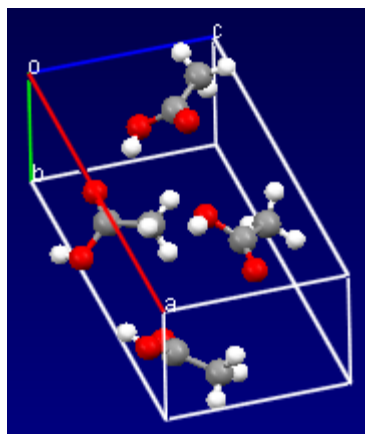
Как правило, наиболее полная информация находится в таблицах *Structure* (кристаллографические параметры), *Bond List* (межъядерные расстояния), *All Angles List* (валентные углы между связями), *All Torsions List* (двугранные углы).

Генерирование рентгенограммы

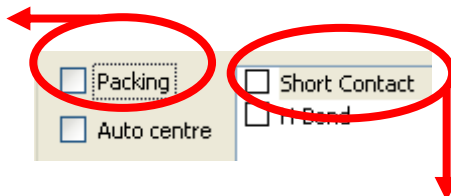
Щелкните по кнопке Powder (нижняя панель)



Пространственное заполнение



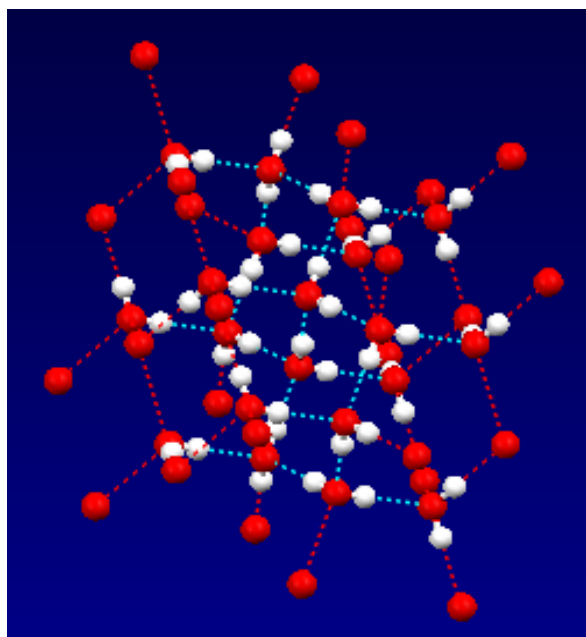
Инструменты находятся на нижней панели.



Пунктир указывает на близко расположенные атомы соседних молекул (межъядерное расстояние меньше суммы Ван-дер-Ваальсовых радиусов).

Голубой цвет пунктира — если соседняя молекула отображена целиком.

Красный цвет пунктира — если отображен только один атом соседней молекулы.



2.2.3. Тренировочные упражнения

Проанализируйте кристаллические структуры, хранящиеся в файлах *1.cif* (уксусная кислота), *2.cif* (вода), *3.cif* (соединение цезия и фуллерена).

1. Определите межъядерные расстояния, углы между связями, двугранные углы в молекуле уксусной кислоты, находящейся в кристаллической фазе.

Ответы для самопроверки:

$$d(\text{C}=\text{O}) = 1,216 \text{ \AA}$$

$$d(\text{C}-\text{O}) = 1,321 \text{ \AA}$$

$$\angle \text{OCO} = 121,90^\circ$$

$$\text{Угол между плоскостями HOC и OCO: } 8,32^\circ$$

2а. Определите межъядерные расстояния в молекулах воды в кристаллической фазе.

Ответ для самопроверки:

$$d(\text{O}-\text{H}) = 0,867 \text{ \AA}; 0,821 \text{ \AA} \text{ и } 0,862 \text{ \AA}$$

2б. Чему равна длина водородной связи $\text{H}(3)\dots\text{O}(1)$?

Ответ для самопроверки:

$$d(\text{O}\dots\text{H}) = 1,897 \text{ \AA}.$$

3. Оцените диаметр молекулы фуллерена.

Ответ для самопроверки:

$$D = 7,14\dots 7,15 \text{ \AA}$$

2.3. Примеры источников кристаллографической информации

Самостоятельная работа.

1. Изучите приемы ведения информационного поиска в каждом из перечисленных ниже источников.

2. Выполните проверочные задания, указанные в инструкциях:

- извлеките нужные *CIF*-файлы;
- проанализируйте особенности структуры заданных веществ;
- запишите (на бумаге) численные значения требуемых молекулярных параметров;
- предъявите преподавателю для проверки.

CIF-файлы можно удалять с диска только после получения зачета по соответствующему этапу работы.

2.3.1. Inorganic Crystal Structure Database (ICSD)

Inorganic Crystal Structure Database (<http://icsd.ill.eu/icsd/index.php>) — наиболее полный в мире архив кристаллографической информации о неорганических веществах.

В WWW имеется бесплатный доступ к части этой базы данных — к *Demo database*.

На основном сайте (<http://icsd.fiz-karlsruhe.de/>) для работы с *Demo database* требуется предварительная бесплатная регистрация. Мы же воспользуемся предпоследней версией сайта — здесь регистрация не требуется.

Краткая инструкция

При обращении по адресу <http://icsd.ill.eu/icsd/index.php> на экран выводится поисковый бланк:

Поиск можно проводить по многим параметрам; мы сконцентрируем наше внимание на некоторых из них:

- **Chem/Mineral Name** — поиск по названию химического вещества или минерала;
- **Elements** — поиск по элементному составу;
- **Element Count** — число химических элементов в искомом веществе.

Примеры заполнения поля *Elements* (символы химических элементов записываются через пробелы):

Cu O — В **брутто**-формуле должны присутствовать Cu и O.

Y Ba₂ Cu₃ O — В **брутто**-формуле должны присутствовать Y, Ba, Cu, O, причем должно быть 2 атома Ba и три атома Cu.

Zn "P₂ O₇" — В **брутто**-формуле должны присутствовать Zn, P, O, причем в **структурной** формуле должна быть группировка P₂O₇.

При щелчке по названию графы бланка на экран выводится краткая инструкция по заполнению данного поля (с примерами).

После заполнения поискового бланка и нажатия кнопки *Search* появляется окно авторизации, где следует нажать кнопку *Demo*.

Так выглядит фрагмент списка результатов поиска:

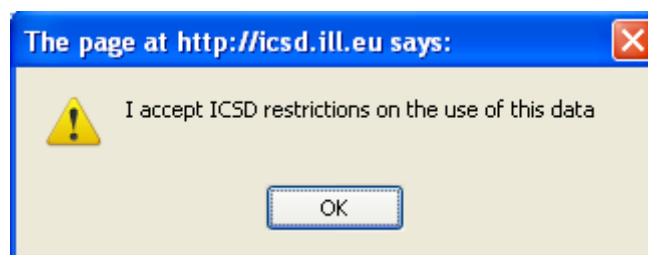
1. Отметить нужную строку галочкой

2. Загрузить CIF-файл

Year	Authors	Title	Struct. Formula	sg	Mineral
<input type="checkbox"/> 1980	Guth, H.; Heger, G.; Klein, S.; Treutmann, W.; Scheringer, C.;	Strukturverfeinerung von Harnstoff mit Neutronenbeugungsdaten bei 60, 123, 293 K und X-N- und X-X (1S2)-Synthesen bei etwa 100 K	C O (N H ₂) ₂	P4-21M	Urea

Page : [1](1 results) 10 results per page.

После вызова CIF-файла потребуется подтвердить, что пользователь не будет нарушать правила работы с базой данных:



Внимание! Прежде чем записать новый запрос на поисковом бланке, следует очистить бланк нажатием кнопки **Reset**.

Контрольное задание 1а. Поиск информации по названию вещества.

Найдите информацию о структуре мочевины в кристаллическом состоянии. (мочевина = urea).

Проанализируйте структуру и определите молекулярные параметры:

- длины связей CN, CO, NH;
- валентные углы OCN, NCN, CNH, HNH;
- двугранный угол между плоскостями, в которых лежат, соответственно, атомы H, N, C и N, C, O.

Контрольное задание 1б. Поиск по элементному составу.

Найдите информацию о структуре гидрата оксалата аммония.

Подсказка: в поисковом бланке заполните графы *Elements* и *Element Count*.

- Под каким углом карбоксильные группы развернуты по оси C—C друг относительно друга?
- Какова форма иона аммония?
- Каков валентный угол HOH в молекуле воды?

2.3.2. Crystallography Open Database (COD)

[Crystallography Open Database \(COD\)](http://www.crystallography.net/) (<http://www.crystallography.net/>) — большая и растущая бесплатная база данных. Она содержит материал из *American Mineralogist Crystal Structure Database* и постоянно пополняется структурами, которые присылают авторы недавно опубликованных работ.

Поисковый бланк расположен на странице:

<http://www.crystallography.net/search.html>

text (1 or 2 words)	<input type="text"/>
journal	<input type="text" value="acta"/>
year	<input type="text"/>
volume	<input type="text"/>
issue	<input type="text"/>
1 to 8 elements	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
NOT these elements	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
volume min and max	<input type="text"/> <input type="text"/>
number of distinct elements min and max	<input type="text"/> <input type="text"/>
filters	<input type="checkbox"/> has Fobs
<input type="button" value="Reset"/>	<input type="button" value="Send"/>

Запрос может состоять из слов и химических символов.

Некоторые особенности **заполнения поискового бланка**:

- В текстовых полях можно указывать название химического вещества или минерала, а также отдельные фрагменты библиографического описания той статьи, в которой был опубликован CIF-файл.

- Каждое поле групп "**1 to 8 elements**" и "**NOT these elements**" предназначено для указания только одного химического элемента.

В случае однобуквенных химических символов (таких как O, C и др.) рекомендуют ставить один пробел перед символом и один пробел после — тогда уменьшается вероятность неадекватной работы программы.

- Можно указывать число атомов элемента в формульной единице.

На сайте говорится, что между химическим символом и числом атомов должен быть пробел. Как показывает практика, иногда такой пробел не способствует, а препятствует поиску.

- В графах "number of distinct elements, min and max" указывают минимально и максимально допустимые числа химических элементов в каждом из извлекаемых веществ.

Результаты поиска содержат некоторые кристаллографические характеристики, библиографическое описание публикации — источника структуры — и ссылку на CIF-файл, например:

CIF file	Formula : - H 2.00 O 1.00 - Comments : McFarlan, R L Journal of Chemical Physics The Structure of Ice II Journal of Chemical Physics (1936) 60 64 Space group :C 2 2 21 Cell volume: 195.20 Cell parameters:7.8000; 4.5000; 5.5600; 90.000; 90.000; 90.000;
----------	--

Контрольное задание 2а. Структура пероксида водорода.

Извлеките файл с кристаллической структурой пероксида водорода.

- Чему равен двугранный (торсионный) угол между плоскостями $H_{(1)}OO$ и $H_{(2)}OO$ в молекуле пероксида водорода?
- Одинаковы ли длины связей $H-O$? Чему они равны ?
- Каковы длины водородной связи $O...H-O$? (Измерьте в пределах одной элементарной ячейки).

Контрольное задание 2б. Лед под высоким давлением.

Извлеките файлы со структурами льда II и льда III (лед = ice).

- Каковы длины связи $H-O$ в обеих кристаллических структурах?
- Каковы валентные углы HOH в обеих кристаллических структурах?

2.3.3. Журнал Acta Crystallographica Section E: Structure Reports Online

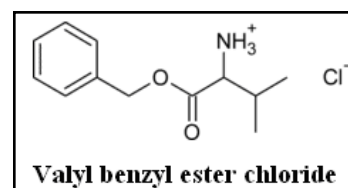
[Acta Crystallographica Section E: Structure Reports Online](http://journals.iucr.org/e/journalhomepage.html) (<http://journals.iucr.org/e/journalhomepage.html>) — пример онлайн-журнала, содержащего кристаллографическую информацию.

Это официальное издание Международного союза кристаллографии (*International Union of Crystallography*).

В журнале полностью открыт весь материал начиная с 2008 г., а также статьи категории *Open Access* за период 2004-2007 гг.

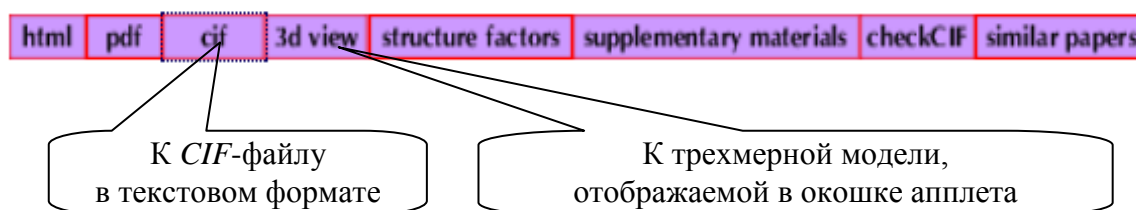
Изучение сайта и проверочное задание

Наша задача — найти в журнале и проанализировать структуру следующего вещества:



Перейдите к поисковому бланку, сформулируйте запрос. В результатах поиска выберите нужную статью.

Вы увидите подобный навигационный блок:



Щелкнув по кнопке **cif**, выведете на экран *CIF*-файл в **текстовом формате**. Ознакомьтесь со строением *CIF*-файла.

Контрольное задание 3.

Найдите в *CIF*-файле те информационные блоки, в которых содержится количественная информация о структуре исследованного сложного эфира.

Ответьте на следующие вопросы:

- Чему равны длины связей C—O и C=O в сложноэфирной функциональной группе?
 - Чему равен валентный угол в структурном фрагменте O—C=O ?
 - Справедливо ли утверждение, что группировка NH₃⁺ имеет форму правильной треугольной пирамиды?
- Приведите соответствующее доказательство.

3. СТРУКТУРНЫЙ ПОИСК. ФОРМИРОВАНИЕ ЗАПРОСА С ПОМОЩЬЮ АППЛЕТА

3.1. Структурный поиск

Во многих базах данных, содержащих сведения о химических структурах, имеется возможность вести информационный поиск, используя в качестве запроса двумерную структурную формулу.

В таких базах данных можно встретить три варианта действий поисковой программы; обычно пользователь сам выбирает нужный ему вариант из имеющегося набора.

1. Поиск по структуре — Structure (Exact) Search.

Поисковая программа ищет структуры, полностью совпадающие со структурой запроса.

2. Поиск по подструктуре (субструктуре) — Подструктурный (субструктурный) поиск — Substructure Search.

Поисковая программа ищет структуры, в остовах которых есть участок, *целиком* совпадающий с *остовом* запроса.

3. Поиск по подобию — Similarity Search.

Поисковая программа ищет структуры, в которых имеются такие же структурные фрагменты, какие есть в запросе.

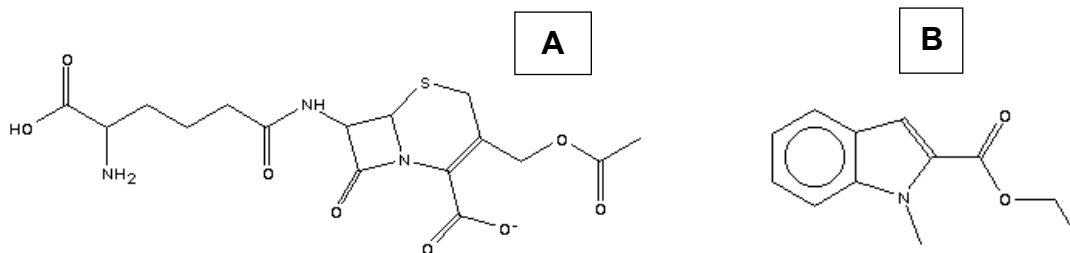
На поисковом бланке обычно можно задать минимально допустимую степень соответствия (в %).

Для расчета степени соответствия создатели баз данных часто используют коэффициент Танимото.

3.2. Контрольное задание 1. Расчет коэффициента Танимото

Предположим, вы — разработчик базы данных, в которой возможен структурный поиск по подобию.

База данных содержит, в частности, эти вещества:



Ваша задача:

- создать словарь структурных фрагментов, характерных для этой базы данных;
- сформировать структурные коды (отпечатки пальцев) для двух веществ А и В;
- рассчитать коэффициент Танимото для этой пары веществ.

Примем, что в нашей базе данных отпечатки пальцев будут иметь размер 10 битов.

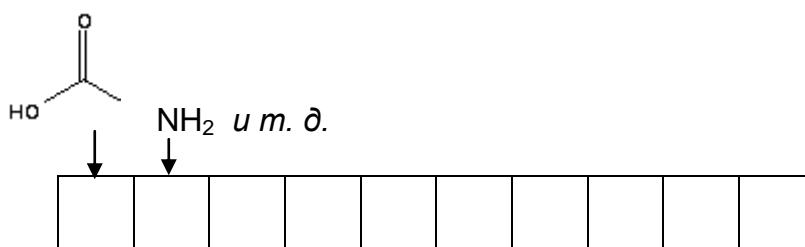
Какие именно структурные фрагменты следует включать в словарь?

Ответ: Любые на ваше усмотрение, но пригодные для формирования отпечатков пальцев структур A и B.

Решение приведите на листе бумаги по следующей схеме:

Словарь фрагментов.

Пример начала словаря:



Отпечатки пальцев:

A

--	--	--	--	--	--	--	--	--	--

B

--	--	--	--	--	--	--	--	--	--

Коэффициент Танимото:

$S_{AB} = \dots$ (расчетная формула и численное значение).

3.3. Апплет

Апплет — это маленькая исполняемая программа, которая загружается вместе с веб-страницей.

Рассмотрим апплеты, предназначенные для создания двумерных химических структур. С помощью таких апплетов конструируют запросы в тех базах данных, где возможен поиск по структурной формуле.

Приемы работы с апплетом весьма схожи с приемами работы с "большим" молекулярным редактором, таким как *Symyx Draw* — только намного проще.

Основная задача апплета — правильно отобразить молекулярный граф, на основании которого должен проводиться поиск. Красота формируемого изображения не имеет никакого значения в поисковом процессе, поэтому у таких апплетов нередко отсутствуют дизайнерские функции (выравнивание длин связей, валентных углов, поворот в плоскости листа и т. п.).

Инструментарий апплета, как правило, ограничивается управляющими командами (стереть, выделить) и кнопками химических символов (главным образом, "органического" набора), химических связей (главным образом, не отражающих симметрию), простейших структурных фрагментов (например, циклов).

Стандартизация в работе различных апплетов отсутствует, но приемы работы просты настолько, что при минимуме навыков они оказываются интуитивно понятными.

3.4. Апплет NIST Molecular Editor

[NIST Chemistry WebBook \(http://webbook.nist.gov/chemistry/\)](http://webbook.nist.gov/chemistry/) — это один из авторитетнейших онлайн-источников термодинамической, спектроскопической и структурной информации.

Запрос для информационного поиска в *NIST Chemistry WebBook* может быть сформирован несколькими способами. Рассмотрим, как для этих целей применяется апплет.

С Главной страницы *NIST Chemistry WebBook* к нужному бланку следует перейти по цепочке ссылок: *Structure* (в группе ссылок *Search Options / General Searches*) → *Use applet to draw a structure*.

Самостоятельная работа.

Ознакомьтесь с приемами формирования запроса для структурного поиска. Выполните тренировочные упражнения и контрольные задания.

Краткая инструкция к NIST Molecular Editor

Из имеющихся в WWW инструментов подобного назначения апплет *NIST* характеризуется, по-видимому, самым простым строением.

The screenshot shows the NIST Molecular Editor applet interface. At the top, there is a toolbar with buttons: Mode, Select, Draw, Erase, Template, Draw Element, a dropdown menu showing 'H', Text, and Done... Above the toolbar, five callout boxes point to specific features: 'Кнопки переключения режимов' (Mode buttons), 'таблица шаблонов' (Template table), 'список химических элементов' (Chemical elements list), 'вывод MOL-файла в текстовом формате' (MOL file output), and 'Кнопка "Готово"' (Done button). The main drawing area contains a chemical structure of salicylic acid (benzene ring with two adjacent hydroxyl groups). A callout box points to the structure with the text: 'Ничего страшного, если рисунок будет выглядеть не очень симпатичным' (Nothing to worry about if the drawing doesn't look very nice). At the bottom left, a callout box points to the molecular formula 'C7H6O3' with the text: 'Апплет автоматически определяет брутто-формулу' (The applet automatically determines the gross formula). The bottom status bar displays 'C7H6O3', 'Draw Mode', and 'NIST Molecule Editor'.

Апплет работает в нескольких режимах (*Mode*), которые включаются кнопками:

- **Select** — выделить;
- **Draw** — рисовать;
- **Erase** — стереть;
- **Template** — шаблоны.

Во всех режимах атом или связь можно выделить, установив курсор над ним (ней).

Draw — режим рисования.

Химическая структура наращивается тем атомом, который выбран в списке *Draw Element*.

- **Отобразить атом** — щелкнуть по полю.
- **Добавить** второй атом, связанный с первым — установить курсор на первый атом, нажать левую клавишу мышки и, не отпуская, тянуть курсор в нужном направлении на нужное расстояние, после чего клавишу отпустить.
- **Создать связь** между атомами — установить курсор на первый атом, нажать левую клавишу мышки и, не отпуская, перенести курсор на второй атом, после чего клавишу отпустить.
- **Изменить порядок** связи — установить курсор на связь и щелкать клавишей мышки нужное число раз.
- **Изменить свойства** атома — двойным щелчком по атому открыть вспомогательное окошко, в котором сделать необходимые замены.

Некоторые операции редактирования можно выполнять с помощью клавиатуры.

Список "горячих клавиш" приведен в таблице, расположенной на той же странице, где находится поисковый бланк.

- *Пример:* заменить в структуре атом углерода на атом азота.
Выполнение операции с помощью клавиатуры: установить курсор на атом углерода и нажать клавишу *n*.

Select — режим выделения.

- Выделенный атом можно передвигать в нужном направлении.
Остальные действия над атомами и связями такие же, как в режиме *Draw*.

Erase — режим стирания.

- Для удаления атома или связи необходимо щелкнуть по нему (ней).

Template — режим шаблонов (циклов).

- **Выбрать цикл** из таблицы — щелкнуть по нему.
- **Отобразить** этот цикл — щелкнуть в нужном месте редактируемого поля.
- **Добавить** нужный цикл к структуре — установить курсор на атом или связь и щелкнуть клавишей мышки.
(Проверьте, как добавляется цикл при щелчке по *атому*, а как — при щелчке по *связи*).

Об атомах водорода.

По умолчанию, на экране формируется осто́в молекулы, а присутствие атомов водорода подразумевается в соответствии с правилами валентности в *неявной* форме.

Например, молекула $\text{H}_3\text{C}-\text{CH}_3$ по умолчанию отображается так: —.

Для отображения водорода в *явной* форме необходимо установить курсор на атом, к которому присоединены атомы водорода, и на клавиатуре нажать соответствующую цифровую клавишу.

Тренировочное упражнение.

- Изобразите этан в такой форме: $\text{H}_3\text{C}-\text{CH}_3$
- Изобразите этил в такой форме: $\text{H}_3\text{C}-\text{CH}_2$

Явная и неявная формы отображения водорода по-разному интерпретируются в ходе субструктурного поиска.

- Неявно присутствующий водород никак не учитывается в ходе субструктурного поиска.

Например, по запросу $\text{C}-\text{C}$ извлекаются все структуры, содержащие два атома углерода, соединенные одинарной связью.

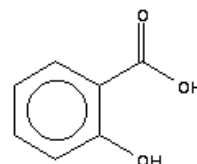
- Явно присутствующий водород учитывается в субструктурном поиске.

Например, по запросу $\text{H}_3\text{C}-\text{CH}_2$ извлекаются все структуры, содержащие группировку C_2H_5 .

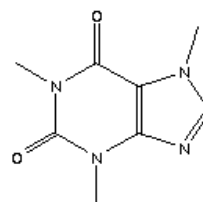
Выбор "поиск по структуре" или "поиск по субструктуре" проводится после нажатия кнопки *Done* ("Готово").

Контрольные задания 2.

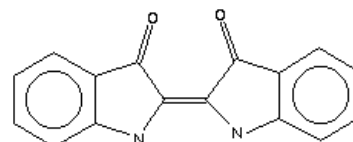
2а. Определите температуру плавления этого вещества:



2б. Определите энергию ионизации (eV) этой молекулы в газовой фазе:



2в. Проведите субструктурный поиск по такому запросу:



Чему равна энтальпия сублимации бромпроизводного, имеющегося в списке результатов поиска?

Результаты запишите на листе бумаги и сдайте преподавателю.

3.5. Апплет JME

Апплет *JME* используется на большом числе сайтов.

Функциональные возможности JME рассмотрим в [каталоге реактивов компании Sigma-Aldrich](#).

Зайдите на сайт:

<http://www.sigmaaldrich.com/catalog/AdvancedSearchPage.do>

и по ссылке *Structure Search* перейдите к поисковому бланку с апплетом.

Самостоятельная работа.

Ознакомьтесь с приемами формирования запроса для структурного поиска. Выполните тренировочные упражнения и контрольные задания.

Краткая инструкция к JME

Структура рабочего окна апплета

Шаблоны функциональных групп

1 2 3 4 5 6 7

Select substituent

Кнопки верхнего ряда:

- 1 – вывести в отдельном окошке код *SMILES*,
- 2 – стереть последнюю нарисованную молекулу
- 3 – начать рисовать еще одну структуру,
- 4 – удалить выделенный фрагмент,
- 5 – переместить структуру при нажатой левой клавише мышки, вращать структуру при нажатой правой клавише,
- 6 – превратить выделенный атом в ион,
- 7 – отменить последнее действие.

Химические связи

Углеродная цепь

Циклы

Нарисовать указанный атом. Заменить выделенный атом на данный.

Нарисовать "нестандартный" атом. Заменить выделенный атом на иной, вводя его символ на клавиатуре (открывается вспомогательное окошко).

JME Molecular Editor

JME Editor courtesy of Peter Ertl, Novartis

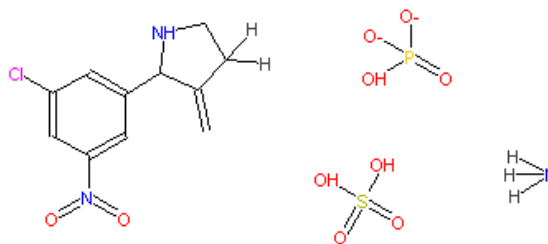
В иных версиях апплета на этих кнопках могут быть нанесены не пиктограммы, а буквенные обозначения.

Самостоятельная работа

Методом проб и ошибок изучите функциональные возможности всех кнопок.

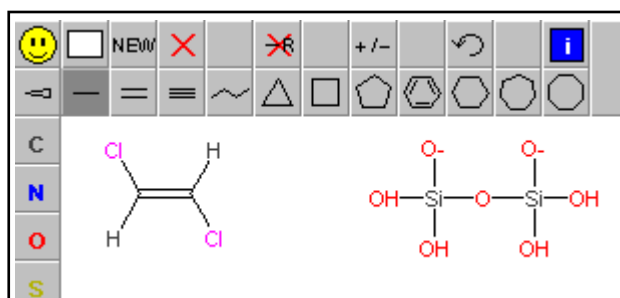
Потренируйтесь в наборе структурных формул, например, таких, которые изображены справа.

Проведите любой структурный поиск в каталоге.



Контрольное задание 3а.

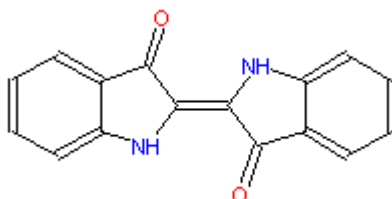
Создайте на экране точно такое же изображение:



Контрольное задание 3б.

Ваш коллега мельком упоминал, что видел в каталоге *Sigma-Aldrich* **редокс-индикатор**, пригодный для решения некоей исследовательской задачи.

Точного названия индикатора вы не помните, но знаете, что речь шла о производном такого вещества:



Итак, поздний вечер. Вы срочно составляете смету проекта, в котором потребуется именно этот индикатор. Время поджимает, а телефон коллеги недоступен.

Найдите в каталоге стоимость индикатора.

3.6. Апплет MarvinSketch

Функциональные возможности MarvinSketch рассмотрим на примере работы поискового бланка на сайте chemicalize.org.

chemicalize.org - поисковая система нового типа. Она предназначена для обнаружения химически значимой информации, причем запрос здесь можно составить в виде структурной формулы.

Самостоятельная работа.

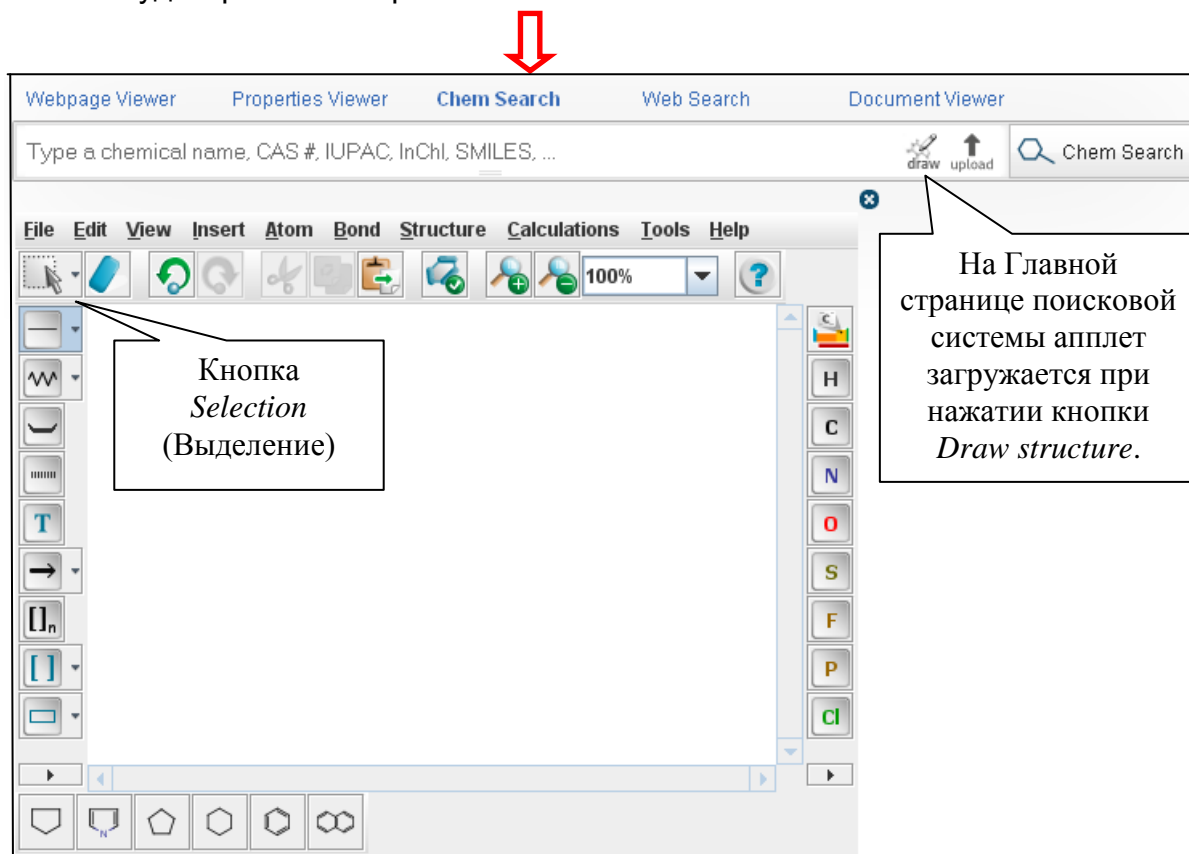
На сайте *chemicalize.org* перейдите в режим работы **Chem Search** (ссылка находится над поисковым бланком).

Изучите апплет *MarvinSketch*. Для вызова апплета служит кнопка **draw** на поисковом бланке (справа от графы, предназначенной для текстового запроса). Выполните проверочное задание, результат предъявите преподавателю.

Краткая инструкция к MarvinSketch

Поисковая система **chemicalize.org** использует простейшую версию *MarvinSketch*; в иных базах данных подобный апплет способен выполнять больше действий.

Мы будем работать в режиме *Chem Search*.



Ознакомление с апплетом

В отличие от апплетов, рассмотренных ранее, в данном случае в интерфейсе мы видим:

- пункты меню, многие команды которого дублируются кнопками;
- пункты меню, которые только формально присутствуют, но в апплете не работают (в одноименной прикладной программе — работают);
- инструмент для написания текстовых комментариев (кнопка T) (в поиске не используется);
- инструмент для отображения схемы реакции (кнопка со стрелкой в левом столбце) (в этой базе данных не используется).

Замечания перед работой:

- При низкой скорости сети апплет с запаздыванием реагирует на действия пользователя.
- Если в ходе конструирования структуры под атомом появляется красная черточка — значит, атом оказался в валентном состоянии, большем допустимого.
- По умолчанию, атомы Н присутствуют в структуре в неявной форме и компенсируют свободные валентности.

Самостоятельная работа.

Ваша задача — исследовать апплет самостоятельно и найти способы выполнения стандартных операций.

Исследуйте, каким образом можно:

- отобразить связь С–С (при нажатой кнопке С; при нажатой кнопке "связь");
- изменить порядок связи;
- заменить атом С на атом Н, N или О;
- заменить атом С на атом любого иного химического элемента;
- выделить атом или связь;
- выделить группу атомов (обратим внимание на кнопку со стрелкой в верхнем ряду и три алгоритма ее работы);
- выделить всю структуру;
- снять выделение;
- удалить атом;
- удалить группу атомов;
- удалить всю структуру.

К чему приводит двойной щелчок по структуре при включенной кнопке *Selection*?

Что происходит, если курсор поместить на середину выделенной структуры?

Как использовать это для перемещения структуры; для вращения структуры?

Как отобразить атом водорода в явной форме?

(Подсказка. Проанализируйте в меню группу команд *View*).

Как привести в опрятный вид неаккуратно нарисованную структуру?

(Подсказка. Проанализируйте в меню группу команд *Structure*).

Как двумерную структуру превратить в трехмерную?

Не пройдите мимо указанных ниже полезных функций.

Функциональные клавиши F6 и F7 назначают два режима работы:

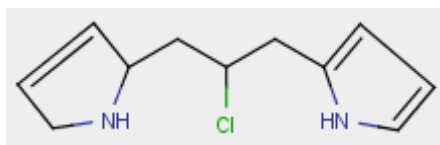
- F6 — изменение масштаба (при нажатой левой клавише мышки следует передвигать курсор по вертикали),
- F7 — вращение в трехмерном пространстве.

Буквенные клавиши можно использовать для отображения атомов. Например, N — атом азота, Cl — атом хлора и т. п. (допускается печатать и в нижнем регистре: n, cl и т. п.).

При щелчке правой клавишей мышки по атому или связи открывается контекстное меню. Поэкспериментируйте с контекстным меню.

Проверочное задание.

Изобразите на экране точно такую же структуру в точно таком же виде:



3.7. Знакомство с элементами семантического веба

Далее мы ознакомимся с идеями авторов *chemicalize.org*, пытающихся найти новые подходы в области работы с химической информацией.

Самостоятельная работа. Шаг 1.

Сформируйте запрос - изобразите структурную формулу **дифосфат-иона**.

Проведите поиск в базе данных поисковой системы — нажмите кнопку **Chem Search**.

Для самопроверки: список результатов поиска содержит около трех десятков записей (по состоянию на февраль 2013 г.).

Самостоятельная работа. Шаг 2.

Проанализируйте, куда ведут ссылки из списка результатов поиска.

Выведите на экран список документов, в которых упоминается **дифосфат кальция**.

Самостоятельная работа. Шаг 3.

В списке документов имеются патенты, обнаруженные в базе данных "freepatentsonline".

Выведите на экран текст любого из этих патентов (в формате *HTML*).

Самостоятельная работа. Шаг 4.

Проанализируйте вид текста патента, выведенного на экран.

Найдите такие особенности текста, которые явно отсутствовали бы в патенте, извлекаемом из обычной патентной базы данных (*USPTO, espacenet* и т. п.).

Перед вами **семантически обогащенный документ**.

Возможно, подобным образом будут конструироваться все научные документы в будущем — когда на смену нынешнему вебу придет *Web 3.0*.

Извлеченный патент **предъявите преподавателю**.

Перечислите те элементы и приемы, которые *chemicalize.org* использовала для обогащения исходного текста химической информацией.

4. ЛИНЕЙНЫЕ НОТАЦИИ. CAS RN, SMILES, InChI

4.1. Регистрационные номера химических веществ.

Одно и то же вещество можно назвать или письменно отобразить многочисленными способами (*какими?*).

Проблема однозначности имени вещества решалась и решается в химии разными путями. Один из них — присвоение веществу регистрационного номера.

Метод регистрации имеет и плюсы, и минусы.

С одной стороны, регистрационный номер действительно может служить уникальным идентификатором; с другой стороны, в нем совершенно не отражается состав вещества.

Главное же — понятие "вещество" настолько многогранно, что единственный регистрационный номер далеко не всегда пригоден при систематизации сведений о реальном химическом объекте.

В мире отсутствует единая система регистрации химических веществ; практически каждая большая база данных осуществляет регистрацию в пределах своего информационного массива.

Идентификационные коды полезны в информационном поиске — но как правило только внутри конкретного массива. У каждого правила должно быть исключение; в данной области исключением являются регистрационные номера *Chemical Abstracts Service*.

4.2. CAS Registry Number и справочник Common Chemistry

CAS Registry Number (CASRN, CAS RN, CAS Number, CAS#) — это номер, под которым химическое вещество (или смесь веществ) зарегистрировано в *Chemical Abstracts Service* (CAS).

Формат записи CASRN: группа из трех чисел, объединенных дефисами; в этом наборе последнее число обязательно состоит из одной цифры, а предпоследнее — из двух.

Примеры CASRN: 74-84-0 (этан); 548491-80-1 (ZnFe₂O₄).

В тексте документа код CASRN нередко заключают в квадратные скобки, например: [7647-14-5].

Сотрудники *Chemical Abstracts Service*, просматривающие свежую научную литературу, регистрируют каждый химический объект, обнаруженный в публикациях. Номера при этом присваиваются веществам простым и сложным, органическим, неорганическим и элементоорганическим, координационным, низкомолекулярным и высокомолекулярным (в том числе, белкам), ионам, минералам, сплавам, характерным смесям (в том числе, смесям неопределенного состава).

Присвоение номеров идет в хронологическом порядке, по мере обнаружения объектов, поэтому в сам номер никакой химический смысл не закладывают.

Ценность CASRN заключается в том, что регистрационный номер позволяет лаконично и однозначно указать на некое вещество, которое может иметь

множество названий и состав которого может быть отображен несколькими способами.

Пример: Na_2HPO_4 .

В литературе это вещество называют по-разному: гидроортофосфат натрия, натрий-гидрофосфат, кислый фосфат натрия двузамещенный, динатрийгидротетраоксофосфат(V), гідрафасфат натрию, disodium orthophosphate, DSP, hidrogenoortofosfato de disodio и т. д.

В зависимостей от целей, его химическую формулу иногда записывают, например, так: $\text{HNa}_2\text{O}_4\text{P}$; $\text{Na}_2\text{O} \cdot 0,5\text{H}_2\text{O} \cdot 0,5\text{P}_2\text{O}_5$; $2\text{Na}^+ \text{HPO}_4^{2-}$.

Количество вариантов графических изображений структуры вообще неисчислимо.

CASRN данного вещества однозначен: 7558-79-4.

Chemical Abstracts Service занимается систематическим учетом веществ с 1957 года; ежедневно количество присвоенных номеров возрастает на 15 тыс. На сайте [CAS](http://www.cas.org/) (<http://www.cas.org/>) есть счетчик зарегистрированных химических объектов, работающий в режиме реального времени:



Долгое время коды CASRN выполняли функции неофициальных стандартных идентификаторов химических веществ. Их можно встретить в научных статьях, технических спецификациях, каталогах реактивов, например:

Sodium chloride, crystalline powder, PDV, 99+%
CINa FW 58.44 [7647-14-5] mp 802° bp 1413° RTECS
EINECS 231-598-3 TSCA Fieser 6,285 Merck 12,8742

Поскольку коды CASRN интенсивно используются в химической литературе, их применяют в качестве поисковых терминов в информационном поиске, и не только в специализированных базах данных.

Пример. Необходимо найти "сертификаты безопасности материала" (*Material Safety Data Sheet, MSDS*) для Na_2HPO_4 .

Универсальная поисковая система *Google* успешно справляется с этой задачей, причем запрос **"7558-79-4" msds** оказывается более эффективным, чем текстовые запросы, такие как **"disodium orthophosphate" msds** .

Повсеместное распространение CASRN сдерживается двумя причинами:
(1) в этом коде нет информации о составе и структуре вещества;
(2) регистрационная система является коммерческой.

Исчерпывающий список всех CASRN содержится в *CAS REGISTRY* — платной базе данных. *Chemical Abstracts Service* разрешает некоммерческое использование CASRN — но при условии, что в информационный источник будет включено менее 10 тыс. регистрационных номеров. В результате такой политики в печатной литературе (особенно в каталогах реактивов) и в онлайн-справочных базах данных накопилось значительное количество кодов CASRN, однако это лишь мизерная доля той информации, которой владеет CAS.

Для широкой научной аудитории доступ к первоисточнику (*CAS REGISTRY*) затруднен, поэтому она склонна пользоваться вторичными документами. В результате в онлайн-литературе из одного документа в другой копируются

коды с опечатками и отмененные коды, а главное, нередко встречается некорректное соотнесение *CASRN* и вещества.

Правильное соотнесение *CASRN* и вещества — задача не такая простая, какой она кажется при беглом взгляде на проблему. CAS присваивает регистрационные номера не только химическим веществам в строгом понимании этого термина, но и в нестрогом, и даже смесям.

Пример 1. Цис-1,2-дихлорэтен, транс-1,2-дихлорэтен и 1,2-дихлорэтен (т. е. без указания взаимного расположения атомов хлора) имеют три разных регистрационных номера. Если в молекуле атомы ^1H заменять атомами ^2H , получаются новые объекты, которые тоже будут иметь свои уникальные номера.

Пример 2. Одно-, двух-, трехзамещенные ортофосфаты натрия, как и следовало ожидать, имеют три уникальных регистрационных номера. Оказывается, свой номер есть и у "замещенного фосфата натрия" (т.е. смеси веществ, состав которой можно описать брутто-формулой $\text{Na}_x\text{H}_y\text{PO}_4$, где $x+y=3$). Кроме того, регистрационные номера присвоены каждому обнаруженному кристаллогидрату ($\text{Na}_2\text{HPO}_4 \cdot 2\text{H}_2\text{O}$, $\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$ и т. д.).

С точки зрения регистрации реальных химических объектов, о которых имеются сведения в научной литературе, такой подход вполне логичен.

Трудности возникают у авторов документов и у составителей справочных баз данных, вознамеривших дополнить свои материалы кодами *CASRN*. Для правильного выбора кода необходимо профессионально ориентироваться в регистрационной системе и иметь доступ к первоисточнику *CAS REGISTRY*. Как показывает анализ качества бесплатных веб-ресурсов, эти условия выполняются далеко не всегда.

Трудности возникают у пользователей, намеревающихся найти в Интернете информацию о веществе по коду *CASRN*. Пользователь должен знать "правильный" код, да к тому же еще и уметь предвидеть, какие "частично правильные" коды могут встретиться в онлайн-литературе (присвоение кристаллогидрату регистрационного номера безводной соли — самый безобидный пример).

Так имеет ли смысл использовать *CASRN* в информационном поиске? Конечно же, да.

Во-первых, без оговорок эффективен поиск по *CASRN* в тех источниках, для которых эти коды изначально были предназначены — в базах данных *Chemical Abstracts Service* (увы, платных).

Во-вторых, в бесплатном Интернете имеется значительный информационный пласт, где *CAS Registry Numbers* применяются издавна и в основном выверены. Это каталоги реактивов ведущих химических компаний, сертификаты безопасности материалов (*MSDS*), базы данных *National Institute of Standards and Technology (NIST)*, указатели к печатным справочникам по свойствам веществ. (Именно здесь пользователю целесообразно решать проблему соотнесения "вещество \leftrightarrow *CASRN*", по крайней мере, для более-менее распространенных веществ).

Неплохими стартовыми точками для поиска кодов *CASRN* являются базы данных, в которых собрана информация из каталогов разных производителей

химреактивов, например: [BuyersGuideChem](http://www.buyersguidechem.de/) (<http://www.buyersguidechem.de/>), [ChemExper](http://www.chemexper.com/) (<http://www.chemexper.com/>).

В тех случаях, когда требуется расшифровать серию кодов CASRN, удобен указатель справочника [Landolt-Börnstein Substance / Property Index](http://lb.chemie.uni-hamburg.de/static/RN/) (<http://lb.chemie.uni-hamburg.de/static/RN/>).

Chemical Abstracts Service до недавних пор был монопольным держателем первичной информации о CASRN. В конце 2008 года он сделал первый шаг в предоставлении свободного доступа к этому типу своих ресурсов и разместил официальную выборку из CAS REGISTRY на специальном сайте [Common Chemistry](http://www.commonchemistry.org/) (<http://www.commonchemistry.org/>).

Common Chemistry содержит сведения (CASRN, названия, молекулярные и структурные формулы) примерно о 8 тыс. химических веществ — таких, которые не менее тысячи раз упоминаются в базах данных CAS.

Предполагается, что сайт должен служить инструментом поиска названий и регистрационных номеров важнейших химических товаров.

Поисковый бланк, находящийся на Главной странице *Common Chemistry*, состоит из двух граф:

Chemical Name — здесь запросом может быть полное название вещества или фрагмент названия;

CAS Registry Number — здесь запросом может быть только полный код CASRN, но не его фрагмент.

Пример. По запросу **sodium phosphate** в список результатов поиска попадают одно-, двух-, трехзамещенные ортофосфаты, их смесь неопределенного состава, один ортофосфат-кристаллогидрат, ди-, мета-, полифосфаты, гипофосфит и соль с фосфорорганическим анионом.

По запросу **disodium phosphate** извлекается только запись, касающаяся вещества Na_2HPO_4 .

Принимая во внимание небольшой объем базы данных, результат можно считать удовлетворительным.

You Searched: sodium phosphate
11 results found
1 2

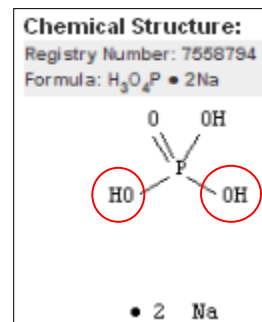
Registry Number: 7558-79-4
Molecular Formula: $\text{H}_3\text{O}_4\text{P} \cdot 2\text{Na}$
Chemical Name: Phosphoric acid, disodium salt
[Show Synonyms](#)

Registry Number: 7558-80-7
Molecular Formula: $\text{H}_3\text{O}_4\text{P} \cdot \text{Na}$
Chemical Name: Phosphoric acid, monosodium salt
[Show Synonyms](#)

При работе с базой данных следует учитывать некоторые принятые здесь условности, на первый (да и на второй) взгляд кажущиеся химику весьма странными.

(1) Химическую формулу соли и в символьной ("*Molecular Formula*"), и в графической ("*Chemical Structure*") записях отображают как некий комплекс кислоты и атомов металла, например, $\text{H}_3\text{O}_4\text{P} \cdot 2\text{Na}$, а не $\text{HNa}_2\text{O}_4\text{P}$ (химические символы приведены в алфавитном порядке).

Печально, что эту особенность записей, принятую в регистрационной системе CAS, бездумно тиражируют

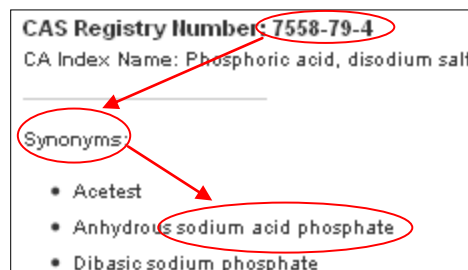


авторы веб-сайтов, действующие по принципу *Copy&Paste*. В итоге в некоторых онлайн-справочных базах данных появляются необычные объекты типа $\text{H}_3\text{Na}_2\text{O}_4\text{P}$, для которых там же приводятся "молекулярные веса" и даже физико-химические характеристики.

(2) Все названия, под которыми данное вещество было обнаружено в литературе, CAS перечисляет в рубрике, неудачно поименованной "Синонимы".

Некритично настроенный пользователь, подавленный авторитетом *Chemical Abstracts Service*, может и на самом деле поверить, что "двунариевая соль фосфорной кислоты" и "кислый фосфат натрия" — синонимы.

Кстати, и этот термин, именно с таким искаженным смыслом, бродит по Интернету.



Можно предположить, что "антихимические" условности были временно внесены программистами CAS на заре компьютеризации, да так и остались в базах данных. (Нет ничего более постоянного, чем временное?)

Абстрагируясь же от терминологических вопросов, следует признать, что сайт *Common Chemistry* как первичный источник информации о CASRN является ценным и перспективным онлайн-ресурсом.

4.3. CAS RN в базах данных.

Если речь идет о веществах, достаточно распространенных, для соотнесения CAS RN и названия вещества рекомендуется обращаться к сайтам, где содержится официальная информация:

[Common Chemistry](http://www.commonchemistry.org/) (<http://www.commonchemistry.org/>)

[Wikipedia](http://en.wikipedia.org/) (<http://en.wikipedia.org/>)

Википедия и "официальная информация"? Это шутка или опечатка? Дайте мотивированный ответ.

Упражнение.

Найдите достоверный CAS RN воды.

В ЕС существует своя система регистрации химических товаров *EINECS* (*European INventory of Existing Commercial chemical Substances*); соответствующие коды записываются в формате EC # (в ходу и прежние коды *EINECS* #, *ELINCS* #).

Соответствующая база данных находится на сайте:

[ESIS : European chemical Substances Information System](http://esis.jrc.ec.europa.eu/) (<http://esis.jrc.ec.europa.eu/>).

Самостоятельная работа.

Ознакомьтесь с сайтом *ESIS*. Изучите, как по названию вещества можно найти код EC #, код CAS RN.

Какая информация о веществе имеется в базе данных?

Найдите CAS RN воды в базе данных *ESIS*. Правильный ли код попал в эту базу данных?

Как правило, хорошим справочником по CAS RN может быть каталог реактивов солидной компании.

Рассмотрим с этой точки зрения каталог компании [Sigma-Aldrich](http://www.sigmaaldrich.com/catalog/AdvancedSearchPage.do) (<http://www.sigmaaldrich.com/catalog/AdvancedSearchPage.do>).

Контрольное задание 1.

В каталоге *Sigma-Aldrich* проведите поиск по термину "вода" (*water*). Задание формулируйте в разделе "*Data Specific Search*" усложненного поискового бланка.

Список результатов поиска содержит, кроме самой воды, огромное число разных водных растворов. Чтобы сделать список более удобным для анализа, отсортируйте вещества по увеличению молярной массы.

Найдите в списке химические объекты, к которым совершенно обоснованно можно применить термин "вода". Разные ли это химические вещества?

Выпишите на лист бумаги названия и CAS RN этих объектов и предъявите преподавателю.

Продолжаем знакомство с сайтом, у которого, похоже, великое будущее:

[ChemSpider](http://www.chemspider.com/) (<http://www.chemspider.com/>).

В 2009 г. этот прежде частный сайт перешел в собственность *Royal Society of Chemistry*. RSC создает здесь крупнейший онлайн-центр структурной и спектральной информации.

База данных *ChemSpider* постоянно растет, расширяется инструментарий сайта, что не может не радовать пользователей.

С другой стороны, подход *ChemSpider* к формированию базы данных должен того же пользователя настораживать.

ChemSpider — характерный представитель *Web 2.0*. Его база данных — это сумма многочисленных коммерческих и некоммерческих баз данных плюс репозиторий, наполняемый онлайн-сообществом; отсюда вытекают проблемы редактирования, рецензирования и расплывчатая достоверность.

Упражнение.

В базе данных *ChemSpider* проведите поиск по термину "вода".

В результатах поиска откройте раздел *Names and Identifiers* — здесь приведено то многообразие имен и кодов, под которыми воду можно обнаружить в литературе.

Как и полагается в *Web 2.0*, в списке отражены результаты работы онлайн-рецензентов: видно, что ими проверено, что — нет (кстати, а откуда это видно?).

Найдите в списке коды CAS RN, приписанные воде.

Контрольное задание 2.

Определите, каким химическим объектам принадлежат коды CAS RN, имеющиеся в той записи, которая соответствует воде.

Какие онлайн-справочники вы использовали для проверки?

Если бы вы были экспертом *ChemSpider*, какие бы изменения вы внесли в эту часть раздела *Names and Identifiers*?

Результат изложите на бумаге в письменном отчете.

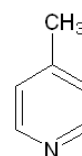
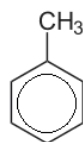
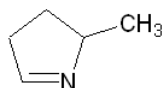
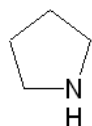
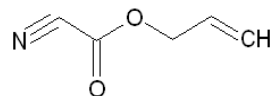
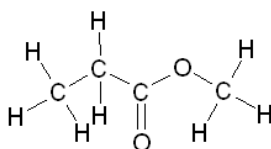
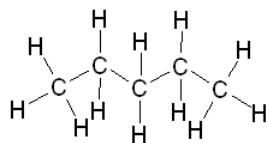
4.4. SMILES

Линейная нотация *SMILES* отображает химическую структуру в форме одномерной строки символов.

(Основные правила формулирования *SMILES* рассматривались на лекции).

Тренировочное упражнение. Формирование кода *SMILES*.

а) Не прибегая к помощи компьютера, сформируйте коды *SMILES* для каждого из указанных ниже веществ и запишите эти коды на листе бумаги.



б) Проведите самопроверку сформированных вами кодов *SMILES*, как указано ниже.

В каталоге [Sigma-Aldrich](#) на странице поискового бланка *Structure Search* имеется инструмент для конвертации кода *SMILES* в двумерную структурную формулу:

SMILES:	<input type="text"/>	Convert
---------	----------------------	----------------

Доступность сайта хорошая, конвертация протекает на лету, поэтому мы воспользуемся этим инструментом для самопроверки.

Вызовите бланк на экран своего компьютера (если нужной графы на странице *Structure Search* вы не увидите, обратите внимание на ссылку *More Options*).

Для самопроверки вносите в бланк сформированный вами код *SMILES*, нажимайте кнопку *Convert* и сравнивайте структуру, появляющуюся в окошке апплета, с заданием. Должно совпадать! (Конечно же, с учетом неявно присутствующих атомов водорода).

Во многих базах данных поисковое задание можно формулировать в виде *SMILES*.

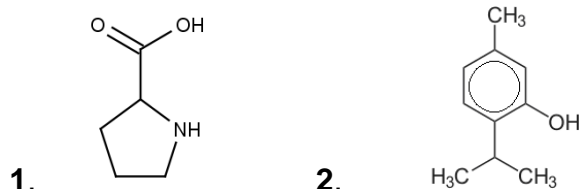
При поиске сведений о веществе код *SMILES* целесообразно использовать в запросе в следующих ситуациях:

- при поиске в текстовых базах данных (где структурную формулу в принципе применить невозможно);
- в структурных базах данных, если код прост в написании (например, легче набрать на клавиатуре строку NCC(O)=O, чем рисовать двумерную структуру глицина);

- в структурных базах данных, если готовый код можно скопировать из другого источника, например, из молекулярного редактора (из-за несовместимости форматов молекулярных редакторов далеко не всегда возможен перенос двумерной формулы; для буквенно-цифровой строки *SMILES* проблема несовместимости отсутствует в принципе).

Контрольное задание 3. Информационный поиск в базе данных ChemSpider.

3а) Составьте коды *SMILES* указанных ниже веществ.



3б) Проведите поиск по *SMILES*-запросам в базе данных *ChemSpider*. (Можно использовать простейший бланк, расположенный на Главной странице).

Simple search
Structure search
Advanced search

Systematic name, synonym, trade name, registry number, SMILES or InChI

Справка. Перевод некоторых терминов сайта.

Properties = свойства;

Predicted Properties = предсказанные (расчетные значения) свойства;

Experimental Properties = экспериментально определенные свойства;

Melting Point = температура плавления;

Density = *Specific Gravity* = плотность.

3в) В базе данных *ChemSpider* найдите ответы на следующие вопросы:

- Как называется вещество № 1?
Чему равна экспериментально найденная температура плавления этого вещества?
 - Как называется вещество № 2?
Насколько сильно различаются значения его плотности, определенные экспериментально и рассчитанные теоретически?
- Веб-страницы с ответами предъявите преподавателю.

4.5. InChI и InChIKey

Линейная нотация *InChI* разрабатывается под эгидой ИЮПАК.

(Структура *InChI* рассматривалась на лекции).

Информационный поиск по запросу, состоящему из кода *InChI* или кода *InChIKey*, пока что не обладает какими-то особыми преимуществами по сравнению с другими методами. Причина кроется в том, что эти коды — явление достаточно новое и, хотя горячо рекламируемое своими сторонниками, не получившее повсеместного распространения.

Тем не менее, есть все основания полагать, что такая ситуация временна.

Научному сообществу необходима стандартная линейная нотация химической структуры, и *InChI* для этой роли вполне годится.

В *WWW* растет количество мест, где можно узнать код *InChI* для химического вещества. Например, графа с кодом *InChI* есть в таблицах свойств веществ на страницах Википедии.

Насколько правильны коды *InChI*, встречающиеся в Интернете — вопрос открытый, но не такой сложный, как с регистрационными номерами *CAS RN*.

В Интернете есть генераторы кодов *InChI*, причем рекомендованные ИЮПАК:

1. ChemSketch

ChemSketch — молекулярный редактор фирмы *ACD/Labs*. Он является частью свободно распространяемого пакета программ *ACDLabs Freeware*; генерирование *InChI* — одна из функций редактора.

ChemSketch — предмет нашего рассмотрения на одном из последующих занятий.

2. ChemSpider (<http://www.chemspider.com/>)

С информационным центром *ChemSpider* мы уже работали.

На сайте *ChemSpider*, с одной стороны, *InChI* можно использовать в качестве запроса; с другой стороны, в результатах поиска, проведенного любым иным методом, всегда можно видеть коды *InChI* и *InChIKey* для извлеченных веществ, например:

Systematic Name:	2-isopropyl-5-methyl-cyclohexan-1-ol
SMILES:	<chem>CC1CC(CCC1C(C)C)C</chem> Сору
InChI:	<u>InChI=1/C10H20O/c1-7(2)9-5-4-8(3)6-10(9)11/h7-11H,4-6H2,1-3H3</u>
InChIKey:	<u>NOOLISFMXDJSKH-UHFFFAOYAE</u>
Std. InChI:	<u>InChI=1S/C10H20O/c1-7(2)9-5-4-8(3)6-10(9)11/h7-11H,4-6H2,1-3H3</u>
Std. InChIKey:	<u>NOOLISFMXDJSKH-UHFFFAOYSA-N</u>

3. PubChem Sketcher (<http://pubchem.ncbi.nlm.nih.gov/edit2/index.html>)

Этот апплет, расположенный на сайте *PubChem*, позволяет генерировать коды *InChI* и *SMILES* по структурной формуле и, наоборот, по кодам *InChI* и *SMILES* генерировать структурную формулу.

Внешне *PubChem Sketcher* выглядит несколько иначе по сравнению с теми апплетами, с которыми мы работали ранее.

Функциональные его особенности в области простейших задач незначительны:

- Перед вводом очередной структуры обязательно следует убрать прежнюю информацию — нажать кнопку *New*.
- Отсутствуют и кнопка "выделить", и отдельная операция "выделить", но в неявной форме "выделение" функционирует. Хотя при наведении курсора на атом или связь в структуре ничего не подсвечивается, смело можно в этом месте впечатывать новый химический символ или изменять порядок связи.
- На панели активная кнопка окрашена в оранжевый цвет, неактивные — в голубой.

Перейти к новой структуре

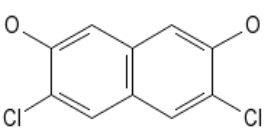
Список типов нотаций

Текст нотации

Broad
InChI
InChI=1/C10H6Cl2O2/c11-7-1-5-2-8(12)10(14)4-6(5)3-9(7)13/h1-4,13-14H

New	Undo	Clm	Del	Qry	↔	↺	↻	⊕	⊖	⊗	⊘	⊙	⊚	⊛	⊜	⊝	⊞	⊟	⊠	⊡	⊢	⊣	⊤	⊥	⊦	⊧	⊨	⊩	⊪	⊫	⊬	⊭	⊮	⊯	⊰	⊱	⊲	⊳	⊴	⊵	⊶	⊷	⊸	⊹	⊺	⊻	⊼	⊽	⊾	⊿	⋀	⋁	⋂	⋃	⋄	⋅	⋆	⋇	⋈	⋉	⋊	⋋	⋌	⋍	⋎	⋏	⋐	⋑	⋒	⋓	⋔	⋕	⋖	⋗	⋘	⋙	⋚	⋛	⋜	⋝	⋞	⋟	⋠	⋡	⋢	⋣	⋤	⋥	⋦	⋧	⋨	⋩	⋪	⋫	⋬	⋭	⋮	⋷	⋸	⋹	⋺	⋻	⋼	⋽	⋾	⋿	⋐	⋑	⋒	⋓	⋔	⋕	⋖	⋗	⋘	⋙	⋚	⋛	⋜	⋝	⋞	⋟	⋠	⋡	⋢	⋣	⋤	⋥	⋦	⋧	⋨	⋩	⋪	⋫	⋬	⋭	⋮	⋷	⋸	⋹	⋺	⋻	⋼	⋽	⋾	⋿
-----	------	-----	-----	-----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Шаблоны



Export

SDF

Done

Hydrogen

Keep AsIs

Help

Import

Browse...

Упражнение 1. Генерирование структуры по известному коду *InChI*.
 Найдите в Википедии код *InChI* для подсластителя аспартама (*aspartame*), скопируйте код в текстовую графу апплета, нажмите клавишу "Ввод".
 Апплет сгенерирует структурную формулу.
 Сравните ее с той структурной формулой, которая приведена в Википедии.

Упражнение 2. Генерирование кода *InChI* по структурной формуле .
 Рисуйте в окне апплета структурную формулу простейшей аминокислоты и наблюдайте, что происходит в графе, предназначенной для кода *InChI*.

Контрольное задание 4а. Информационный поиск по коду *InChI*.
 Скопируйте код аминокислоты, сгенерированный вами в *Упражнении 2*, и проведите по нему поиск в *Google*.
 Сколько обнаружено документов, какова их информативность?
 Результаты поиска предъявите преподавателю.

Вернитесь к *PubChem Sketcher* и проанализируйте, какие типы линейных нотаций этот апплет может генерировать по известной двумерной структуре (в качестве подопытного объекта используйте все ту же простейшую аминокислоту).

Примените полученные знания для выполнения следующего задания.

Контрольное задание 4б. Информационный поиск по коду *InChIKey*.

Сколько химических объектов в базе данных *ChemSpider* имеет точно такой же структурный остов, какой имеется у простейшей аминокислоты?

Извлеките список объектов, используя *InChIKey*, и предъявите преподавателю.

5. КАТАЛОГИ ХИМИЧЕСКИХ РЕАКТИВОВ

Самостоятельная работа 1. Печатный каталог

Ознакомьтесь со структурой печатного каталога реактивов.
Найдите в каталоге реактив, указанный преподавателем.

Самостоятельная работа 2. Онлайновые каталоги

Ознакомьтесь с сайтами, перечисленными ниже.
Найдите поисковые бланки.
Изучите возможности простейших и усложненных поисковых бланков.
Самостоятельно проведите поиск сведений о реактиве (например, *фенол* = *phenol*, *бензол* = *benzene* и т.д.).
Изучите структуру и объем материала, получаемого в результате поиска.

Каталоги крупных компаний

Acros Organics (<http://www.acros.com/>).

Текстовые поисковые бланки находятся в левой колонке.
В меню *Search center* выберите тип термина текстового запроса (CAS, название вещества и пр.).
Для формирования структурного запроса в каталоге используется апплет JME.

Alfa Aesar (<http://www.alfa.com/en/ge100w.pgm>).

Простейший поисковый бланк находится вверху страницы; ссылки на усложненный — вверху страницы, а на бланк структурного поиска — внизу.

Sigma-Aldrich (<http://www.sigmaaldrich.com/catalog/AdvancedSearchPage.do>)

С сайтом вы уже работали и знаете, что здесь можно вести и текстовый, и структурный поиск.

Merck Millipore (<http://www.merckmillipore.com/chemicals>).

Текстовый запрос можно сформулировать на Главной странице.

Ссылка на бланк структурного поиска находится на странице *Advanced Search*.

Метасайты

eMolecules (<http://www.emolecules.com/>).

Поиск проводится по каталогам ведущих компаний, продающих реактивы в Европе.

Из результатов поиска можно перейти сразу на нужную страницу каталога соответствующего продавца.

ChemExper (<http://www.chemexper.com/>).

В результатах поиска перечислены продавцы, адреса их сайтов, но прямые ссылки на соответствующие страницы каталогов не приводятся.

BuyersGuideChem (<http://www.buyersguidechem.de/>).

В списке результатов поиска перечислены продавцы, адреса их сайтов, но прямые ссылки на соответствующие страницы каталогов отсутствуют.

Chemical Book (http://www.chemicalbook.com/ProductIndex_EN.aspx).

В результатах поиска приводится большой список поставщиков химических товаров, причем на экран можно вывести такие кластеры как *China Suppliers* или *Global Suppliers* (здесь упоминаются компании и наших соседей, например, Украины, России).

6. ОРГАНИЧЕСКИЙ СИНТЕЗ: ОНЛАЙНОВЫЕ РЕСУРСЫ

6.1. Введение

Базы данных, с которыми мы знакомились ранее и с которыми встретимся позже, содержат разноплановую информацию о химических веществах: структуру, экспериментально определенные физические свойства, рассчитанные параметры, химические свойства, спектры, биологическую активность и т. д.

Достаточно часто перед химиком стоит конкретная задача — найти методику синтеза вещества. В общем случае, для решения такой проблемы потребовалось бы проанализировать содержание многих баз данных, в том числе, статьи в журналах, где публикуются соответствующие сведения. Понятно, что для решения специфической задачи полезно было бы использовать соответствующую базу данных, специализирующуюся именно в этой области.

Рассмотрим инструменты, которые могут быть полезны при поиске методик синтеза органических веществ.

6.2. ChemSynthesis

Базу данных [ChemSynthesis](http://www.chemsynthesis.com/index.html) (<http://www.chemsynthesis.com/index.html>) можно рассматривать как путеводитель по почти полутора десяткам авторитетных журналов (*Journal of the American Chemical Society*, *Chemistry Letters*, *Tetrahedron* и др.).

В своей основе *ChemSynthesis* — это библиографическая база данных, обогащенная химическими структурами, возможностью структурного поиска и некоторыми полезными вспомогательными инструментами.

Основная область ее применения: — обнаружение библиографического описания той статьи, в которой описывается методика синтеза заданного вещества.

В режиме *Browse* можно просмотреть все записи, извлеченные из каждого из журналов; записи, сгруппированные по классам соединений и по функциональным группам.

Эти списки не структурированы, поэтому в обычной поисковой работе ценность их невелика.

В режиме *Text Search* возможен поиск:

- по названию вещества (и тривиальному, и систематическому),
- по CAS RN и некоторым иным регистрационным номерам,
- по InChIKey (и по полному, и по первой части кода),
- по брутто-формуле (и по полной, и по ее фрагменту).

Самостоятельная работа.

Ознакомьтесь с сайтом; проведите пробный поиск информации.

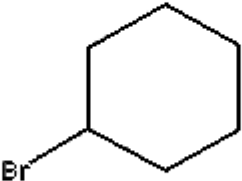
От простейшего бланка, находящегося на Главной странице, ссылка *Text Search* ведет к еще двум бланкам текстового поиска: простому и усложненному.

Усложненный бланк:

Chemical Name <input type="text"/>	Cas Number <input type="text"/>	Synthesis Reference <input type="text"/>
Molecular Formula <input type="text"/>	Molecular Weight <input type="text"/>	InChIKey <input type="text"/>
<input type="button" value="Search!"/>		

Графа для записи названия журнала

Типичная запись в базе данных имеет вид:

Cyclohexyl bromide	
Molecular Formula:	C ₆ H ₁₁ Br
Molecular Weight:	163.057 g/mol
Cas Number:	108-85-0
InChIKey:	AQNQQHJNRPDOQV-UHFFFAOYAD
Smiles:	BrC1CCCCC1
	
Synthesis Reference	
Canadian Journal of Chemistry, 50, p. 3109, 1972 Journal of the American Chemical Society, 112, p. 8212, 1990 The Journal of Organic Chemistry, 45, p. 1638, 1980	

Справочная информация

Ссылки на статьи, где содержатся методики синтеза

Структурный поиск.

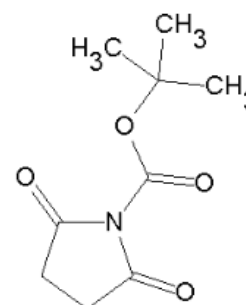
Для формирования структурного запроса используется апплет *JME*.

Апплет содержит полезное дополнение: меню с функциональными группами.

Для того, чтобы изобразить функциональную группу, следует выбрать ее в меню, а затем щелкнуть в структуре по тому атому, к которому эта группа должна быть присоединена.

Контрольное задание 1.

- Проведите поиск в базе данных;
- найдите ссылку на методику синтеза;
- найдите соответствующую статью;
- загрузите статью;
- покажите преподавателю то место в статье, где описывается методика синтеза следующего соединения:



6.3. Organic Chemistry Portal

Organic Chemistry Portal (<http://www.organic-chemistry.org/>) содержит несколько взаимосвязанных частей, в том числе:

- **Organic Chemistry Highlights.**

Этот раздел построен в форме онлайн-журнала кратких тематических обзоров новых методов получения органических веществ.

Полностью сформирован материал по темам "Многокомпонентные реакции", "Микроволновая химия"; в стадии формирования находятся разделы "Тотальный синтез", "Ультразвуковая химия".

Следует учитывать, что в обзорах реферируются новейшие разработки и не рассматриваются методики, опубликованные в прошлом.

- **Organic Reactions.**

Этот раздел состоит из подразделов *Name Reactions* (Именные реакции), *Organic Synthesis Search* (Поиск в области органического синтеза), *Protecting Groups* (Защитные группы).

Рассмотрим структуру подраздела **Organic Synthesis Search** (<http://www.organic-chemistry.org/synthesis/>).

Organic Synthesis Search — инструмент обнаружения методик проведения таких реакций, в ходе которых образуется заданная химическая связь.

Organic Synthesis Search имеет структуру многоуровневого каталога.

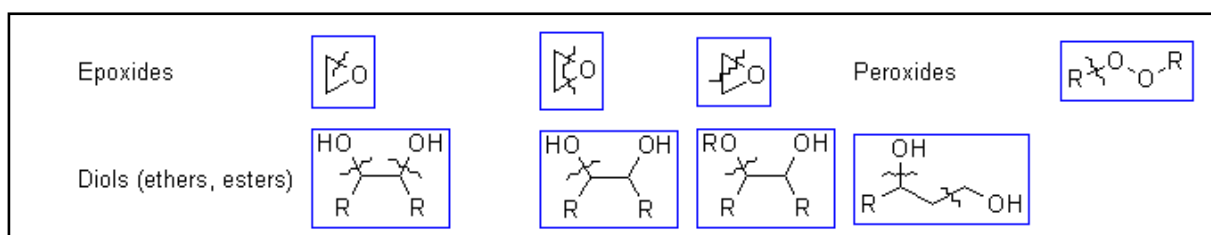
Извлечение информации проводится методом *Browse*.

На первой стадии в соответствующей ячейке таблицы следует выбрать связь, которая должна образоваться в результате реакции.

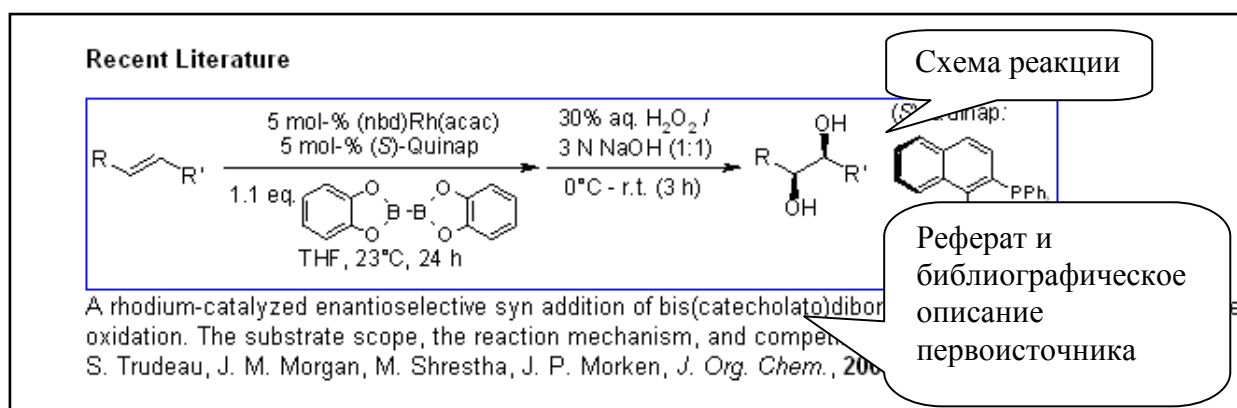
	B	C	N	O	Si	P	S
H / :		—	—	—		—	—
B		—					
C	—	—=—≡	—=—≡	—=—	—	—	—=—
N		—=—≡	—=—≡	—=—		—	—=—
O		—=—	—				—=—
Si		—		—			
P		—	—				—=—
S		—=—	—=—	—=—		—=—	—
F		—					
Cl		—	—				—
Br		—					
I		—		—			
Sn		—					
Div			Heterocycles	Heterocycles			Heterocycles

На второй стадии запрос уточняется — здесь необходимо выбрать класс образующегося соединения.

Так, например, выглядит фрагмент страницы "Образование связи C–O":



На следующих стадиях работы пользователю предьявляется информация о некоторых общих методах синтеза, а также конкретные методики в схематическом изложении с указанием первоисточника (без гиперсвязи), например:



Самостоятельная работа.

Ознакомьтесь с сайтом *Organic Chemistry Portal*.

Контрольное задание 2. Дисульфид из тиола.

- Найдите информацию о способе получения $\text{CH}_3\text{-S-S-CH}_3$ из CH_3SH .
- Перейдите к первоисточнику, содержащему методику этого синтеза.
- Сравните объемы информации, касающейся этого синтеза, в базе данных *Organic Chemistry Portal* и в первоисточнике (материал, доступный бесплатно).

Форма отчетности о выполнении задания:

- выведите на экран сведения из первоисточника;
- на листе бумаги укажите те детали методики, которые доступны бесплатно в *Organic Chemistry Portal*, но на сайте первоисточника доступны только за плату.

6.4. ChemSpider SyntheticPages

6.4.1. Структура базы данных

В 2010 году [ChemSpider](http://cssp.chemspider.com/) (<http://cssp.chemspider.com/>) приступил к интегрированию в свою базу данных методик органического синтеза, содержащихся на сайте *Synthetic Pages*.

Пока что сервис небольшой по объему, но есть основания полагать, что со временем он станет одним из интереснейших онлайн-ресурсов в области

органической химии.

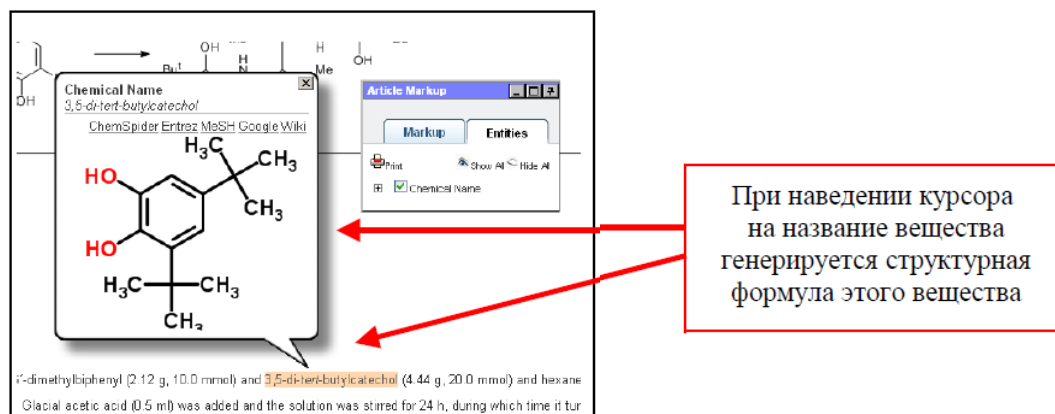
База данных *ChemSpider SyntheticPages* содержит методики синтеза органических веществ.

ChemSpider SyntheticPages работает по принципам *Web 2.0* — материал формируется из сообщений пользователей и рецензируется пользователями.

Полный текст записи содержит подробное описание методики синтеза, а также сведения о реагентах, физико-химические и спектроскопические параметры продукта реакции, комментарии автора и читателей, библиографические ссылки на первоисточники, файлы химических структур (в нескольких форматах).

Текст методики семантически обогащен; за основу взят подход, который применяется в *RSC Project Prospect*.

В тексте методики цветом выделены слова, несущие химическую смысловую нагрузку (главным образом, названия веществ). При наведении курсора на название вещества во всплывающем окошке появляется соответствующая структурная формула и ссылки, ведущие к информации об этом веществе в иных базах данных. Информация из внешних источников извлекается по коду *InChI*.



Управление семантически обогащенной разметкой осуществляется из всплывающего окошка *Article Markup*.

Принципы ведения информационного поиска

Пользователь формулирует либо текстовый запрос (*Text Search*), либо структурный (*Structure Search*).

В режиме *Text Search* можно вести поиск по ключевым словам, названию вещества, фамилии автора.

Для того, чтобы сформулировать запрос в режиме *Structure Search*, следует выполнить ряд действий.

- При щелчке по полю, на котором изображена графическая формула с карандашом, открывается еще одно окно браузера.
- Здесь следует перейти на закладку *Draw Structure* (если она не открылась по умолчанию).
- По умолчанию, загружается один из апплетов. Пользователь может переключиться на любой иной апплет из набора, указанного на бланке.

На данном занятии вам необходимо будет сформировать запрос с помощью апплета *ACD/Labs SDA*.

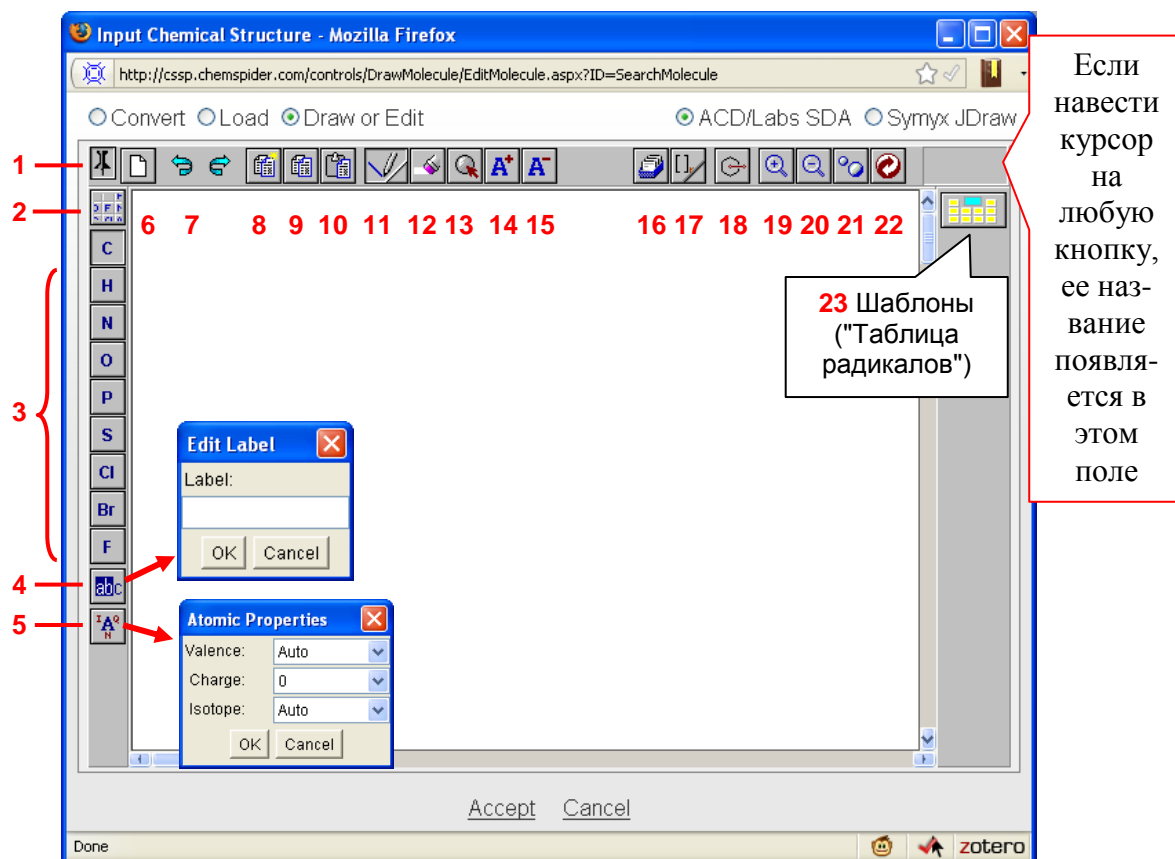
6.4.2. Апплет ACD/Labs SDA на сайте ChemSpider

Апплет *ACD/Labs SDA* создан на основе графического редактора из пакета *ACD/Labs Freeware*, с которым мы еще будем работать позже.

С помощью апплета можно нарисовать химическую структуру и сохранить ее в графическом или в *MOL*-формате.

Для формулирования запроса в *ChemSpider* пока что используются не все функциональные возможности апплета.

Управляющие кнопки апплета в основном интуитивно понятны.



- 1 — Закрепление апплета в окошке браузера.
- 2 — Кнопки всех химических элементов.
При выборе химического элемента на панели появляется новая кнопка, которая сохраняется до конца сеанса.
- 3 — Атомы химических элементов.
- 4 — Бланк для набора сокращенных названий структурных фрагментов, например, *Me* (в запросе не использовать).
- 5 — Бланк для указания валентности, заряда, массового числа атома (в запросе не использовать).
- 6 — Очистить поле апплета.
- 7 — Кнопки *Отменить* и *Восстановить*.
- 8 — Нарисованное скопировать в память в графическом формате.
Скопированное можно внести в виде рисунка, например, в документ *MS Word*.
- 9 — Нарисованное скопировать в память в *MOL*-формате.

Скопированное можно сохранить в виде *MOL*-файла, используя текстовый редактор.

10 — Преобразовать текст *MOL*-файла в графическое изображение.

11 — Нормальное Рисование связей
(а также кнопки отображения стереосвязей).

12 — Стереть.

13 — Выделить/Передвинуть.

Все нарисованное выделяется после двух щелчков по полю.

Для выделения фрагмента структуры обвести фрагмент курсором при нажатой левой клавише мышки.

Передвинуть можно выделенный фрагмент или всю структуру, перемещая курсор при нажатой левой клавише мышки.

Снять выделение — еще раз щелкнуть по полю.

14 — Щелчком по атому приписать ему положительный заряд.

Повторные щелчки увеличивают заряд.

15 — Щелчком по атому приписать ему отрицательный заряд.

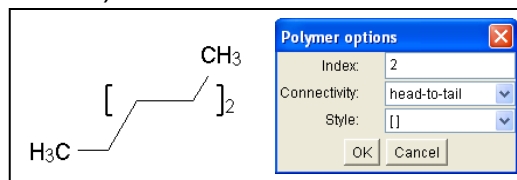
Повторные щелчки увеличивают отрицательный заряд.

16 — Шаблоны (в *ChemSpider* временно не работают).

17 — Бланк отображения числа

полимеризации.

Если выделен фрагмент структуры, число полимеризации указывается для этого фрагмента, например:

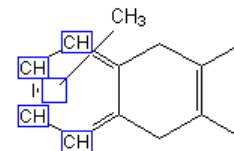


18 — Структура Маркуша (в запросе не использовать).

По умолчанию, заместителем является метил, который затем можно изменять стандартным путем.

Если в исходной структуре есть выделенный фрагмент, то заместитель будет относиться к этому фрагменту.

Для того, чтобы определить, к какому фрагменту относится заместитель в готовой формуле, следует поместить курсор на конец связи, направленной от заместителя.



19,20 — Увеличение/уменьшение размера формулы.

21 — Согласовать размер формулы с размером окна.

22 — Исправить изображение химической структуры.

В двумерной графической формуле стандартизируются длины связей и валентные углы (в предположении, что применим метод гибридизации).

23 — Шаблоны структурных фрагментов.

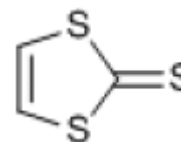
Для присоединения фрагмента к существующей структуре необходимо совместить букву R шаблона с тем атомом исходной структуры, с которым должна образоваться связь.

Сформулировав структурный запрос, следует нажать на кнопку *Асcept (Принять)*. Структурная формула вносится в поисковый бланк.

В бланке необходимо указать режим поиска (точное совпадение — *Exact Match* — или иной) и, наконец, нажать на кнопку *Search*.

Контрольное задание 3.

Найдите методику синтеза следующего вещества:
Результат предъявите преподавателю.



6.5. Organic Syntheses

[Organic Syntheses](http://www.orgsyn.org/default.asp) (<http://www.orgsyn.org/default.asp>) — это многотомный сборник выверенных методик синтеза органических веществ.

Тексты методик находятся на сайте; первоисточники представлены в форме библиографических описаний.

Информация о способах получения вещества в *Organic Syntheses* излагается более подробно, чем это принято в журналах. Воспроизводимость каждой методики перед публикацией проверяется не менее двух раз в лабораториях членов редколлегии.

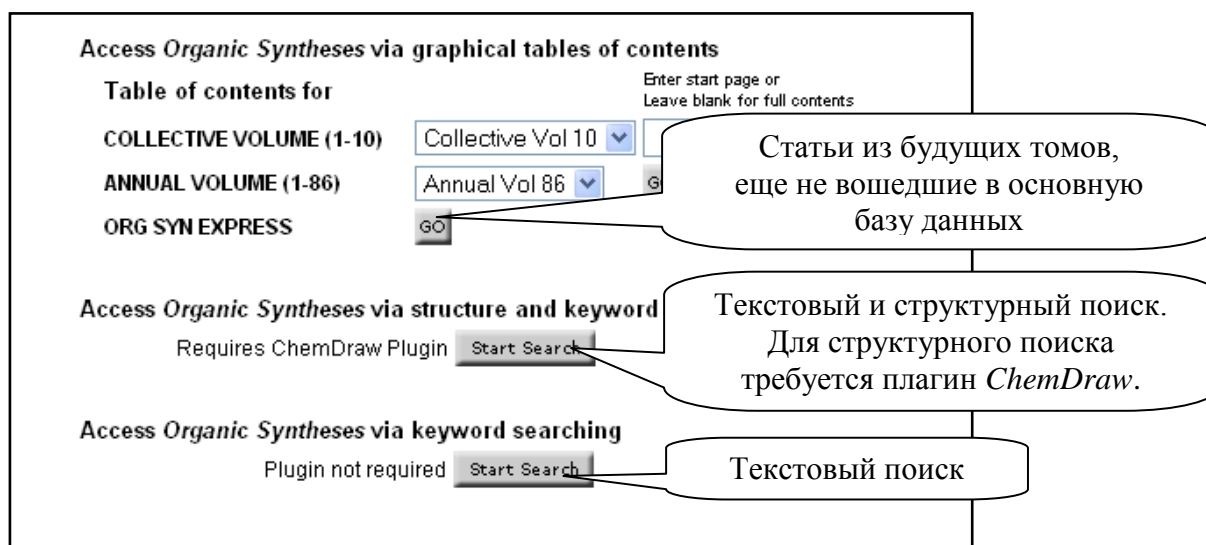
Organic Syntheses издается в форме ежегодников (с 1921 г.).

Статьи повторно публикуются в Сводных томах (*Collective Volume*). По этой причине у каждой давней статьи имеется по два библиографических описания. Ранее, по 59-й том включительно, Сводные тома выходили каждые десять лет, а впоследствии — каждые пять лет.

Некоторые методики могли быть существенно улучшены после их опубликования в ежегоднике — в таком случае в Сводный том включают обновленную информацию.

Перед началом работы пользователь должен прочитать предупреждение о необходимости соблюдать технику безопасности и о том, что составители не несут ответственность за его действия; он также должен подтвердить свою квалификацию как химика — щелкнуть по ссылке *Accept and Continue* (Принимаю. Продолжить).

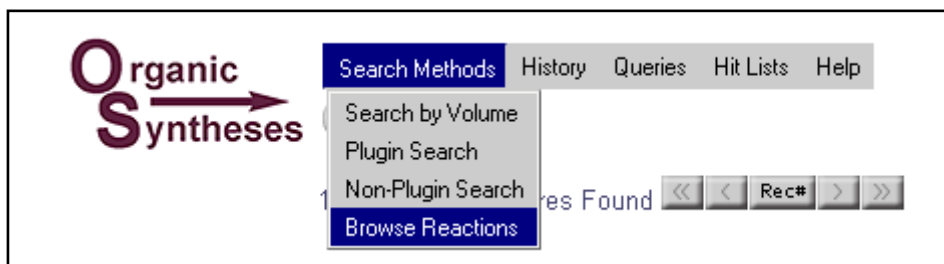
Режим работы выбирают на Главной странице:



The screenshot shows the main navigation area of the Organic Syntheses website. It includes three main sections:

- Access Organic Syntheses via graphical tables of contents**: This section allows users to select a volume. It has dropdown menus for 'COLLECTIVE VOLUME (1-10)' (set to 'Collective Vol 10') and 'ANNUAL VOLUME (1-86)' (set to 'Annual Vol 86'). There is a 'GO' button. An annotation bubble points to the 'GO' button, stating: 'Статьи из будущих томов, еще не вошедшие в основную базу данных'.
- Access Organic Syntheses via structure and keyword**: This section requires the ChemDraw Plugin and has a 'Start Search' button. An annotation bubble points to this button, stating: 'Текстовый и структурный поиск. Для структурного поиска требуется плагин ChemDraw.'
- Access Organic Syntheses via keyword searching**: This section does not require a plugin and has a 'Start Search' button. An annotation bubble points to this button, stating: 'Текстовый поиск'.

В результатах поиска, в перечне статей тома по команде *Browse Reactions* (в меню — в группе *Search Methods*) можно перейти к тематическому каталогу, где опубликованная информация (не вся!) рассортирована по типам реакций:



Поисковая программа может обнаруживать:

- реагенты,
- продукты реакции,
- схемы превращений.

Запрос может содержать не только текст, но и структурную формулу.

Поисковый бланк:

The screenshot displays the Organic Syntheses search interface. At the top is the logo and a navigation bar with 'Search Options', 'Search Methods', 'History', 'Queries', 'Hit Lists', and 'Help'. Below the logo are 'Search' and 'Clear Form' buttons. A prompt reads 'Search for articles containing the substructure or reaction'. A large rectangular box for drawing chemical structures is shown, with a red box and arrow pointing to it from a text label: 'Панель инструментов появляется при щелчке по редактируемому полю.' To the right of this box is a vertical 'Tools' toolbar containing various chemical drawing icons. Below the drawing area are several search filters, each with a text input field and a help icon: 'and has this phrase anywhere', 'and molecular formula is', 'and chemical name contains', 'and chemical's CAS RN is', 'and index terms contains', 'and title contains', 'and author's name starts with', 'and in collective volume', 'and starts on collective volume', 'and in annual volume (1-8)', and 'and starts on annual volume page'. At the bottom, the 'Order Results' section includes four radio button options: 'Collective Volume & Page Descending' (selected), 'Annual Volume & Page Descending', 'Collective Volume & Page Ascending', and 'Annual Volume & Page Ascending'.

Основные правила формулирования текстового запроса

Бланк для текстового запроса содержит графы для поиска по перечисленным ниже полям.

Has this phrase anywhere — полнотекстовый поиск по всем полям.

- Внимание! Не обращайтесь на слово *phrase* в названии поля; запрос здесь — не фраза, а набор терминов.
- Шаблоны в запросе использовать нельзя.
- Разрешены булевы операторы *AND*, *OR*, *AND NOT*. Правила использования булевых операторов стандартны.
- Для поиска близко расположенных слов рекомендуется применять оператор *NEAR*. Расстояние между словами не указывается. Чем ближе расположены слова, объединенные оператором *NEAR*, тем выше ранг документа, тем выше документ расположится в списке результатов поиска.
- Фраза обозначается кавычками. Стоп-слова, находящиеся внутри кавычек, учитываются в ходе поиска.

Molecular Formula — Молекулярная (брутто-) формула.

- Поиск проводится по всему документу.
 - Химические символы должны быть в стандартном написании: первая буква заглавная, вторая — строчная.
 - Порядок химических символов в формуле может быть любым.
 - Цифры записываются не в нижнем индексе, а в строке.
 - Формула может содержать пробелы.
- Примеры записей: Si_2O_7 или Si_2O_7 — допустимы оба варианта.

Chemical Name — название вещества.

- Поиск по названию вещества проходит по всему документу.
- Автоматический учет синонимов не проводится. Поэтому, например, по термину *mercaptan* (меркаптан) не извлекаются документы с термином *thiol* (тиол).

CAS RN

- Регистрационные номера *CAS Registry Numbers* должны быть в формате *NNNNN-NN-N*.

Index Terms — термины указателя (индекса)

- На сайте имеется вспомогательный предметный указатель (индекс), который не предназначен для перелистывания, но может применяться в поиске.
- Указатель содержит названия химических приборов, типы реакций, термины техники безопасности, названия реактивов, например: *voltmeter* (вольтметр), *transesterification* (переэтерификация), *exchange resin* (ионообменная смола), *stench* (зловоние).

Title — название методики

- Поиск проводится по полю "Название методики".

Author's Name — персональные данные автора

- Персональные данные автора следует записывать в обращенном порядке: на первом месте указывается фамилия, затем имя и второй инициал; после фамилии следует ставить запятую, затем пробел перед именем.
- Поисковая программа ищет авторов, персональные данные которых начинаются с букв, указанных в запросе, поэтому в запросе допускается и обрубание фамилии (без использования шаблона).
Например, по запросу *Schmi* будут обнаружены *Schmidt, Steven A.*, *Schmitz, Ernst* и *Schmitz, William R.*
По запросу *Schmitz, E* будут обнаружены *Schmitz, Ernst* и *Schmitz, E.*, но не *Schmitz, William R.*

Volume and Page Number — том и номер страницы

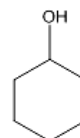
- *Organic Syntheses* публикуется в форме ежегодников. Каждые десять лет, по том 59 включительно, и впоследствии каждые пять лет статьи повторно публикуются в Сводных томах (*Collective Volume*). По этой причине у каждой статьи, опубликованной в прежние годы, имеется два библиографических описания.
Например, синтез *ethyl 4-hydroxycrotonate* изначально был опубликован в т. 64, с. 104, позже повторно — в сводном томе 7, с. 221. Методика может быть извлечена, если указать 7 в графе номера сводного тома и 221 в графе страниц сводного тома. Второй вариант — внести числа 64 и 104 в графы номеров томов и страниц ежегодника.

Структурный поиск

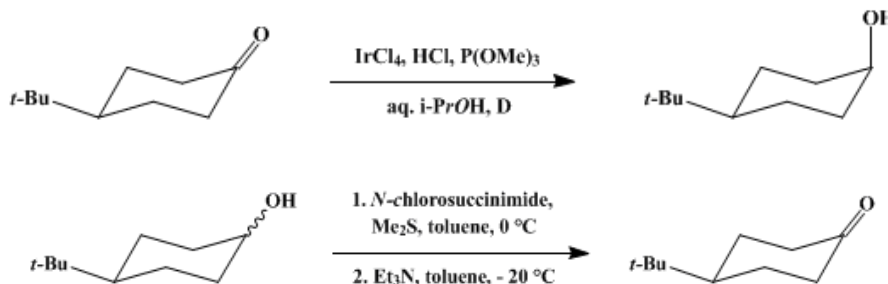
Для формулирования структурного запроса применяется плагин *ChemDraw*. В качестве запроса можно использовать одну или несколько структурных формул; кроме того в запросе может присутствовать стрелка как часть записи схемы реакции (как и для чего — об этом будет сказано ниже).

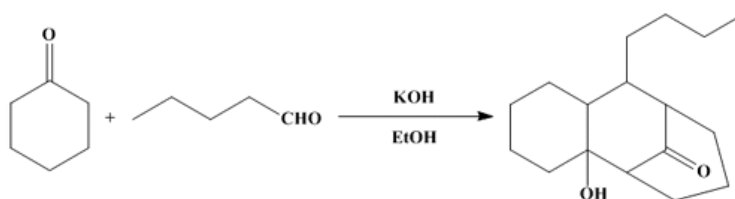
Если запросом является структурная формула **без стрелки**, поисковая программа ищет все схемы реакций, в которых есть **реагент или продукт** с такой структурой или с таким структурным фрагментом.

Например, если запросом является *циклогексанол*



будут извлекаться все реакции, в которых реагент или продукт содержит гидроксильную группу, присоединенную к насыщенному шестичленному углеродному кольцу, включая сопряженные циклы, в том числе такие:

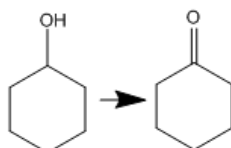




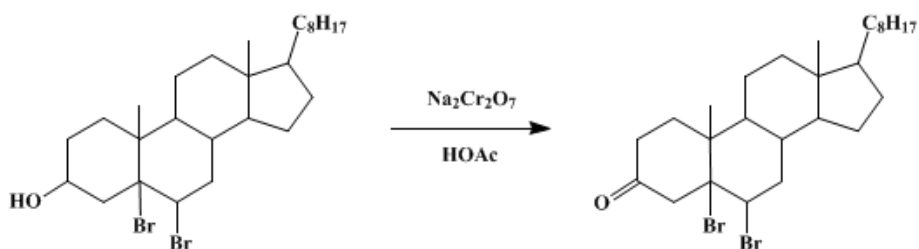
Чтобы извлечь **только** реакции с изображенным **реагентом** и любыми продуктами, следует нарисовать **стрелку после** формулы.

Чтобы извлечь **только** реакции с изображенным **продуктом** и любыми реагентами, следует нарисовать **стрелку перед** формулой.

Запрос может содержать одновременно **и реагент, и продукт** реакции. Например, по запросу (*обратите внимание на наличие стрелки*):



будет извлечена методика синтеза:



По умолчанию, в ходе реакции допускается замещение в любом месте исходной структуры.

Пользователь может **запретить замещение** на каком-нибудь атоме.

Место в структуре, доступное для замещения, здесь обозначают термином *Free Site*.

Если в свойствах атома остова указать *Free Sites: 0*, то это означает запрет на замещение на данном атоме.

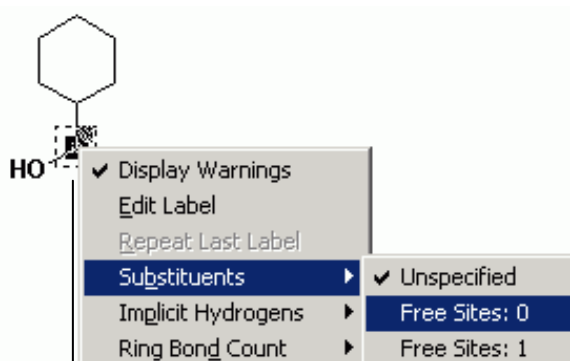
Порядок действий:

Выделить атом в структуре.

Открыть контекстное меню.

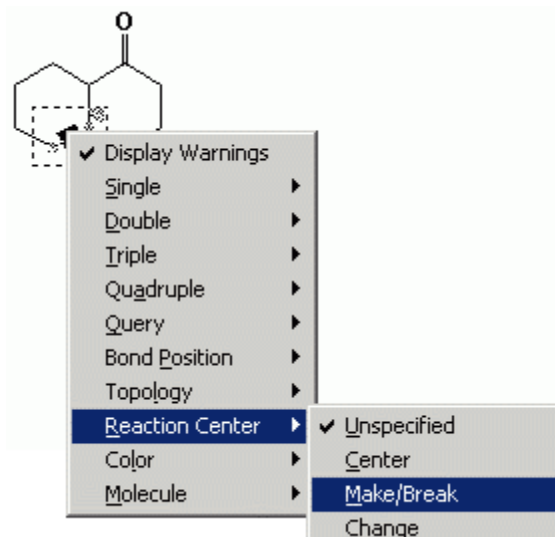
Перейти в строку *Substituents* (заместители).

Щелкнуть в подменю по строке *Free Sites: 0*.



В запросе можно указать место, где **должна образовываться или разрываться** химическая связь.

Порядок действий.
 Выделить химическую связь в структуре.
 Открыть контекстное меню.
 Перейти в строку *Reaction Center* (Реакционный центр)
 Щелкнуть в подменю по строке *Make/Break*.

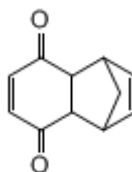


В запросе должна присутствовать стрелка (перед или после формулы), так как поисковая программа должна знать, о чем идет речь — о реагенте или о продукте реакции.

Примечание. Практика показывает, что нерелевантные документы достаточно часто встречаются среди результатов поиска.

Контрольное задание 4.

4а) Найдите методику синтеза следующего вещества:



4б) Представим себе следующую вполне реальную ситуацию.

Вам необходимо синтезировать 160-170 г того вещества, о котором идет речь в пункте 4а.

Время не поджидает, некоторое количество валюты имеется. Исходных реактивов у вас нет, но вы можете подождать, пока получите эти вещества по почте от продавца.

Задача:

Найдите оптимальный вариант покупки полного комплекта реагентов для синтеза.

(оптимальный вариант = минимум затрат + надежный продавец).

Сдайте преподавателю расчеты, оформленные по следующей схеме:

Реагент	Масса реагента, необходимая для синтеза	Масса вещества в одной упаковке	Стоимость одной упаковки	Число упаковок	Сумма расходов на этот реагент	Фирма-поставщик
Всего:						

6.6. Послесловие: Spotlight Reagents.

Журнал *SYNLETT* содержит любопытный раздел [Spotlight Reagents](http://www.thieme-chemistry.com/en/products/journals/synlett/access/online-edition/spotlight-reagents-list.html) (<http://www.thieme-chemistry.com/en/products/journals/synlett/access/online-edition/spotlight-reagents-list.html>).

В этот раздел принимают к публикации работы студентов-старшекурсников, магистрантов, аспирантов, молодых ученых. Каждая публикация — это двухстраничное краткое сообщение о каком-нибудь известном реактиве (органическом или неорганическом), имеющем важное значение в органическом синтезе.

Самостоятельная работа.

Ознакомьтесь со структурой одной из статей, например, S375, *Silver Carbonate*.

Spotlight Reagents функционирует с 1999 года, здесь опубликовано около 400 работ, но еще ни разу не было автора из Беларуси (статистику 2009 г. можно посмотреть в юбилейной статье S300).

Кто первый из вас, коллеги?

7. СПЕКТРАЛЬНЫЕ БАЗЫ ДАННЫХ: ЯМР

Спектроскопические исследования применяются в химии для идентификации веществ, для качественного и количественного анализа смесей и т.д.

Рассмотрим один из методов спектроскопического исследования — ЯМР-спектроскопию.

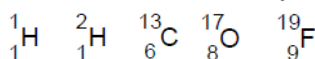
7.1. Краткое описание метода ЯМР-спектроскопии.

Ядерный магнитный резонанс (ЯМР) = Nuclear Magnetic Resonance (NMR)

Ядра атомов с нечетным зарядом или нечетным массовым числом имеют ненулевой спин и характеризуются ненулевым магнитным моментом.

Примеры:

Магнитный момент не равен нулю:



Магнитный момент равен нулю:



Во внешнем магнитном поле ядра с ненулевым магнитным моментом могут находиться в нескольких энергетических состояниях.

Переход с нижнего энергетического уровня на верхний сопровождается поглощением энергии ($\Delta E = h\nu$), которое можно обнаружить с помощью прибора — ЯМР-спектрометра.

Протонный магнитный резонанс (ПМР) — Proton Nuclear Magnetic Resonance, ^1H -NMR

Протонный магнитный резонанс — частный случай ядерного магнитного резонанса.

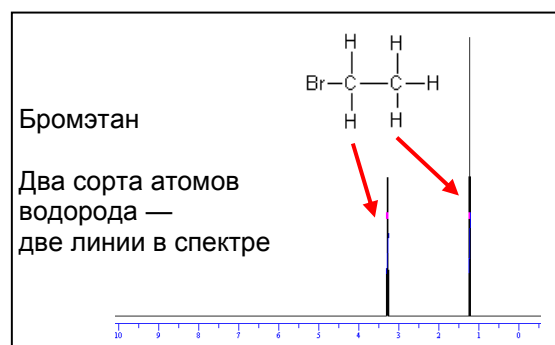
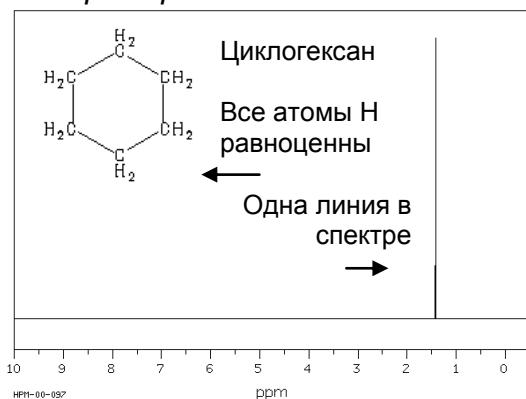
Ядро ^1H может находиться в двух состояниях: его собственный магнитный момент может ориентироваться по направлению внешнего поля и против поля.

При ЯМР-исследовании ампулу с растворенным исследуемым веществом помещают в сильное магнитное поле и облучают мощным узкочастотным импульсом, основная частота которого характерна для данного ЯМР-спектрометра (например, 100 МГц). Атомы H резонируют при некоторых частотах ($\Delta E_i = h\nu_i$). Прибор обнаруживает поглощение энергии при этих частотах и выводит результаты (в табличной форме, в виде спектра).

При прочих равных условиях частота поглощения зависит от состояния электронной оболочки атома ^1H .

- Ядра химически идентичных атомов водорода поглощают излучение одной и той же частоты.
- Ядра химически различающихся атомов водорода поглощают излучение разной частоты.

Примеры.



Для того, чтобы можно было сравнивать результаты разных экспериментов, в ЯМР-спектроскопии используется следующая стандартизация:

а) спектр снимают в присутствии вещества-эталоны.

В ПМР-спектроскопии в качестве эталона чаще всего используют $\text{Si}(\text{CH}_3)_4$ (тетраметилсилан, *TMS*), дающий интенсивный одиночный сигнал в стороне от большинства иных сигналов.

б) положение остальных сигналов прибор рассчитывает относительно сигнала *TMS* по формуле:

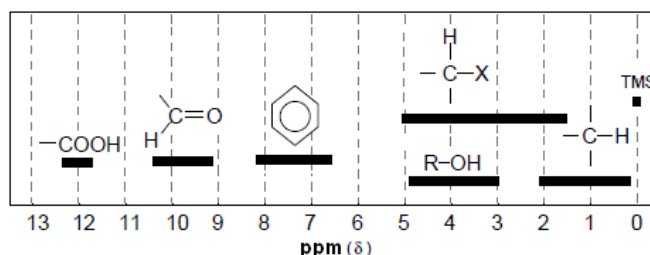
$$\delta = \frac{(\text{частота данного сигнала} - \text{частота сигнала } TMS), \text{ Гц}}{\text{частота спектрометра, Гц}} \times 10^6$$

δ называется **химическим сдвигом**; это безразмерная величина, выражаемая в миллионных долях (*ppm = parts per million*) от основной частоты прибора.

Запомним термин: **Химический сдвиг = Chemical Shift**

Химический сдвиг сигнала атомов водорода, входящих в состав *TMS*, по определению равен нулю: $\delta=0,00$.

Химический сдвиг каждого типа атомов водорода лежит во вполне определенном интервале, например: →



Химический сдвиг δ зависит от ближайшего окружения атомов.

Химический сдвиг δ тем больше, чем сильнее смещена электронная плотность с атома водорода.

Посмотрите на приведенные выше спектры циклогексана и бромэтана.

Химические сдвиги для атомов водорода циклогексана и CH_3 -групп бромэтана близки по величине ($\delta \approx 1$).

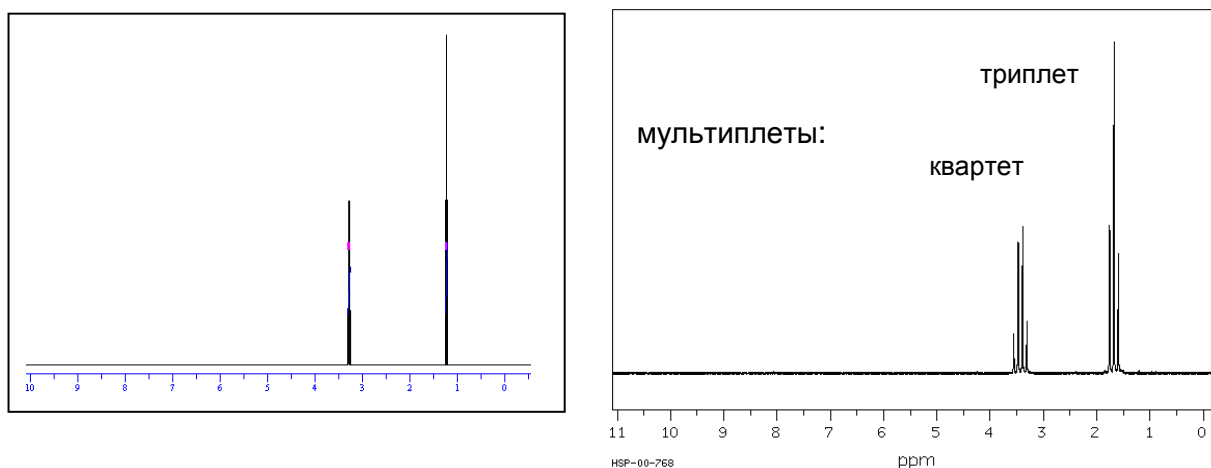
В бромэтано из-за смещения электронной плотности к атому брома химические сдвиги двух типов протонов разные и соотносятся следующим образом: $\delta(\text{CH}_2) > \delta(\text{CH}_3)$.

В спектрах ЯМР низкого разрешения, о которых шла речь до сих пор, каждой группе идентичных атомов ^1H соответствует один сигнал.

В спектрах ЯМР высокого разрешения обнаруживаются более тонкие эффекты.

При практически том же химическом сдвиге, при котором в спектре низкого разрешения наблюдается одиночный пик (синглет), в спектре высокого разрешения может находиться мультиплет — группа пиков, расположенных очень близко друг возле друга.

Пример — спектры низкого разрешения (слева) и высокого разрешения (справа) для бромэтана:



Мультиплет появляется оттого, что на энергетическое состояние данного протона влияют магнитные моменты иных близко расположенных протонов — происходит расщепление энергетических уровней.

Поскольку энергия таких взаимодействий очень мала, этот эффект можно обнаружить только в спектре высокого разрешения.

Химически одинаковые протоны не расщепляют энергетические уровни друг у друга.

Пример: Циклогексан.

Все атомы H в молекуле циклогексана химически идентичны. В спектре низкого разрешения наблюдается один пик. В спектре высокого разрешения тоже наблюдается только один пик — расщепление отсутствует.

Расщепление уровней и появление мультиплета вызывается протонами, химически отличающимися от данного и расположенными, главным образом, по соседству.

Число пиков в мультиплете = (число соседних химически иных протонов) + 1.

Пример.

Бромэтан $\text{Br}-\text{CH}_2-\text{CH}_3$ (спектр приведен на предыдущей странице)
Два типа протонов: А В

В спектре высокого разрешения имеется два мультиплета:

- сигнал протонов группы А под действием протонов группы В расщепляется на квартет ($3+1=4$);
- сигнал протонов группы В под действием протонов группы А расщепляется на триплет ($2+1=3$);

Дополнительные замечания.

Высоты пиков в мультиплете соотносятся следующим образом:

2 пика	(дублет)	1:1
3	(триплет)	1:2:1
4	(квартет)	1:3:3:1
5	(квинтет)	1:4:6:4:1 и т. д.

Площадь под мультиплетом тем больше, чем большему числу атомов водорода принадлежит этот сигнал.

Соотношение площадей под мультиплетами равно соотношению протонов в соответствующих химически разных группах атомов.

Например, в спектре бромэтана площади под квартетом и триплетом соотносятся как 2:3 (они формируются из фрагментов CH_2 и CH_3).

Из-за высокой подвижности протона в гидроксильной группе пик протона гидроксильной группы не расщепляется и виден в спектре высокого разрешения как синглет.

Если атом водорода заместить на атом дейтерия, спектр поменяется, потому что атом дейтерия резонирует при частотах, сильно отличающихся от частот резонанса атомов водорода.

Сравнение спектров недейтерированного и дейтерированного соединения помогает идентифицировать вещество в тех случаях, когда спектры оказываются очень сложными.

Ядерный магнитный резонанс на атомах углерода ^{13}C (^{13}C -NMR, Carbon-13 NMR)

Ядра углерода C-12 имеют нулевой спин и поэтому неактивны в ЯМР-эксперименте.

Природный углерод содержит около 1 % нуклида C-13, который в соответствующих условиях можно обнаружить в веществе методом ЯМР-спектроскопии.

Исследование магнитного резонанса ядер ^{13}C проводят в таких условиях, в которых не проявляется расщепление сигналов. Поэтому в спектрах ^{13}C -ЯМР каждой группе химически эквивалентных атомов углерода соответствует один пик.

Примеры.

а) Бромэтан $\text{Br}-\text{CH}_2-\text{CH}_3$: два неэквивалентных атома углерода, два пика.

б) 1-бромпропан $\text{Br}-\text{CH}_2-\text{CH}_2-\text{CH}_3$: три неэквивалентных атома углерода, три пика.

в) 2-бромпропан $\text{CH}_3-\text{CH}_2\text{Br}-\text{CH}_3$: две группы неэквивалентных атомов углерода, два пика.

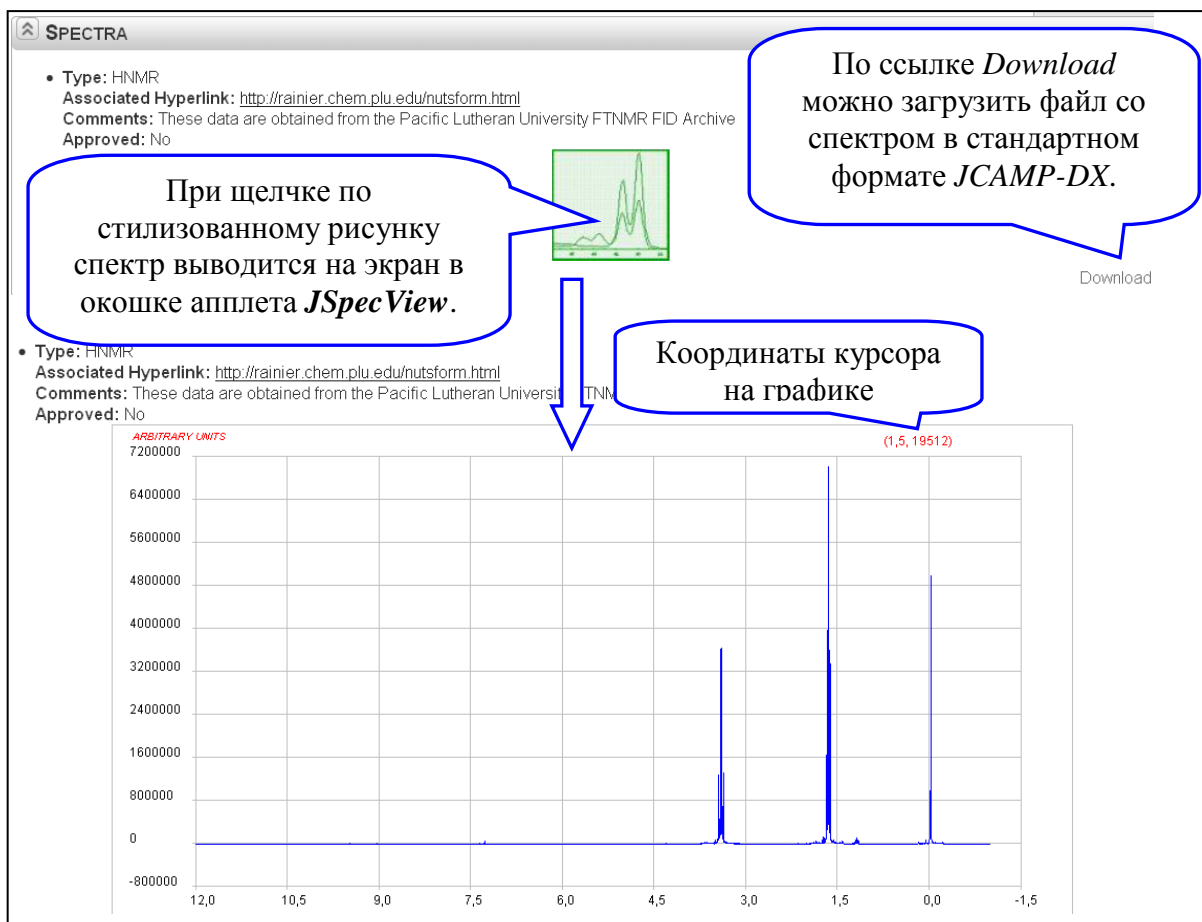
Надеюсь, читатель осознает, что на этих страницах приведено упрощенное и схематичное описание метода ЯМР-спектроскопии.

7.2. ЯМР-спектры в ChemSpider

Знакомство с онлайнowymi репозиториями спектроскопической информации начнем с сайта **ChemSpider** (<http://www.chemspider.com/>). Этот портал собирает в свою базу данных спектры — и уже имеющиеся на других сайтах, и новые, присылаемые исследователями.

ChemSpider начал наполняться спектральной информацией сравнительно недавно, и пока что далеко не каждое соединение здесь охарактеризовано спектрами.

На странице, содержащей сведения о конкретном веществе, ЯМР-спектр может присутствовать в разделе *Spectra*, например:



Кроме этого, на странице вещества в разделе *Data Sources* (закладка *Spectral Data*) приводятся ссылки на внешние спектральные ресурсы.

Самостоятельная работа.

Ознакомьтесь с сайтом; проведите пробный поиск информации.

Контрольное задание 1.

Найдите в базе данных и выведите на экран ^1H -NMR-спектр бромэтана.

(Подсказка. Код *SMILES* — самая простая форма конструирования запроса в этом случае).

Проанализируйте возможности апплета *JSpecView* для работы со спектром (изучите контекстное меню).

Что происходит, если при нажатой левой клавише мышки выделить участок спектра?

Спектр высокого разрешения бромэтана предъявите преподавателю.

Покажите, каким образом можно определить точное значение δ для одного из пиков.

7.3. Справочный сайт Organic Chemistry Info

На сайте [Organic Chemistry Info](http://www.chem.wisc.edu/areas/organic/index-chem.htm) (<http://www.chem.wisc.edu/areas/organic/index-chem.htm>) — кстати, во многих отношениях весьма полезном для химика-органика — размещен комплект справочников, в которых имеется систематизированная информация по ЯМР-спектроскопии, в том числе:

Proton NMR Chemical Shifts

<http://www.chem.wisc.edu/areas/reich/Handouts/nmr-h/hdata.htm>

C-13 NMR Chemical Shifts

<http://www.chem.wisc.edu/areas/reich/handouts/nmr-c13/cdata.htm>.

Справочник представляет собой пакет из пяти частей, каждая из которых содержит значения химических сдвигов для одного типа ядер (^1H , ^{13}C , ^{19}F , ^{31}P , ^{77}Se).

Материал онлайн-справочника собран из печатных справочников по ЯМР-спектроскопии, а также из первоисточников — публикаций в научных журналах.

Структуру справочника рассмотрим на примере части, относящейся к протонному магнитному резонансу.

Справочник предназначен для работы только в режиме *Browse*.

Материал скомпонован в виде двух блоков:

- *Shift Tables* — здесь соотнесены значения δ и структурные фрагменты, дающие соответствующий сигнал;
- классы, группы соединений — здесь приведены значения химических сдвигов протонов индивидуальных веществ (учтите, что информация далеко не всегда сгруппирована в интуитивно понятной форме).

Вся справочная информация приведена только в графическом формате.

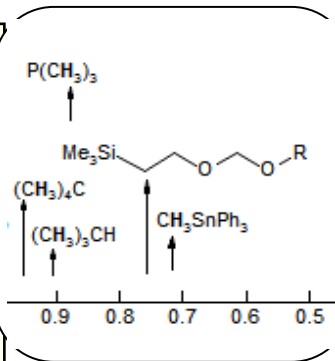
Самостоятельная работа.

Ознакомьтесь с сайтом; проведите пробный поиск спектральной информации, выполните тренировочные упражнения.

Пример:

К другим разделам

Shifts: [C-13](#), [F-19](#), [P-31](#), [Se-77](#)
 Coupling: [J\(HH\)](#), [J\(CH\)](#)
[NMR IR Spectra database](#)



Список сокращений

Proton NMR Chemical Shifts -
[-Explanation](#)
[-Reference Abbreviations](#)
[-Shift Table 1: \$\delta\$ -10 to 2](#)
[-Shift Table 2: \$\delta\$ 2 to 6](#)
[-Shift Table 3: \$\delta\$ 6 to 10](#)
[-Shift Table 4: \$\delta\$ 10 to 20](#)

Соотнесение δ и структурного фрагмента

[Acetal](#)
[Acetylene](#)
[Acridine](#)
[Adamant](#)
[Alcohol](#)
[Aldehyde](#)

Химические сдвиги для индивидуальных веществ

Alcohol See: **Ketone**

<p>3.40 Me-OH JA-79-5361</p>	<p>1.22 (t) OH V-14</p>	<p>1.57 OH V-43</p>	<p>2.3 OH V-44</p>
<p>0.87 S-1897M</p>	<p>3.20 OH V-423</p>	<p>3.82 OH V-423</p>	<p>1.13 OH Reich (Cohen)</p>

Протоны группы CH₂
 $\delta = 3,20$

Протоны групп CH₃
 $\delta = 0,87$

Протон гидроксильной группы: $\delta = 3,82$

Ссылка на источник данных (*Sadtler Spectra Collection*)

Расшифровка по ссылке *Reference Abbreviations* в оглавлении

Ядерный магнитный резонанс на ядрах ¹H.

Тренировочное упражнение.

Найдите значения химического сдвига протонов уксусной кислоты (ищите в группе "Карбоновые кислоты").

Ответ для самопроверки:

CH₃ : $\delta = 2,10$; OH : $\delta = 11,37$.

Контрольное задание 2а.

Найдите значения химического сдвига протонов 2-хлорпропана (ищите в группе "Галиды").

Ядерный магнитный резонанс на ядрах ^{13}C .

Тренировочное упражнение.

Найдите значения химического сдвига ядер ^{13}C в уксусной кислоте.

Ответ для самопроверки:

CH_3 : $\delta = 22,0$; COOH : $\delta = 178,1$.

Контрольное задание 2б.

Найдите значения химического сдвига ядер ^{13}C в 2-хлорпропане.

7.4. The Sadtler Spectral Handbooks

Справочник *Sadtler Spectral Handbooks* создан информационной службой авторитетной компании *Bio-Rad Laboratories*, специализирующейся в области спектроскопии и хемоинформатики.

Справочник состоит из трех частей:

ИК-спектры (*The Sadtler Handbook of Infrared Spectra*)

http://knowitall.informatics.bio-rad.com/handbook/ir/default_ir.asp

^1H -ЯМР-спектры (*The Sadtler Handbook of Proton NMR Spectra*)

http://knowitall.informatics.bio-rad.com/handbook/hnmr/default_hnmr.asp

^{13}C -ЯМР-спектры (*The Sadtler Handbook of Carbon NMR Spectra*)

http://knowitall.informatics.bio-rad.com/handbook/cnmr/default_carbon_nmr.asp

Поисковая программа здесь отсутствует; для обнаружения нужной информации следует открыть оглавление и затем перейти к соответствующему разделу.

Страницы справочника содержат спектры, значения химического сдвига, а также тексты, кратко поясняющие причины формирования спектра именно такой структуры.

Самостоятельная работа.

Ознакомьтесь с сайтом; проведите пробный поиск информации.

Найдите значения химического сдвига для протонов в 1,4-дихлорбензоле.
(Ответ для самопроверки: 7,23).

Контрольное задание 3.

Найдите значения химического сдвига для протонов в метилацетате.

Сколько значений δ необходимо обнаружить в данной задаче?

Веб-страницу с решением предъявите преподавателю.

7.5. NMRShiftDB2

База данных [NMRShiftDB2](http://nmrshiftdb.nmr.uni-koeln.de/) (<http://nmrshiftdb.nmr.uni-koeln.de/>) содержит ^1H - и ^{13}C -ЯМР спектры.

Самостоятельная работа.

Ознакомьтесь с сайтом *NMRShiftDB2*. Проведите пробный поиск информации. Выполните контрольное задание.

База данных предназначена для решения двух задач:

- обнаружение спектра заданного вещества;
- идентификация вещества по известному спектру.

Запрос может быть составлен на поисковом бланке или внесен в виде готового спектрального или структурного файла.

База данных имеет два бланка: основной и усложненный (*expert mode*).

При переходе с Главной страницы по ссылке *Search* по умолчанию открывается основной поисковый бланк:

The screenshot shows the NMRShiftDB2 search interface with several annotations in Russian:

- К усложненному бланку.** (To the complex form.) - points to the "Switch to expert search mode" button.
- Поиск вещества по величинам δ .** (Search for a substance by δ values.) - points to the "Input list" field containing "1. 679" and "3. 427".
- Использовать в качестве запроса файл в формате JCAMP-DX.** (Use as a query file in JCAMP-DX format.) - points to the "Upload file" button.
- Бланк для поиска вещества по спектру.** (Form for searching a substance by spectrum.) - points to the "Search by Spectrum" section.
- Поиск спектра заданного вещества. Бланк текстового запроса.** (Search for the spectrum of a given substance. Text query form.) - points to the "Search by Molecule/Spectrum Properties" section.
- Название, CAS RN и др.** (Name, CAS RN, etc.) - points to the "Chemical Name" dropdown menu.
- Поиск спектра заданного вещества. Бланк структурного запроса.** (Search for the spectrum of a given substance. Structural query form.) - points to the "Search by Structure" section.

Результаты поиска оформлены в виде двух фреймов:

- в левом фрейме приводятся структурные формулы обнаруженных веществ (отсортировано по релевантности);
- в правом фрейме на закладке "*Spectral Data*" приводятся спектры, а на

закладке "Additional Data" — сведения о веществе и о параметрах съемки спектра.

The screenshot displays a search interface for chemical structures. It includes a search bar, a list of results with chemical structures and similarity percentages, and two NMR spectra plots. The 1H NMR spectrum is at the top, and the 13C NMR spectrum is at the bottom. Both spectra show peaks at various chemical shifts. The 1H NMR spectrum has peaks at 3.42, 1.70, and 1.679 ppm. The 13C NMR spectrum has peaks at 27.5, 19.4, 25.82, 27.63, and 27.9 ppm. The interface also includes a table of chemical structures and their similarity percentages.

1H NMR Data:

Atom No.	Mult.(coupling const.)	Meas. Shift	Input Shift	Diff. M-I
1-H		3.42 (H 4, H 5)	3.427	0.01,0.00,
2-H		1.70 (H 6, H 7, H 8)	1.679	0.02,0.00,0.00,

Threshold is 50.00

13C NMR Data:

Atom No.	Mult.(coupling const.)	Meas. Shift	Input Shift	Diff. M-I	expt-0 SpecLib	expt-1 ocmainz inhouse
1	T	27.5	1.679	25.82	27.63	27.9
2	Q	19.4	3.427	15.97	19.53	19.4

Threshold is 50.00

Примечание

По неясным причинам поиск по названию вещества проходит некорректно. В дальнейшей работе рекомендуем проводить текстовый поиск по молекулярной (брутто-) формуле.

Упражнение.

Извлеките ЯМР-спектры монобромэтана.

Рассмотрите структуру страницы результатов поиска.

Контрольное задание 4.

Извлеките ЯМР-спектры бензоат-иона (в запросе заряд иона не указывайте).

Веб-страницу с результатом предьявите преподавателю.

По какой причине в приведенной ниже таблице отсутствуют данные для 5-H и 7-H?

Atom No.	Mult.(coupling const.)	Meas. Shift
4-H		7.894 (H 10)
5-H		
6-H		7.486 (H 12)
7-H		
8-H		7.547 (H 14)

7.6. SDBS — Spectral Database for Organic Compounds

База данных **SDBS** (http://sdb.s.riondb.aist.go.jp/sdb.s/cgi-bin/cre_index.cgi) — представитель того не очень обширного круга солидных бесплатных информационных ресурсов, которые создаются не столько усилиями энтузиастов, сколько научными организациями при мощной поддержке государства. В данном случае мы имеем дело с продуктом японского *National Institute of Advanced Industrial Science and Technology (AIST)*.

В базе данных *SDBS*, содержащей более 30 тыс. органических веществ, постоянно накапливаются спектроскопические данные шести типов: масс-, ЯМР- (на ядрах ^1H и ^{13}C), ИК-, ЭПР- и КР-спектры.

Спектры ЯМР приведены для ок. 15 тыс. веществ; большая их часть получена экспериментально и выверена в лабораториях *A/ST*, меньшая часть рассчитана теоретически.

База данных предназначена для решения двух задач:

- обнаружение спектра заданного вещества;
- идентификация вещества по известному спектру.

Поисковый бланк содержит более 20 полей для формулирования запроса:

SDBS Compounds and Spectral Search

Compound Name:

Molecular Formula:

C, H, then the other elements are alphabetical order, "%*" for the wild card

Molecular Weight:
 to

Numbers between left and right columns
Up to the first place of a decimal point

CAS Registry No.:

"%*" for the wild card.

SDBS No.:

"%*" for the wild card.

Atoms:

C(Carbon)	<input type="text"/>	to	<input type="text"/>
H(Hydrogen)	<input type="text"/>	to	<input type="text"/>
N(Nitrogen)	<input type="text"/>	to	<input type="text"/>
O(Oxygen)	<input type="text"/>	to	<input type="text"/>
F(Fluorine)	<input type="text"/>	to	<input type="text"/>
Cl(Chlorine)	<input type="text"/>	to	<input type="text"/>
Br(Bromine)	<input type="text"/>	to	<input type="text"/>
I(Iodine)	<input type="text"/>	to	<input type="text"/>
S(Sulfur)	<input type="text"/>	to	<input type="text"/>
P(Phosphorus)	<input type="text"/>	to	<input type="text"/>
Si(Silicon)	<input type="text"/>	to	<input type="text"/>

Numbers between left and right columns.

Spectrum:
Check the spectra of your interest.

☐ MS ☐ IR
☐ ¹³C NMR ☐ Raman
☐ ¹H NMR ☐ ESR

IR Peaks(cm⁻¹): Allowance ± 10
", " or space is the separator for multiple peaks.
Use "-", to set a range: eg. 550-750,1650-3000-
Transmittance < %

¹³C NMR Shift(ppm): Allowance ± 2.0
", " is the separator for multiple shifts, eg. 129.3,18.4,...

No shift regions:
Range defined by two numbers separated by a space, eg. 110 78,...

¹H NMR Shift(ppm): Allowance ± 0.2

No shift regions:

MS Peaks and intensities:

Mass and its intensity are a set of data separated by a space, eg. 110 22,...

Hit: hit Sort by:

Почти все поля по умолчанию объединены оператором *AND*.

Исключение составляют поля *CAS Registry No.* и *SDBS No.* — если в одно из этих полей внесен какой-нибудь текст, поиск ведется только по этому полю.

В текстовых графах левой колонки поискового бланка разрешено использование шаблона, причем не только в конце термина, но и в середине, и

в начале.

Формат шаблона: % (знак процента) или * (звездочка).

Шаблон заменяет собой любое количество символов, в том числе, нулевое.

1. Поиск спектра заданного вещества.

Правила заполнения полей бланка.

Compound Name — Название вещества

- Возможно проведение поиска и по систематическому, и по тривиальному названию вещества.
- Разрешено использование шаблона в начале, в середине и в конце текста запроса.
- За пределами шаблона проводится посимвольное сравнение запроса с текстом документа, поэтому, например, задания "benzene" и "benzene " (без пробела и с пробелом) не являются идентичными.

Molecular Formula — Молекулярная (брутто-) формула

- В брутто-формуле элементы должны быть расположены в следующем порядке:
углерод C, водород H, далее остальные элементы по алфавиту.
- Формула может содержать шаблоны.

Molecular Weight — Молекулярная масса

- Допускается заполнение либо обеих граф (минимальное и максимальное значения молекулярной массы), либо только одной графы.

CAS Registry No.

- Если графа содержит текст, поиск ведется только по этому полю.

SDBS No. — регистрационный номер в базе данных SDBS.

- Если графа содержит текст, поиск ведется только по этому полю.

Atoms — Число атомов элемента в брутто-формуле.

- Для каждого из перечисленных на бланке химических элементов допускается заполнение либо обеих граф (минимальное и максимальное значения числа атомов), либо только одной графы.

2. Идентификация вещества по известному ЯМР-спектру.

Правила заполнения полей бланка.

¹³C NMR Shift (ppm)

¹H NMR Shift (ppm)

- Значение химического сдвига (ppm) записывают в графу *NMR Shift*, допустимые отклонения — в графу *Allowance* (по умолчанию здесь установлено значение ± 0.2).
- Для поиска по нескольким сигналам значения химических сдвигов перечисляют через запятую.
Перед и после запятой пробелы ставить нельзя.
- Целая и дробная часть числа разделяются точкой.

Примеры формулирования запроса.

Запрос	Что ищет
верно: 7.65,4.3,2.1	Соединения с ^1H NMR-спектрами, обязательно содержащими сигналы при 7.65 ppm, 4.3 ppm и 2.1 ppm.
неверно: 7.65,4.3, 2.1	Соединения с ^1H NMR-спектрами, обязательно содержащими сигналы при 7.65 ppm и 4.3 ppm. Из-за того, что перед 2.1 есть пробел, поисковая программа проигнорирует это число.

No shift regions

- В этой графе можно указать те части спектра, где не должно быть линий — для этого начало и конец интервала записывают через пробел.
- Если необходимо указать несколько таких интервалов, их перечисляют через запятые.
Перед и после запятой не должно быть пробелов.

Примеры заполнения графы No shift regions.

Запрос	Что ищет
верно: 10 9	Вещества с ^1H NMR-спектрами, которые не имеют пиков в интервале от 10 ppm до 9 ppm.
верно: 10 9,8.5 7,5 3	Вещества с ^1H NMR-спектрами, которые не имеют пиков в интервалах: от 10 ppm до 9 ppm, от 8.5 ppm до 7 ppm, от 5 ppm до 3 ppm.
неверно: 10 9,8.5 7, 5 3	Вещества с ^1H NMR-спектрами, которые не имеют пиков в интервалах: от 10 ppm до 9 ppm, от 8.5 ppm до 7 ppm. Из-за того, что после запятой перед 5 есть пробел, поисковая программа сработает неадекватно.

Список результатов поиска

В списке результатов поиска приводятся обнаруженные вещества и информация о том, какие спектры каждого из веществ имеются в базе данных.

Пример:

SDBS Search Results: 1 - 1 out of 1 hits Sort by: Molecular Weight Ascending Order Search									
SDBS No	Molecular Formula	Molecular Weight	MS	CNMR	HNMR	IR	Raman	ESR	Compound Name
897	C6H12	84.2	Y	Y	Y	Y	Y	Y	cyclohexane

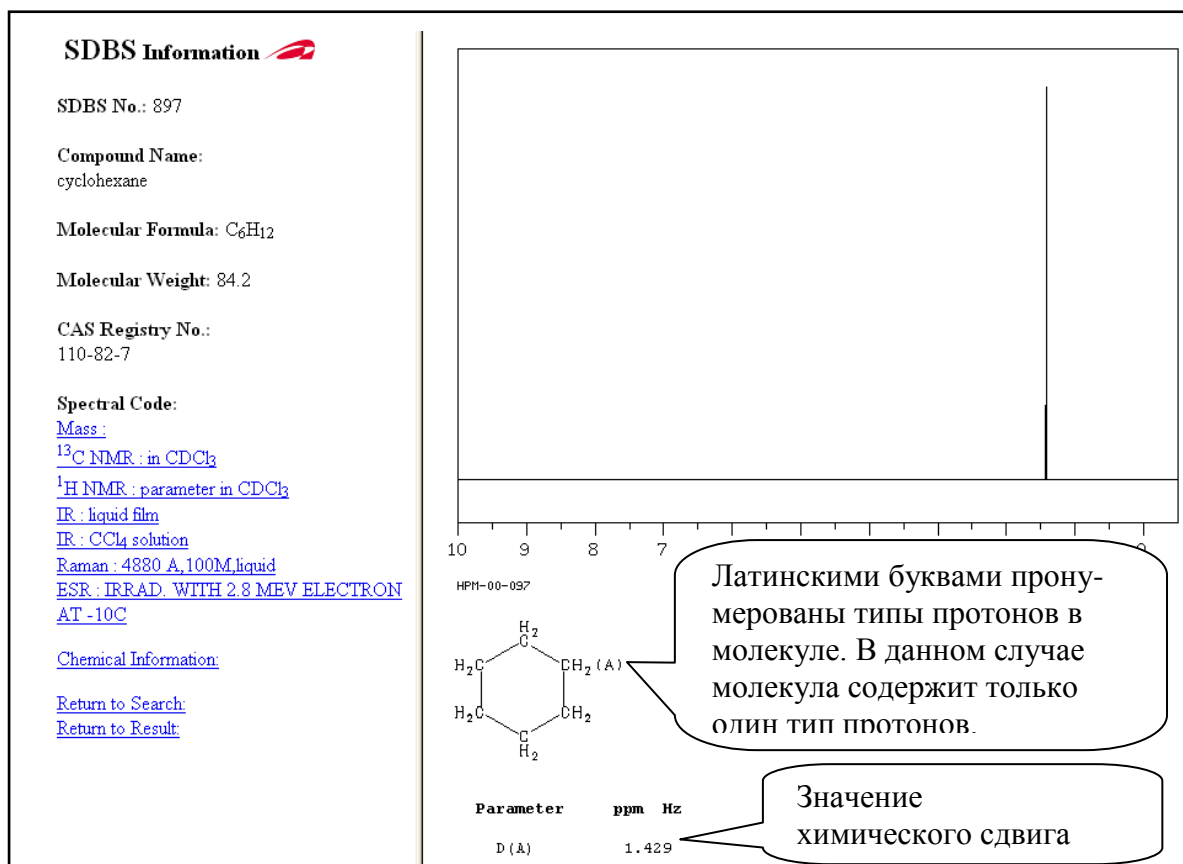
Y (Yes) — спектр этого типа имеется в базе данных;

N (No) — спектр отсутствует.

Тренировочное упражнение

Найдите ^1H -ЯМР спектр циклогексана.

Результат поиска для самопроверки:



Дополнительная информация.

На странице результатов поиска может присутствовать кнопка *peak data* — она выводит на экран спектр в табличной форме, с указанием интенсивности сигнала (в условных единицах) при различных значениях δ .

Сведения об интенсивности сигнала полезно использовать в информационном поиске.

Если возникает необходимость идентификации вещества по известному спектру, содержащему много пиков, то в запрос целесообразно включать сигналы большей интенсивности.

Контрольное задание 5

Сколько бромзамещенных этана имеется в базе данных?

Для какого числа этих соединений в базе данных имеются спектры ^1H -ЯМР?

Для какого числа этих соединений здесь имеются спектры ^{13}C -ЯМР?

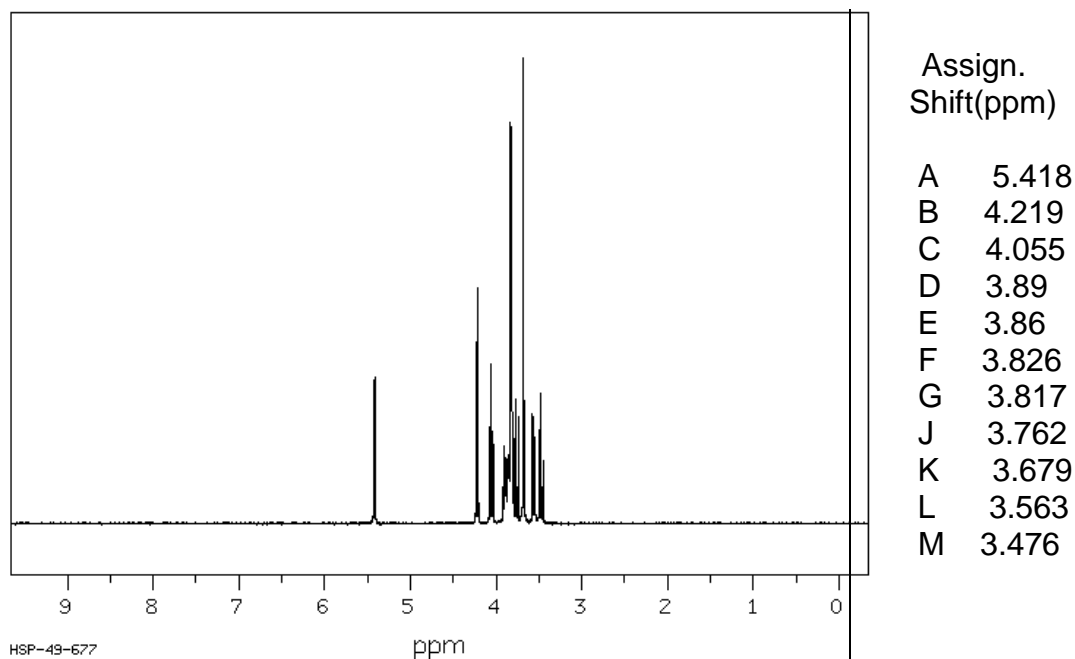
Извлеките ^{13}C -ЯМР-спектр того из бромзамещенных этана, для которого в базе данных отсутствует ^1H -ЯМР-спектр.

Определите значения химических сдвигов для ядер ^{13}C в этом веществе.

Контрольное задание 6

Служба безопасности департамента X во входящей корреспонденции обнаружила конверт со следами белого порошка. Был снят ^1H -ЯМР-спектр подозрительного вещества.

Идентифицируйте это вещество по следующим спектральным данным:



Вспомогательная информация: интенсивность сигнала:

Hz	ppm	Int.	Hz	ppm	Int.
2167.24	5.423	308	1537.60	3.848	99
2163.45	5.414	314	1537.11	3.847	99
1691.28	4.232	391	1536.50	3.845	98
1682.50	4.210	506	1530.40	3.830	862
1628.91	4.076	207	1527.83	3.823	854
1620.48	4.055	344	1524.41	3.815	550
1611.82	4.034	198	1518.07	3.799	55
1565.80	3.918	80	1514.04	3.789	181
1561.89	3.909	96	1504.15	3.764	267
1559.69	3.903	166	1494.87	3.741	231
1557.37	3.897	84	1470.34	3.680	1000
1555.79	3.893	140	1430.42	3.580	235
1553.83	3.888	97	1426.51	3.570	228
1551.27	3.882	135	1420.41	3.555	183
1547.49	3.873	138	1416.50	3.545	187
1542.85	3.861	149	1397.58	3.498	200
1539.92	3.854	140	1388.31	3.474	279

(Int. — интенсивность сигнала, в условных единицах)

Контрольное задание 7.

Вы работаете в лаборатории, контролирующей качество пищевых товаров.

Одна из ваших задач — обнаружение пищевых добавок в продуктах; по этой причине вам нужно иметь под рукой эталонные спектры некоего круга веществ.

Найдите ^1H -ЯМР-спектр соединения, указанного преподавателем. Сформулируйте алгоритм рационального поиска этой информации.

8. СПЕКТРАЛЬНЫЕ БАЗЫ ДАННЫХ: МАСС-СПЕКТРОМЕТРИЯ

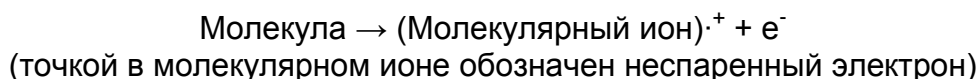
Метод масс-спектрометрии (*Mass Spectrometry*, *MS*) используют для анализа структуры вещества, а в сочетании с хроматографией — для анализа смеси веществ.

8.1. Схема метода масс-спектрометрии.

В основе метода масс-спектрометрии лежит качественный и количественный анализ того, на какие осколки разрушается молекула, если ее бомбардировать электронами.

Исследуемое вещество нагреванием переводят в газообразное состояние и бомбардируют электронами, имеющими высокую энергию.

Исследуемая молекула, получив порцию энергии, теряет при этом собственный электрон и превращается в положительно заряженный молекулярный ион:



Молекулярный ион — агрегат малоустойчивый, поэтому он распадается на части, образуя новые ионы, радикалы и нейтральные молекулы.

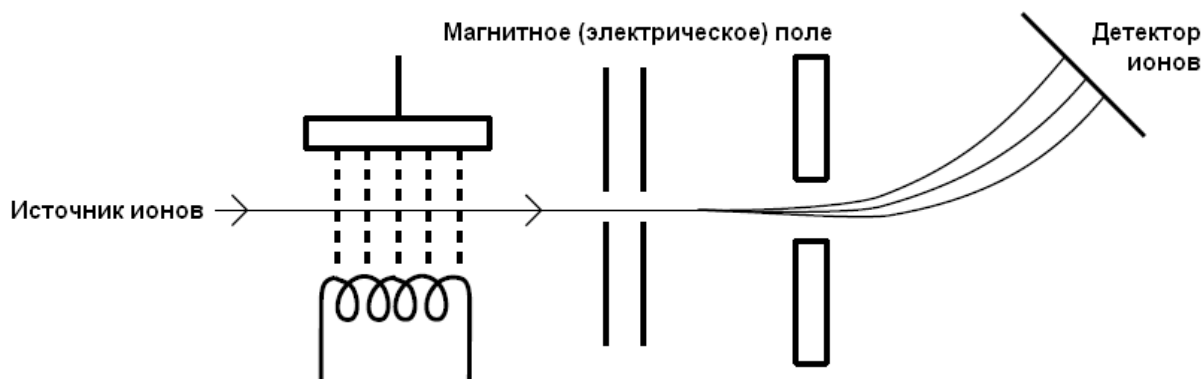
Продукты первичного распада тоже могут быть способны к дальнейшей фрагментации.

В результате эксперимента образуется некий набор заряженных и нейтральных осколков, характерный именно для данного вещества.

Поток осколков направляют в анализатор, где на него воздействуют магнитным или/и электрическим полем.

Заряженная частица в поле отклоняется в большей или меньшей степени. Степень отклонения зависит от величины m/z , где m — масса частицы, а z — ее заряд.

Поток частиц, разделенный на отдельные пучки по величине m/z , попадает в детектор, где каждый пучок регистрируется на определенном участке детектора.

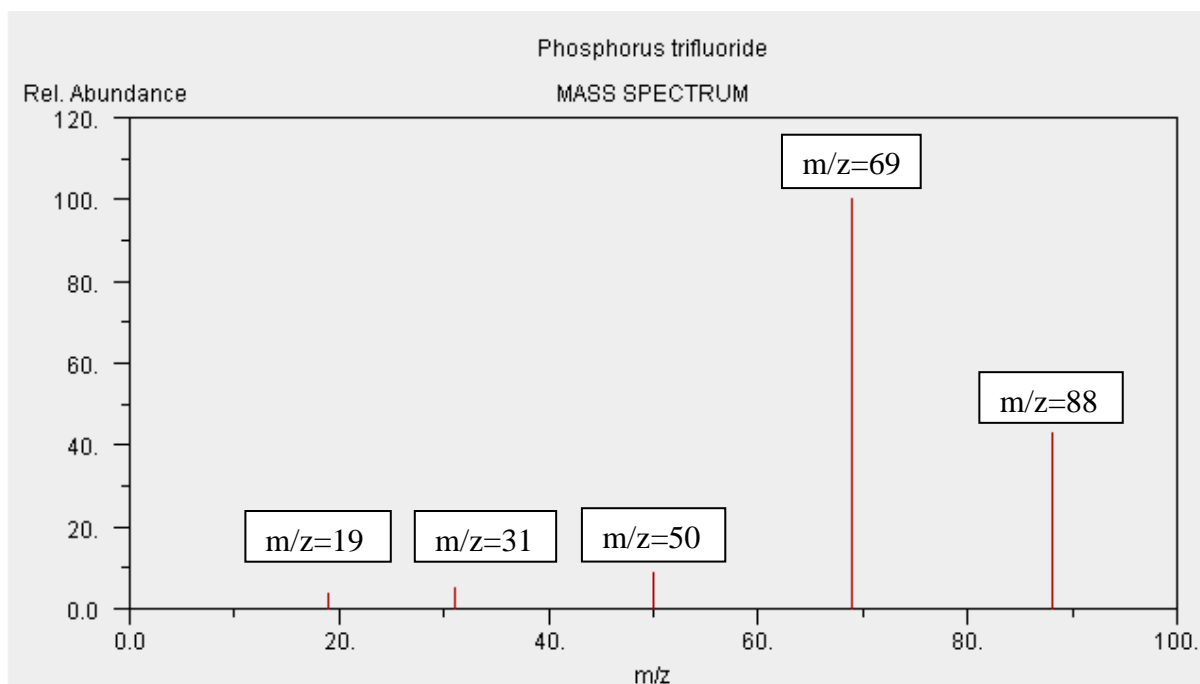


Прибор масс-спектрометр позволяет выделить из всего набора только те осколки, которые несут положительный заряд.

Большая часть осколков несет единичный заряд, но среди фрагментов молекулы могут быть и многозарядные ионы, особенно в тех случаях, когда исходные молекулы большие (например, молекулы белков) или когда многозарядный ион обладает повышенной устойчивостью.

Результат эксперимента выводится в форме спектра, в котором интенсивность сигнала представлена как функция величины m/z .

Пример 1: масс-спектр трифторида фосфора PF_3 :



Интерпретация спектра, приведенного в примере 1.

Учтем, что:

- природный фосфор состоит в основном из одного нуклида: P -31;
- природный фтор состоит в основном из одного нуклида: F -19.

Для рассуждений используем следующие значения относительных молекулярных масс:

- $M_r(P) = 31$.
- $M_r(F) = 19$.
- $M_r(PF_3) = 88$.

Если все частицы, зарегистрированные детектором, однозарядны ($z=1$), то пики спектра можно соотнести следующим образом:

- Сигнал при $m/z=88$ обусловлен молекулярными ионами PF_3^+ .
- Сигнал при $m/z=69$ обусловлен ионами PF_2^+ .
- Сигнал при $m/z=50$ обусловлен ионами PF^+ .
- Сигнал при $m/z=31$ обусловлен ионами P^+ .
- Сигнал при $m/z=19$ обусловлен ионами F^+ .

Как видим, в данном примере сигнал при самом большом значении m/z принадлежит молекулярному иону.

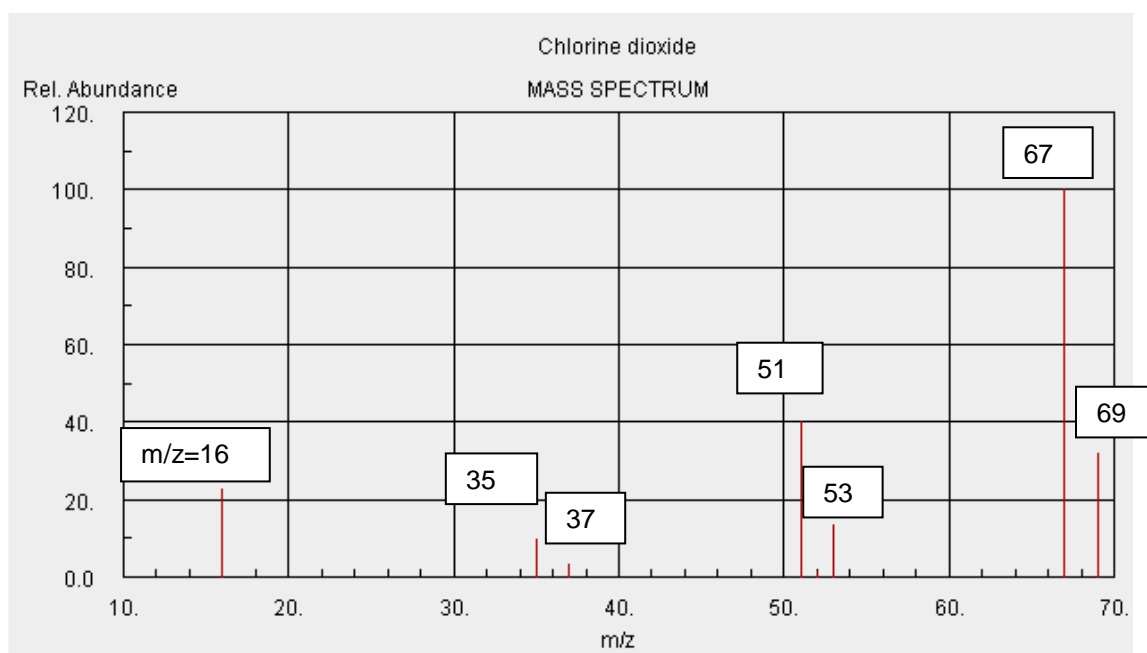
Такая ситуация встречается часто, но не всегда.

- Если молекулярный ион нестабилен и легко распадается, то сигнал с самым большим значением m/z будет принадлежать одному из осколков.
- Если вещество содержит примесь, то сигнал с самым большим значением m/z может принадлежать постороннему соединению.

Тем не менее, в огромном числе случаев максимальная в спектре величина m/z численно равна молярной массе исследуемого вещества. Этот факт полезен в качественном анализе.

Пример 2: масс-спектр диоксида хлора ClO_2 .

Обратите внимание на наличие парных пиков, отстоящих на 2 единицы друг от друга.



Контрольное задание 1.

Соотнесите сигналы спектра (*пример 2*) — определите, какой частице соответствует каждый из пиков.

(Учтите, что природный хлор состоит в основном из двух нуклидов: ^{35}Cl и ^{37}Cl).

8.2. Масс-спектры в NIST Chemistry WebBook

База данных Национального института стандартов и технологии США [NIST Chemistry WebBook](http://webbook.nist.gov/chemistry/) (<http://webbook.nist.gov/chemistry/>) содержит масс-спектры органических и неорганических веществ.

Масс-спектр на веб-странице соответствующего вещества выводится в окошке апплета, поэтому спектр является не просто графическим изображением, а интерактивным объектом.

Для офлайнового использования спектр можно получить в файлах форматов *GIF*, *SVG* и *JCAMP-DX* (ссылки *Additional Data*).

Упражнение.

Извлеките масс-спектр пентакарбонила железа $\text{Fe}(\text{CO})_5$, рассмотрите

структуру веб-страницы и самого спектра.

Исследуйте полезные функции апплета. Обратите внимание, что:

- Координаты курсора отображаются в правом верхнем углу окошка апплета.

Если установить курсор на пик, то можно прочитать соответствующее значение m/z .

- Для того, чтобы отобразить участок спектра в увеличенном масштабе, этот участок следует выделить курсором при нажатой клавише мышки. Для восстановления исходного масштаба следует щелкнуть по полю апплета.

Контрольное задание 2.

Соотнесите пики в спектре **Fe(CO)₅** — укажите, каким частицам соответствует каждый из сигналов.

При соотнесении пиков учтите, что природное железо состоит из нескольких нуклидов.

Для химика, изучающего структуру вещества, весь спектр имеет информационную ценность.

По тому, какие осколки образуются при разрушении молекулы, химик делает выводы о прочности связей в ней, об имеющихся структурных фрагментах, о распределении электронной плотности и т. д.

Наметанный взгляд даже из общего вида спектра извлекает информацию.

Упражнение.

Найдите в NIST Chemistry WebBook и выведите на экран масс-спектры трех углеводородов состава C₆H_n: *n*-гексана, циклогексана и бензола.

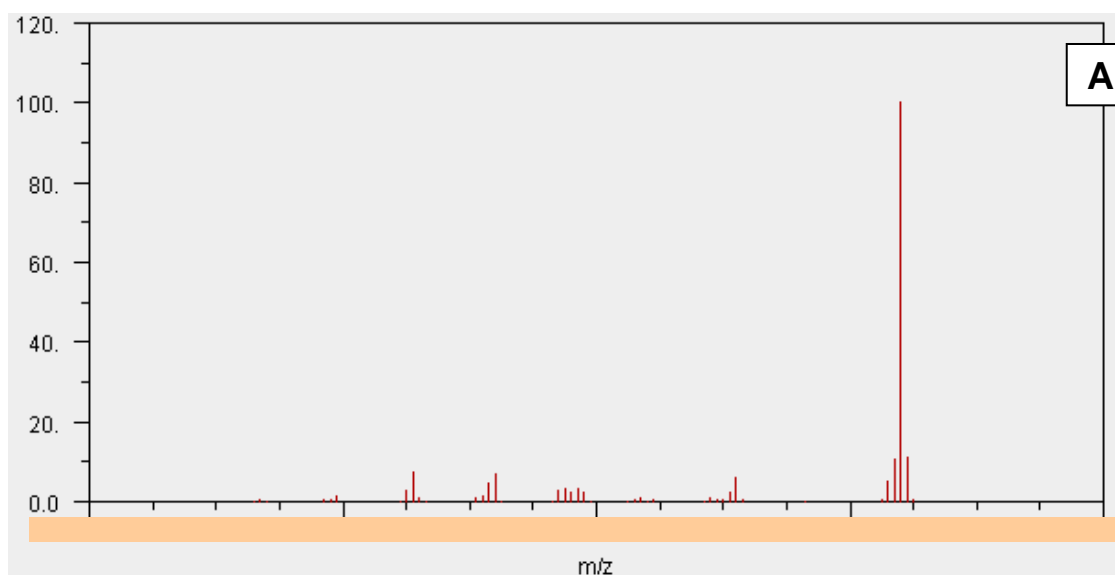
Сравните общий вид трех извлеченных спектров.

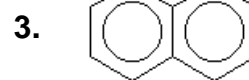
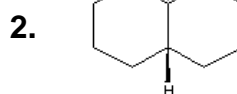
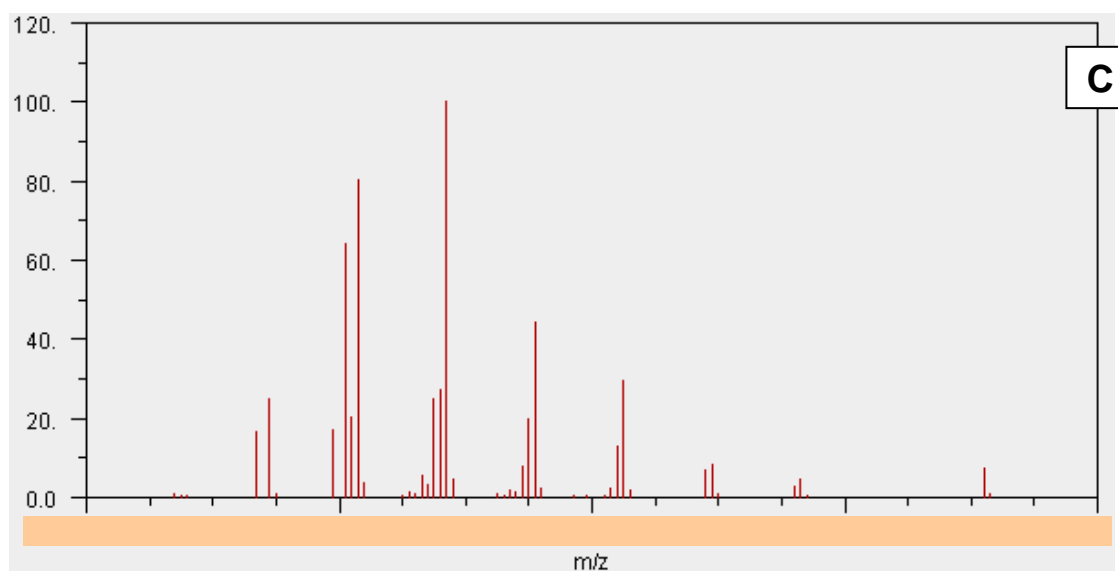
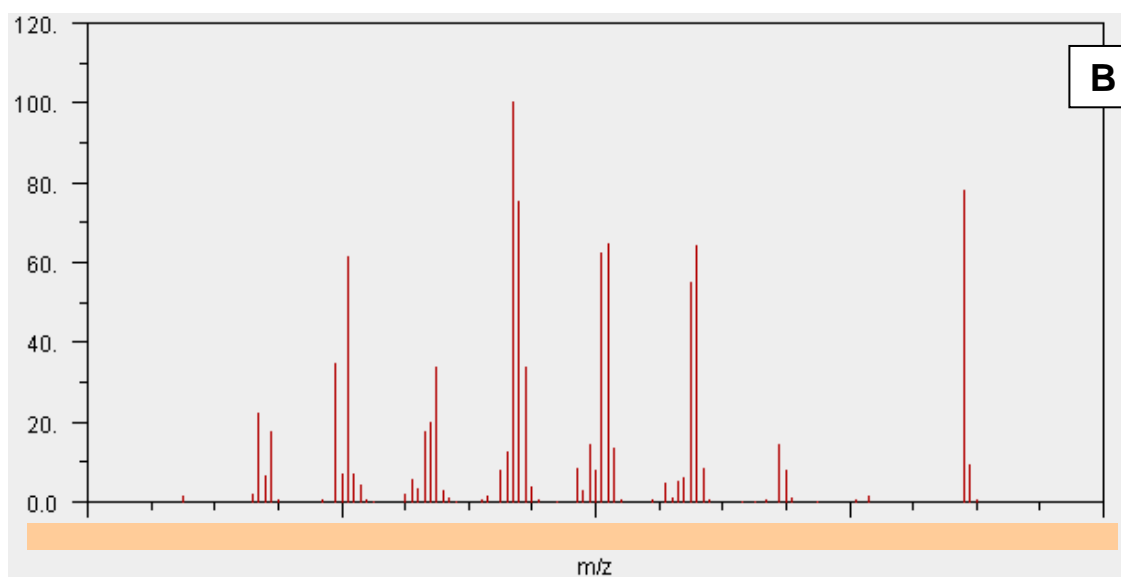
По каким внешним признакам различаются спектры алифатического ациклического, алифатического циклического и ароматического углеводородов?

Предложите разумное объяснение.

Контрольное задание 3.

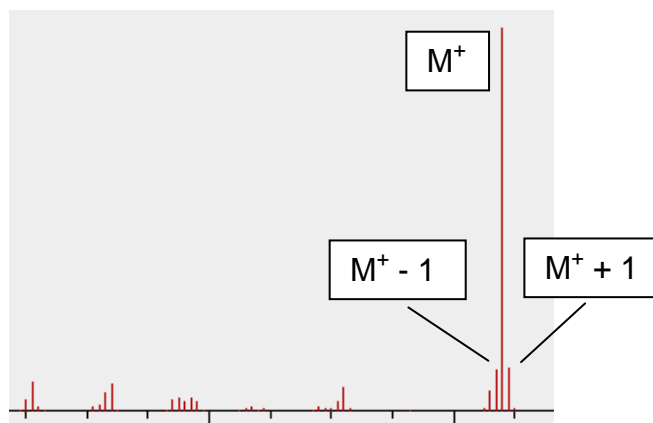
По внешнему виду масс-спектров, *не обращаясь к базе данных*, соотнесите спектры **A, B, C** и структуры **1, 2, 3**.





Контрольное задание 4.

В каждом из приведенных выше спектров правее пика молекулярного иона M^+ наблюдается сигнал небольшой интенсивности $M^+ + 1$, например:



4а) Чем можно объяснить появление сигнала $M^+ + 1$ в спектрах органических соединений?

4б) Чем можно объяснить появление сигнала $M^+ - 1$ в спектрах органических соединений?

8.3. Масс-спектры в Spectral Database for Organic Compounds (SDBS)

База данных спектров органических соединений [SDBS](http://riodb01.ibase.aist.go.jp/sdb/cgi-bin/direct_frame_top.cgi) (http://riodb01.ibase.aist.go.jp/sdb/cgi-bin/direct_frame_top.cgi) содержит около 25 тыс. экспериментально определенных масс-спектров.

Поисковая программа позволяет обнаруживать масс-спектр заданного вещества, а также идентифицировать вещество по известному спектру.

В базе данных *SDBS* приемы работы с масс-спектрами принципиально не отличаются от приемов работы с ЯМР-спектрами.

В списке результатов поиска масс-спектры находятся в колонке *MS*

SDBS No	Molecular Formula	Molecular Weight	MS	CNMR	HNMR	IR	Raman	ESR	Compound Name
898	C ₆ H ₆	78.1	Y	Y	Y	Y	Y	Y	benzene
21952	C ₅ H ₅ N	79.1	N	Y	N	Y	N	N	pyridinium m-nitro

Самостоятельная работа.

Проведите пробный поиск масс-спектра какого-нибудь вещества, например, *n*-гексана. Рассмотрите структуру страницы результатов поиска.

Для **поиска вещества по известному масс-спектру** на поисковом бланке имеется соответствующая графа:

MS Peaks and intensities:

Mass and its intensity are a set of data separated by a space, eg. 110 22,..

Правила формулирования запроса в графе *MS Peaks and intensities*:

- Значения m/z перечисляются через запятую, **перед и после запятой не должно быть пробелов**.
- Рекомендуется указывать относительную интенсивность пика — в этом случае поиск оказывается более эффективным.
- Относительная интенсивность пика записывается через пробел после значения m/z .

Пример 1. По запросу:

78,52,39

извлекаются вещества, в масс-спектрах которых имеются, в частности, такие пики:

- $m/z=78$;
- $m/z=52$;
- $m/z=39$.

Пример 2. По запросу:

78 100,52 18,39 10

извлекаются вещества, в масс-спектрах которых имеются, в частности, такие пики:

- $m/z=78$ (относительная интенсивность 100);
- $m/z=52$ (относительная интенсивность 18);
- $m/z=39$ (относительная интенсивность 10).

Контрольное задание 5.

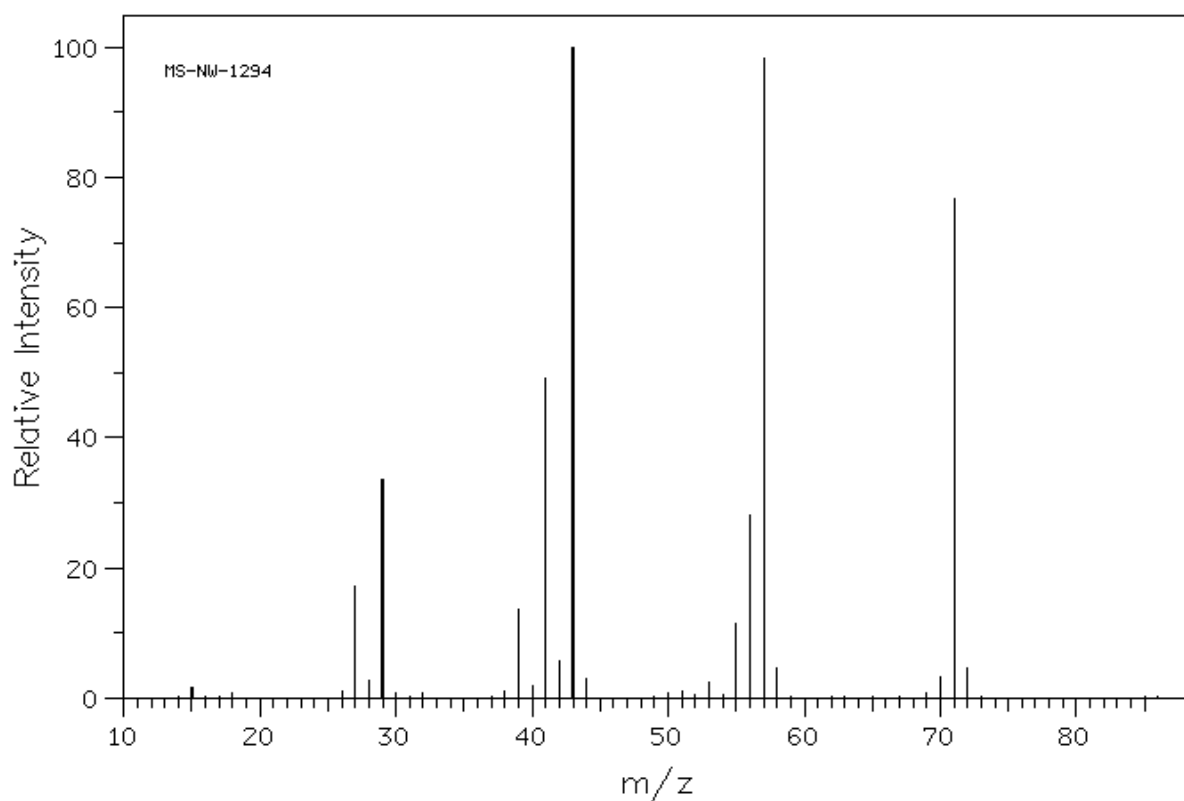
Идентифицируйте вещество по следующим экспериментальным данным (сформулируйте запрос и проведите поиск в *SDBS*):

SDBS-Mass

MS-NW- [REDACTED]

SDBS NO. [REDACTED]

(Mass of molecular ion: [REDACTED])



Source Temperature: 260 °C
 Sample Temperature: 180 °C
 RESERVOIR, 75 eV

(смотрите продолжение на следующей странице)

m/z	Интенсивность	m/z	Интенсивность
15.0	1.6	44.0	3.1
26.0	1.1	51.0	1.1
27.0	17.2	53.0	2.5
28.0	2.6	55.0	11.3
29.0	33.6	56.0	28.0
38.0	1.0	57.0	98.3
39.0	13.6	58.0	4.7
40.0	1.8	70.0	3.3
41.0	49.1	71.0	76.7
42.0	5.6	72.0	4.5
43.0	100.0		

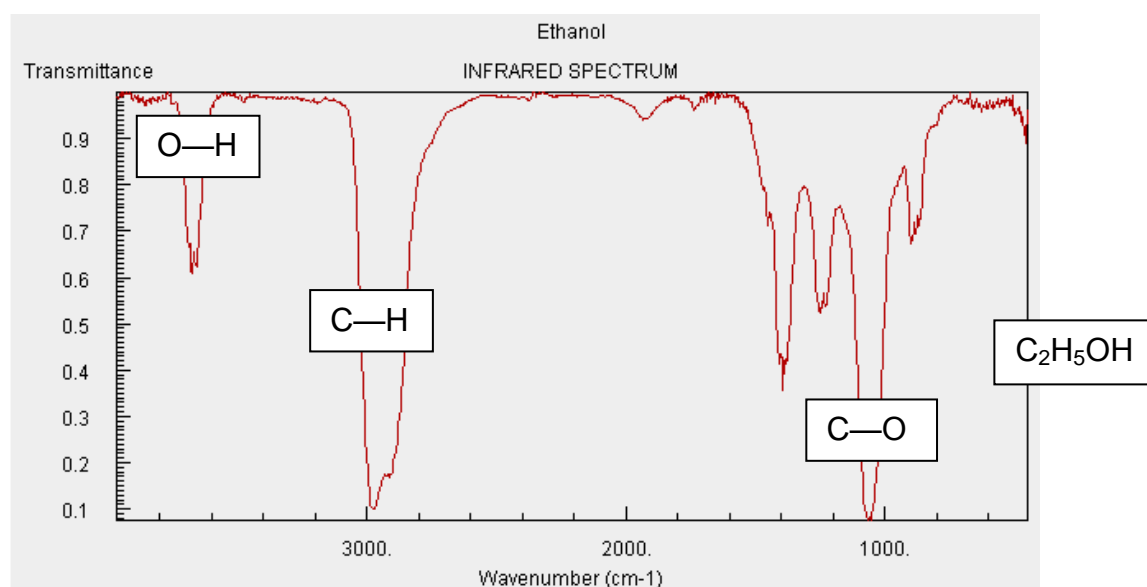
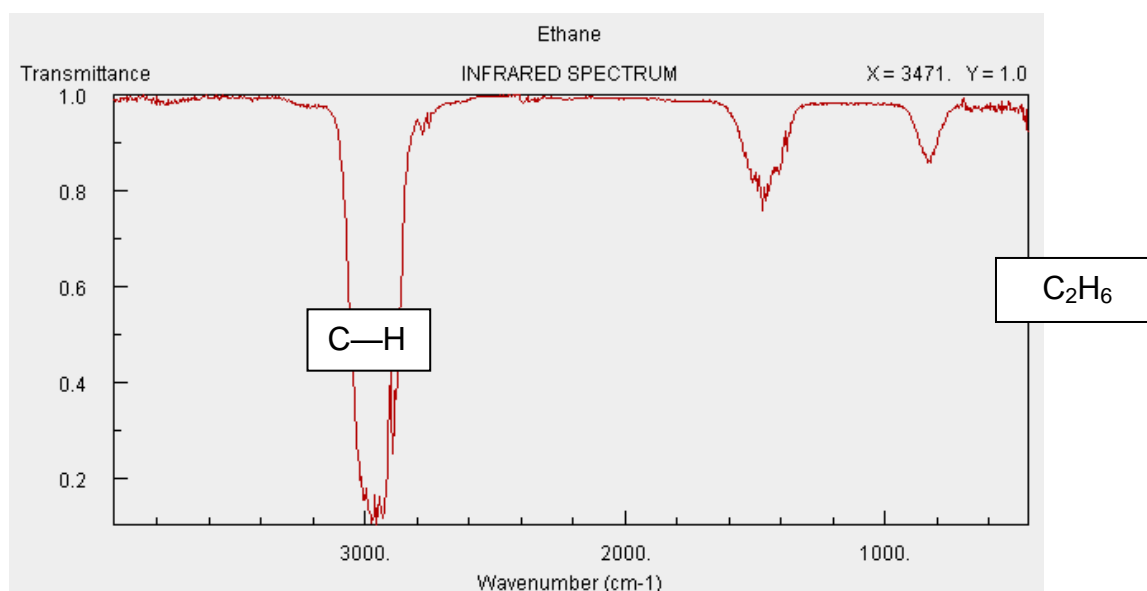
9. СПЕКТРАЛЬНЫЕ БАЗЫ ДАННЫХ: ИНФРАКРАСНАЯ (ИК) СПЕКТРОСКОПИЯ

9.1. Схема метода ИК-спектроскопии (Infra-Red (IR) Spectroscopy)

Атомы в молекулах связаны в единое целое химическими связями и находятся в постоянном колебательном движении. Частота колебаний зависит от силы химической связи и масс связанных атомов, что и используется для распознавания структуры вещества в методе ИК-спектроскопии.

Переходы между колебательными энергетическими уровнями лежат в инфракрасном диапазоне. Если вещество облучать инфракрасным светом разной частоты, оно будет поглощать кванты энергии избранных частот — тех, которые характерны для имеющихся в веществе структурных фрагментов.

Примеры ИК-спектров. Этан и этанол:



Некоторые пути использования ИК-спектров:

- по имеющимся полосам в спектре соединения, состав которого неизвестен, судят о структуре этого вещества;
- по наличию лишних полос в спектре известного соединения судят о чистоте данного вещества.

У органических соединений в области от 1400 см^{-1} до 800 см^{-1} обычно находится много пиков разной интенсивности и формы. Вид спектра здесь настолько специфичен для конкретного вещества, что эту часть спектра называют "отпечатком пальцев". По виду спектра вещество можно идентифицировать (конечно, если имеется библиотека эталонных спектров).

9.2. Онлайн-базы данных

Многие бесплатные базы данных, содержащие ИК-спектры, нам знакомы.

Самостоятельная работа

Посетите перечисленные ниже сайты, обращая внимание на процесс поиска и вывода на экран именно ИК-спектров.

[NIST Chemistry WebBook](http://webbook.nist.gov/chemistry/)

<http://webbook.nist.gov/chemistry/>

[SDBS - Spectral Database for Organic Compounds](http://riodb01.ibase.aist.go.jp/sdbs/cgi-bin/direct_frame_top.cgi)

http://riodb01.ibase.aist.go.jp/sdbs/cgi-bin/direct_frame_top.cgi

[The Sadtler Handbook of Infrared Spectra](http://knowitall.informatics.bio-rad.com/handbook/ir/default_ir.asp)

http://knowitall.informatics.bio-rad.com/handbook/ir/default_ir.asp

Краткая таблица характерных частот поглощения:

[Characteristic IR Absorption Frequencies of Organic Functional Groups](http://www2.ups.edu/faculty/hanson/Spectroscopy/IR/IRfrequencies.html)

<http://www2.ups.edu/faculty/hanson/Spectroscopy/IR/IRfrequencies.html>

Контрольное задание.

Предположим, вы работаете в лаборатории фирмы, выпускающей пищевые товары.

Одна из ваших задач — контроль сырья; по этой причине вам нужно иметь под рукой эталонные спектры некоего круга веществ.

Найдите ИК-спектр соединения, указанного преподавателем.

Сформулируйте алгоритм рационального поиска этой информации.

10. SPECTRAL GAME

Spectral Game (<http://spectralgame.com/>) — проверка умения соотносить структуру и спектр.

Ход работы.

1. Зарегистрируйтесь. В графе *Name* укажите первые пять букв вашей фамилии, остальные поля регистрационного бланка можно оставить пустыми.
2. Переключатель *Viewer* рекомендуем установить в положение *JSpecView*.
В меню *Type*: выберите тип спектра (например, *¹H NMR*).
3. Щелкните по слову *Play* — начнется загрузка первого задания.
4. Ваша задача — щелкать по той из двух предъявляемых структур, для которой показан спектр.
Спектры извлекаются из базы данных *ChemSpider*.
Тестирование продолжается до первой ошибки.
5. По завершении тестирования на экран выводится результат — покажите его преподавателю.

Интересно знать.

Чемпионом и рекордсменом мира по распознаванию ИК-спектров является студент химического факультета Белорусского государственного университета Тимофей ТИХОМИРОВ (*user: Timofei*).

Мировой рекорд — 51 балл, — установленный Т. Тихомировым в 2011 году, до сих пор (июнь 2013 г.) остается наивысшим достижением среди всех номинаций международного чемпионата *Spectral Game* (см. таблицу по ссылке [Scores](#) на Главной странице сайта).

11. ПАКЕТ ПРОГРАММ ACD/Labs FREEWARE

Пакет **ACD/Labs Freeware** состоит из двух автономных, но взаимосвязанных программ:

- **ACD/ChemSketch** — молекулярный редактор двумерных химических структур и графический редактор,
- **ACD/3D Viewer** — программа моделирования и визуализации трехмерных структур.

В основной своей части принципы работы с *ACD/Labs Freeware* аналогичны принципам работы с теми редакторами, которые вам уже знакомы, поэтому вы должны самостоятельно разобраться с основными функциями программ этого пакета. Некоторые особенности пакета *ACD/Labs Freeware v.12.0* поясняются в прилагаемых инструкциях.

11.1. ACD/ChemSketch как молекулярный редактор двумерных химических структур.

11.1.1. Введение

Запуск редактора: *Программы -> ACDLABS 12.0 -> ChemSketch*

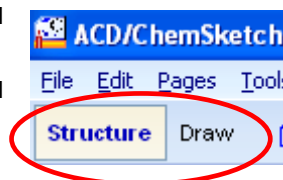
ChemSketch работает в двух режимах:

- **Structure (Структура)** — молекулярный редактор.
Изображаемые атомы и химические связи являются элементами химической структуры и имеют соответствующие свойства.
- **Draw (Рисунок)** — графический редактор.
Все изображаемые элементы являются частями обычного рисунка.

Для переключения между режимами служат кнопки *Structure* и *Draw*:

Переключение также происходит при нажатии клавиши "пробел".

По умолчанию загружается режим *Structure*.



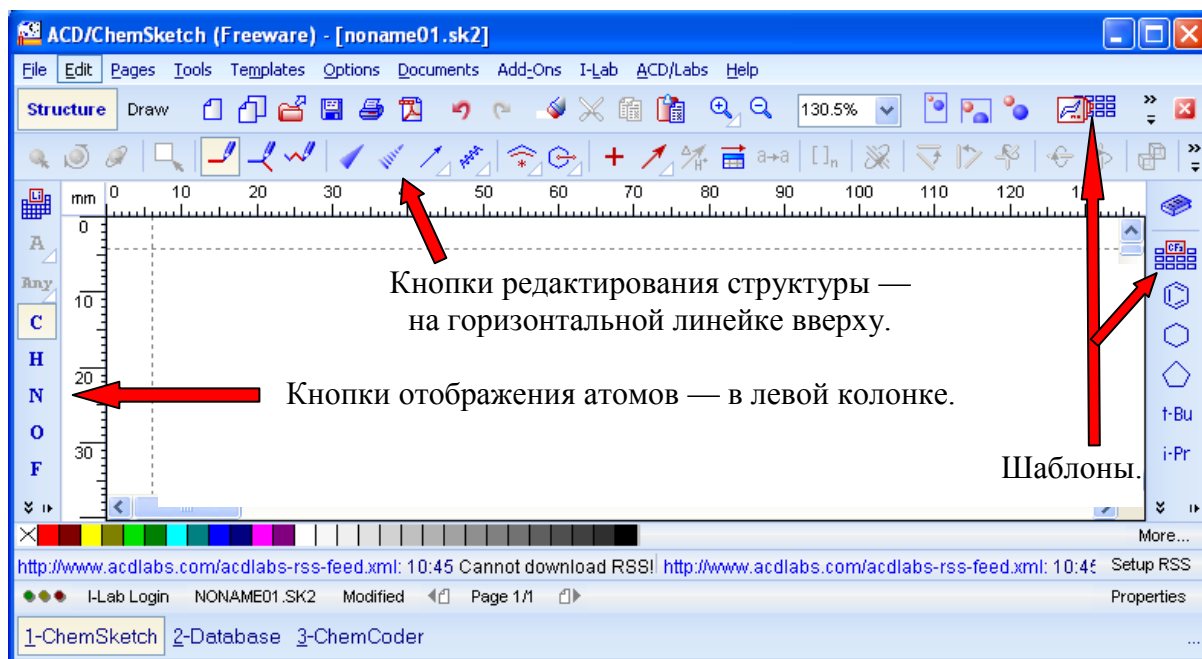
Для справки:

Сайт разработчика: [ACD/Labs](http://www.acdlabs.com/home/) (<http://www.acdlabs.com/home/>)

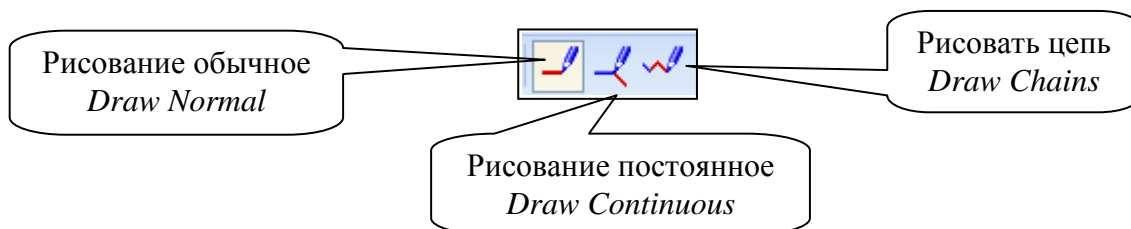
Инсталляторы: <http://www.acdlabs.com/resources/freeware/>

11.1.2. Работа в режиме Structure

В основной своей части интерфейс *ChemSketch* стандартен для такого рода программ:



При загрузке *ChemSketch* по умолчанию включаются кнопки:
С "атом углерода" и **"Обычное рисование"** (*Draw Normal*).



Выполнение основных операций при нажатой кнопке **Draw Normal**:

- **добавить связь** в стандартном направлении — щелкнуть по атому;
- **добавить связь** в заданном направлении — щелкнуть по атому и, не отпуская клавишу мышки, передвинуть курсор в нужном направлении;
- **нарисовать связь между** имеющимися атомами — щелкнуть по первому атому и, не отпуская клавишу мышки, тянуть связь ко второму атому;
- **изменить порядок связи** — щелкнуть по связи.

Действие кнопки **Draw Continuous** аналогичное — кроме того, что:

- при первом щелчке по атому атом выделяется, а добавление связи происходит при втором щелчке.

Draw Chains = "Рисовать цепь":

- щелкнуть по атому и тянуть цепь в нужном направлении на нужную длину.



— кнопка "**Периодическая таблица**"

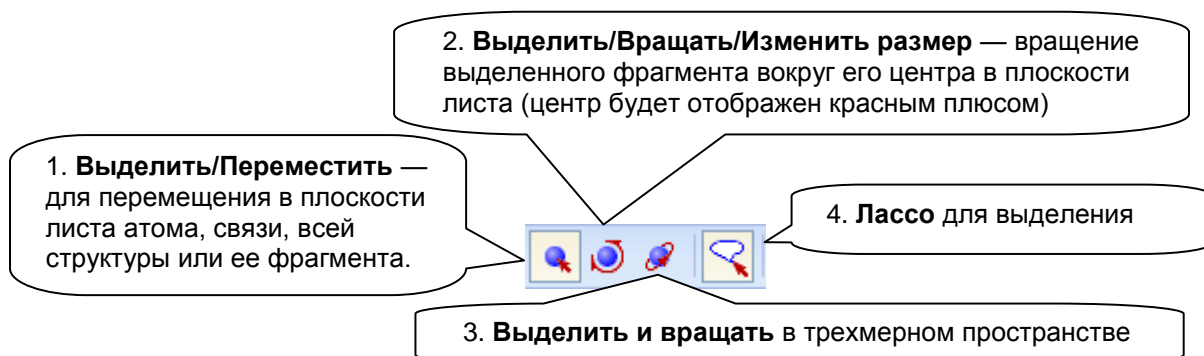
Если в левой колонке отсутствует кнопка какого-нибудь химического элемента, ее добавляют из Периодической системы.

Кнопка остается на панели на весь сеанс.

Чтобы во время сеанса удалить такую кнопку, следует дважды щелкнуть по панели и затем в появившемся окошке подтвердить удаление.

Кнопки выделения.

Программа содержит 4 кнопки для выделения структуры или ее части и последующей манипуляции с выделенным объектом.



При отжатой кнопке "**Лассо**" выделяется прямоугольник, при нажатой — площадь любой конфигурации.

Особенностью кнопок **2**, **3**, **4** является то, что процесс проводится в две стадии:

- при выделении объекта в узлах (на атомах, в центрах связей) появляются маленькие белые квадратики,
- для манипуляции с выделенным объектом необходимо установить курсор на любой из этих квадратиков, и если они закрашиваются в черный цвет, можно выполнять соответствующее действие.



Удалить = *Delete*

Удаление атома или связи:

- Нажать кнопку "**Удалить**" (обратите внимание, что изменяется форма курсора), затем щелкнуть по атому или связи. Атом удаляется со всеми его связями, причем, если в результате этой операции образовались бы одиночные несвязанные атомы, то они исчезают тоже.

Если необходимо удалить только центральный атом, но сохранить периферийные, то удаление проводят при нажатой кнопке *Ctrl*.

Удаление фрагмента или всей структуры:

- Нажать кнопку "**Удалить**", обвести курсором удаляемую структуру или ее часть. Обведенная курсором структура выделяется — на атомах и связях появляются маленькие квадраты. При наведении курсора на квадрат, он становится активным — закрашивается черным цветом. Если щелкнуть по любому черному квадрату, выделенная структура удаляется.

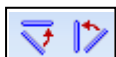
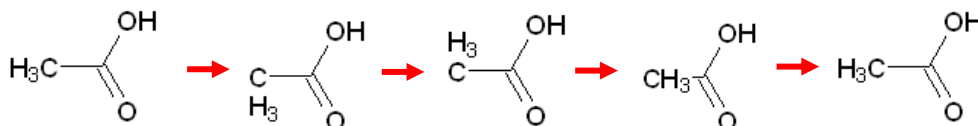


Изменение внешнего вида структуры:



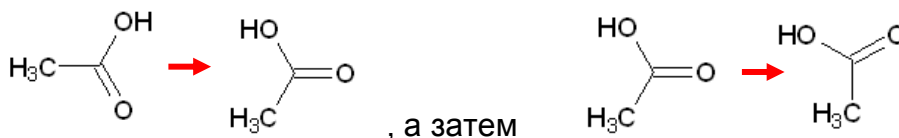
Для изменения **положения атомов водорода**, для изменения способа **отображения двойной связи** следует при нажатой кнопке щелкать по группе CH_n или по связи.

Упражнение. Проведите следующие преобразования:



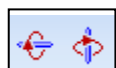
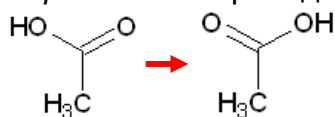
Для преобразования структуры так, чтобы указанная связь оказалась **горизонтальной** или **вертикальной**, при нажатой соответствующей кнопке следует щелкнуть по этой связи.

Упражнение. Проведите следующие преобразования:



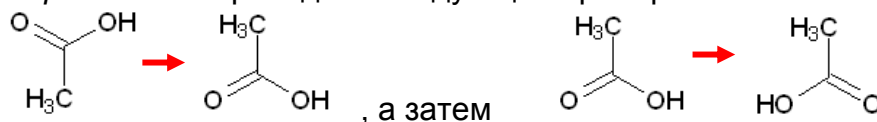
Вращение на 180° вокруг связи — при нажатой кнопке следует щелкнуть по соответствующей связи.

Упражнение. Проведите следующее преобразование:



Вращение на 180° по осям x и y — щелкнуть по одной из этих кнопок.

Упражнение. Проведите следующие преобразования:



"Полимер" — для заключения в скобки звена полимера.

При щелчке по кнопке "Полимер" появляется окошко, в котором следует указать параметры:

Index — число полимеризации,

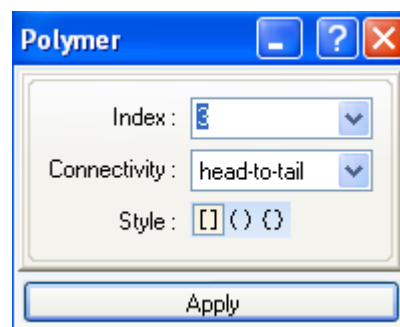
Connectivity — порядок соединения мономерных звеньев:

"голова к хвосту" (*head-to-tail*),

"голова к голове" (*head-to-head*),

"любое" (*either/unknown*).

Style — вид скобок.



После нажатия кнопки *Apply* следует щелкнуть по левой границе звена цепи и, не отпуская клавиши мышки, передвинуть курсор на правую границу звена.

Упражнение. Проведите следующее преобразование:

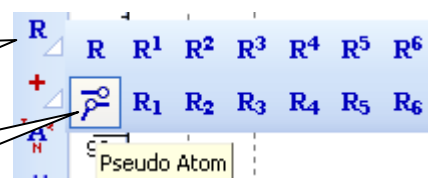


Если при отображении полимера необходимо указать концевую связь без атома, концевой атом заменяют "псевдо-атомом":

1. В левой колонке кнопок щелкнуть по треугольнику кнопки **R** — открывается список

2. Щелкнуть по кнопке "Псевдо-атом"

3. Щелкнуть в структуре по концевым атомам углерода.



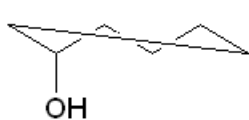
Упражнение:



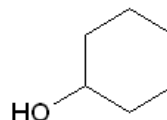
Исправить изображение структуры = Clean Structure

Кнопка позволяет "очистить структуру" — стандартизировать длины связей и углы между связями и сделать ее внешне аккуратной.

Упражнение:



превратите в структуру



Блок кнопок для написания **уравнений химических реакций**:

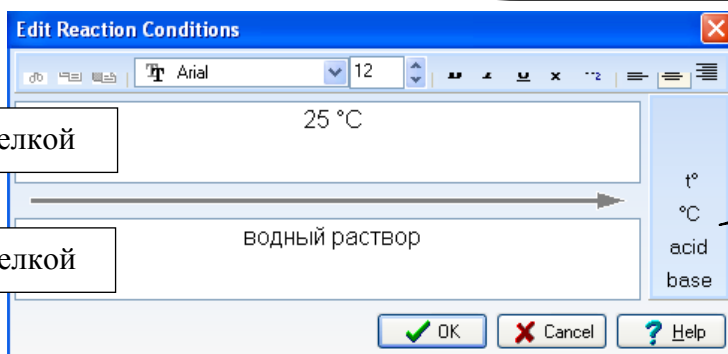
Знак "плюс"



Калькулятор реагентов.

Стрелка между левой и правой частями уравнения

Для написания текста над стрелкой и под стрелкой. При нажатии открывается окошко:



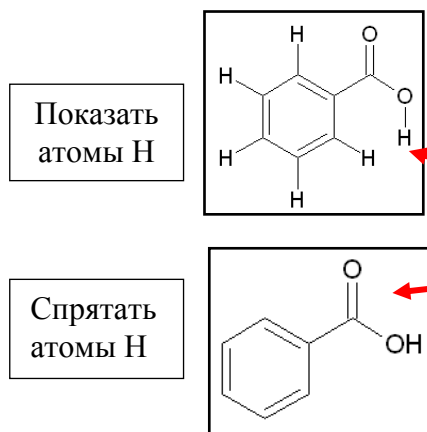
над стрелкой

под стрелкой

готовые шаблоны

Полезные команды в меню. Раздел Tools.

На экране можно показывать или скрывать те атомы водорода, которые насыщают свободные валентности (*Explicit Hydrogens*).



Name for Structure	Ctrl+Shift+I
Stereo Descriptors	
Stereo Descriptors Options...	
SMILES Notation	
Structure from SMILES	
InChI for Structure	
InChI Options...	
Structure from InChI	

Программа умеет генерировать систематическое название, *R/S*-стереодескрипторы, коды *SMILES* и *InChI*, а также генерировать структурные формулы по известным кодам *SMILES* и *InChI*.

Программа умеет рассчитывать физико-химические параметры вещества.

Tools	Templates	Options	Documents	Add-On
Structure Properties				Alt+Shift+S
Clean Structure				F9
Check Tautomeric Forms				Ctrl+Shift+T
3D Structure Optimization				Ctrl+Shift+3
MassSpec Scissors				
Show Aromaticity				Ctrl+Shift+A
Hide Aromaticity				Ctrl+Shift+H
Expand Shorthand Formulae				Ctrl+Shift+F
Add Explicit Hydrogens				Ctrl+Shift+Y
Remove Explicit Hydrogens				Ctrl+Shift+R
Bring Bond(s) to Front				Ctrl+F
Send Bond(s) to Back				Ctrl+K
Auto Renumbering				Ctrl+Shift+N
Clear Numbering				Ctrl+Shift+L
Generate				
Search for Structure...				Ctrl+Shift+C
Calculate				

Molecular Formula
Formula Weight
Composition
Molar Refractivity
Molar Volume
Parachor
Index of Refraction
Surface Tension
Density
Dielectric Constant
Polarizability
Monoisotopic Mass
Nominal Mass
Average Mass
M+
M-
[M+H] ⁺
[M+H] ⁻
[M-H] ⁺
[M-H] ⁻
All Properties
Select Properties to Calculate...
Selected Properties



Кнопка, дублирующая команду *Name for Structure*



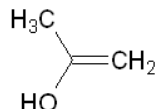
Кнопка, дублирующая команду *InChI for Structure*

Еще несколько особенно полезных кнопок



Check for Tautomeric Forms — Генерирование устойчивых **таутомерных** форм.

Упражнение. Определите, какие таутомерные формы имеются у данной структуры и какая из форм более устойчива:



Формула Маркуша.

Порядок применения.

В структурной формуле выделяют участок, с которым может быть связан радикал, и щелкают по кнопке. Далее, если необходимо, стандартным путем заменяют *R* нужными атомами.



3D Optimization — Генерирование **трехмерной** структуры.

Перед началом процесса генерирования программа может спросить, убирать ли атомы водорода (временно). Следует соглашаться — это заметно ускоряет процесс.

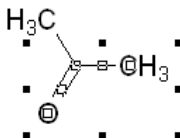
Упражнение. Изобразите двумерную структуру циклогексана и затем сгенерируйте трехмерную его структуру. Вращая структуру в трехмерном пространстве, определите ее конфигурацию: "ванна" или "кресло".



MassSpec Scissors — Расчет **молекулярной массы осколков**, которые могут образоваться в масс-спектрометрическом эксперименте.

Если выделить фрагмент молекулы и нажать эту кнопку, программа подсчитает молекулярные массы возможных осколков. Эта операция полезна при интерпретации масс-спектра.

Упражнение. Выделите фрагмент молекулы ацетона и посмотрите, как программа прогнозирует появление возможных осколков при масс-спектрометрии.



Кнопки для **перехода в базы данных** *PubChem*, *eMolecules*, *ChemSpider*. Изображенная структура используется как запрос в тех базах данных.

Упражнение. Проведите из *ChemSketch* поиск в базах данных *PubChem*, *eMolecules*, *ChemSpider* — извлеките информацию об ацетоне.

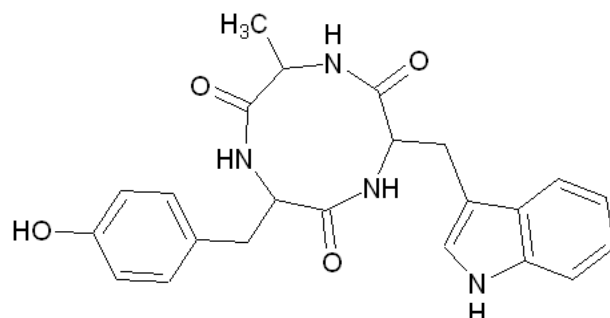


Кнопка загружает **ACD/3D Viewer** — программу работы с трехмерными структурами.

11.1.3. Режим Structure. Контрольные задания

Контрольное задание 1.

Изобразите на экране следующую структуру, предъявите преподавателю.

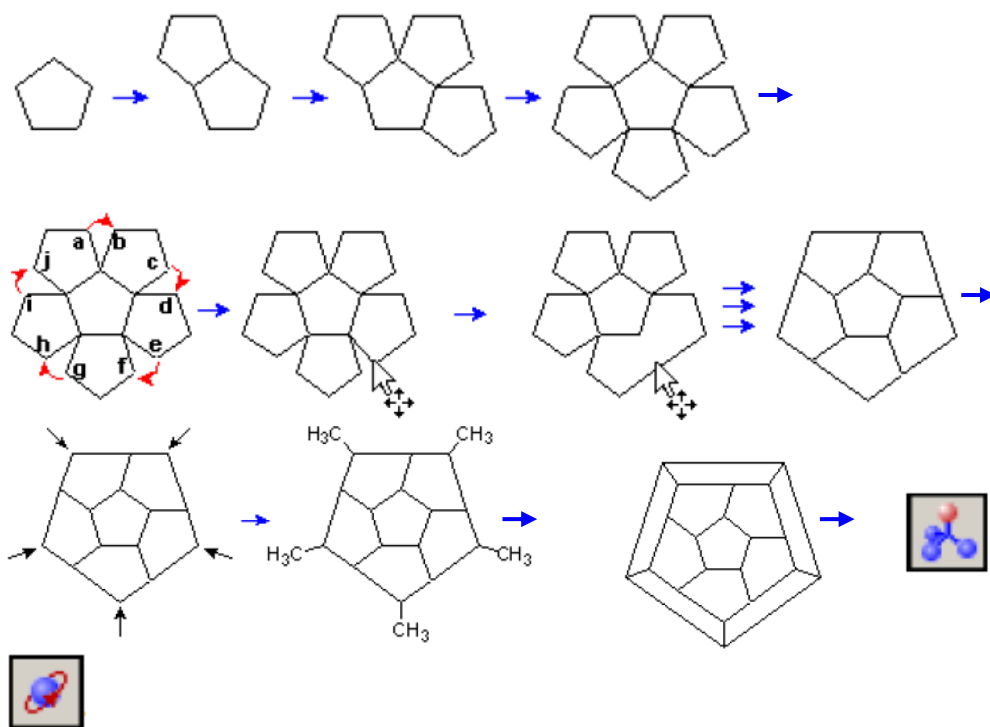


Контрольное задание 2.

Изобразите структуру [5]фуллерена- C_{20} — он имеет форму додекаэдра:

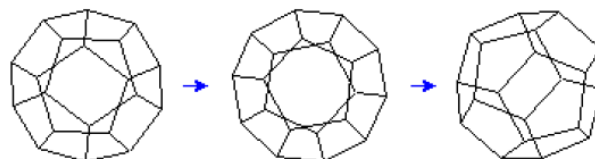


Схема работы:



Рассмотрите
разными углами:

структуру под

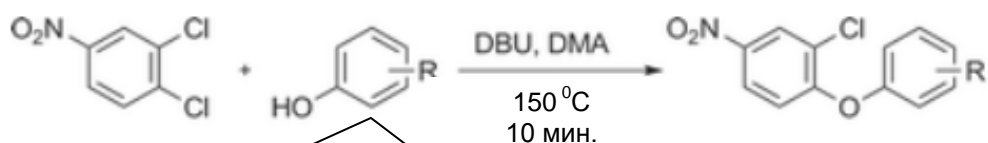


Определите расчетное значение плотности вещества.

Отобразите атомы водорода в явной форме и результат предъявите преподавателю.

Контрольное задание 3.

Изобразите на экране уравнение реакции, предъявите преподавателю.

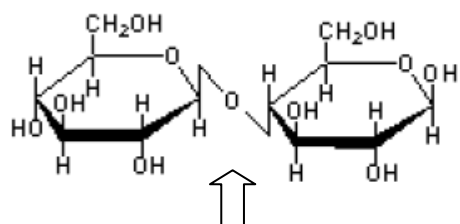


Кстати, а как радикал R
присоединен к бензольному кольцу?
Как называется такой тип формул?

Контрольное задание 4.

Изобразите на экране структуру
следующего дисахарида:

Намек: воспользуйтесь шаблонами
(*Templates -> Sugars -> ...*).



Не забудьте использовать псевдо-атомы на этом участке

Копирование информации в документ MS Word

Все, что изображено в окне *ChemSketch*, может быть стандартным способом перенесено в документ *MS Word*:

Копировать -> Вставить (*Copy -> Paste*) или *Ctrl+C -> Ctrl+V*

В результате такой операции в *DOC*-документе появляется объект программы *ChemSketch*.

При двойном щелчке по объекту открывается *ChemSketch*, объект вносится в исходную программу для редактирования.

Такой метод вставки имеет свои плюсы (возможность редактирования) и минусы (установление связи между программами иногда при некорректных настройках может привести к зависанию компьютера).

Изображенное в *ChemSketch* можно вставить в *DOC*-документ и как обычный рисунок:

Копировать -> Специальная вставка, Рисунок

Такой путь обеспечивает полную автономию *DOC*-документа, но исключает возможность эффективного редактирования химической структуры.

Контрольное задание 5.

Изобразите структурную формулу 8-оксихинолина.

Сгенерируйте его систематическое название, коды *SMILES* и *InChI*.

Сгенерируйте расчетные значения физико-химических параметров.

Все результаты перенесите в документ *MS Word*.

Найдите в литературе экспериментальное значение одного из тех параметров, который вы рассчитали теоретически (разрешается любой онлайн-источник).

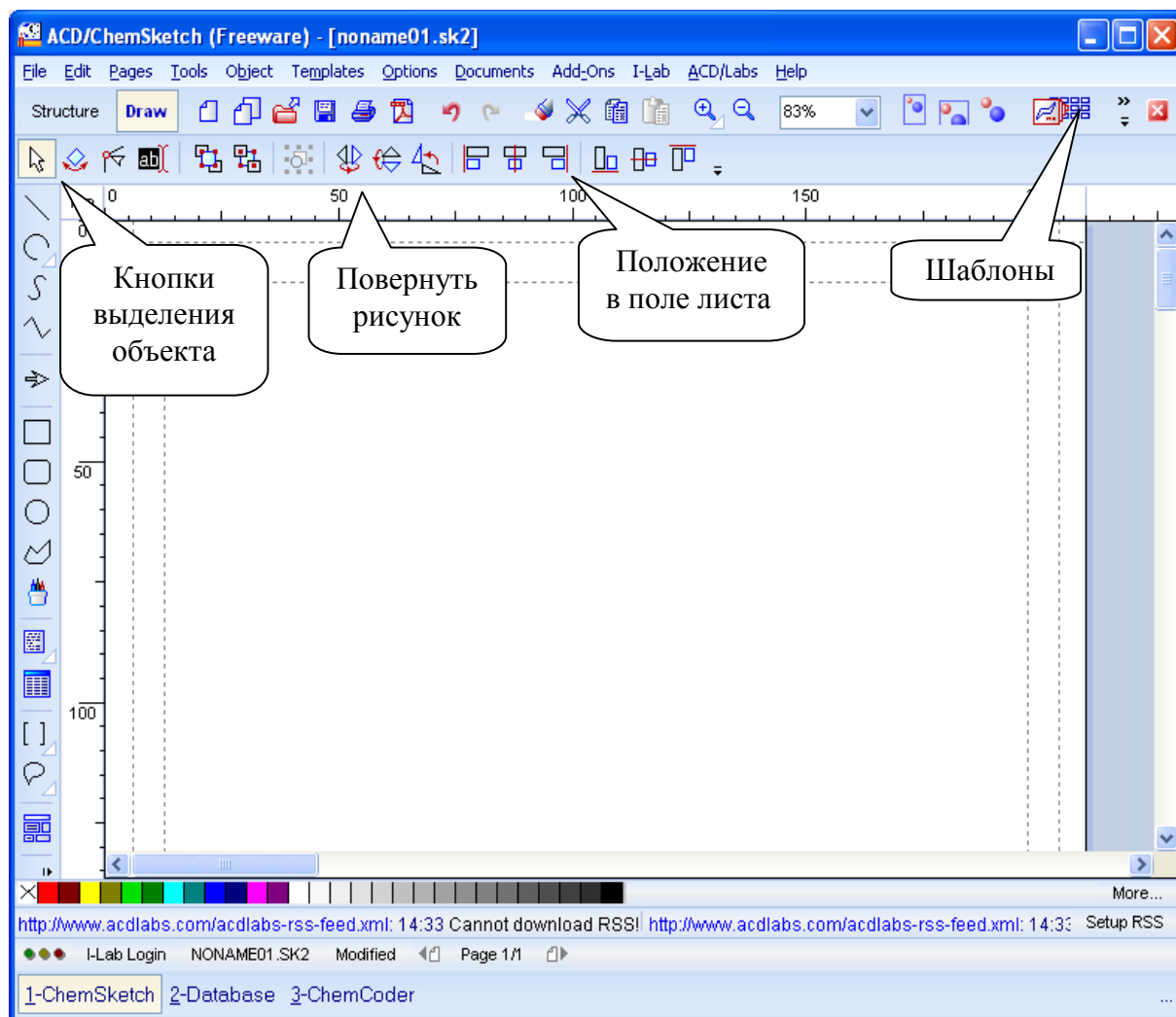
Насколько различаются экспериментальное и расчетное значения?

Результаты внесите в документ *MS Word* и предъявите преподавателю.

11.4. ACD/ChemSketch как графический редактор.

Самостоятельная работа.

Изучите приемы работы с графическим редактором (режим *Draw*), выполните упражнения и контрольные задания.



В левой колонке размещены кнопки с функциями, обычными для графических редакторов.

Упражнение.

Нарисуйте линию, дугу и другие графические элементы, для изображения которых предназначены кнопки левой колонки.

Исследуйте, какое изображение получается при нажатой клавише *Shift*.



— кнопка для рисования **ломанных линий**.

Щелкнуть в начале линии, перенести курсор и далее в точках перелома следует сделать щелчки левой клавишей мышки, в конечной точке — щелчок правой клавишей.

В горизонтальном ряду сверху находятся кнопки управления, в том числе:



Выделить, переместить, изменить размер.



Выделить, **вращать в плоскости** листа.

Выделение нескольких объектов — при нажатой клавише *Shift*.

Выделение всех объектов — *Ctrl+A*.

Исследуйте, что получается, если объект перемещать при нажатой клавише *Ctrl*.



— кнопка для **перемещения узлов** и контрольных точек кривых и ломаных линий.

Узлы — концы сегментов кривых или отрезков прямых.

При щелчке по узлу появляется отрезок касательной с расположенными на нем контрольными точками.

Перемещая контрольную точку, меняют угол наклона кривой.



Упражнение.

Нарисуйте кривую линию и изменяйте ее форму, передвигая узлы и контрольные точки.



При активации этой кнопки на панели появляются дополнительные кнопки, предназначенные для редактирования линий, имеющих узлы:



соединить прямой линией концевые узлы выделенной кривой,



удалить сегмент между выделенными узлами,



добавить сегмент между выделенными узлами,



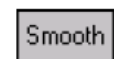
удалить выделенные узлы,



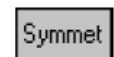
превратить выделенную **кривую или сегмент в прямую**,



превратить выделенную прямую в кривую,



сгладить линию в точке перелома,

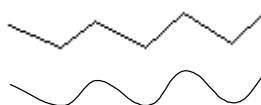


сделать два сегмента симметричными относительно узла.

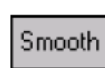
Упражнение.



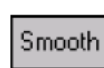
Превратите ломаную линию

в волнистую:



Понадобятся поочередно кнопки



,  ,  ,  .

Изменение цвета объекта:

- выделить объект,
- левой клавишей щелкнуть по цветному квадрату в палитре — закрашивается площадь замкнутого объекта,
- правой клавишей щелкнуть по цветному квадрату в палитре — закрашивается контур или линия.



Вносить и редактировать текст.

11.1.5. Режим Draw. Контрольные задания

Контрольное задание 6.

Изобразите мицеллу:

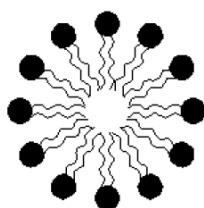
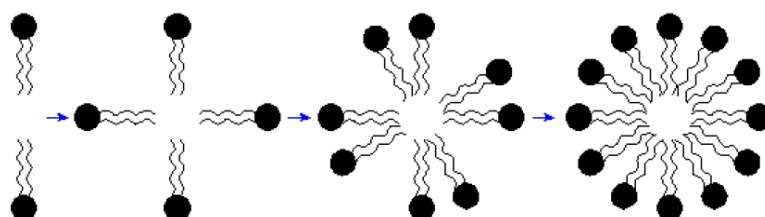
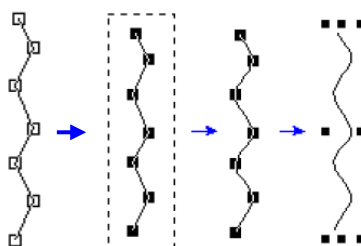
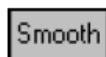
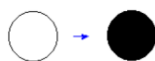


Схема работы:



Контрольное задание 7.

Изобразите строение молекулы гидразина:

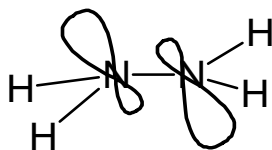


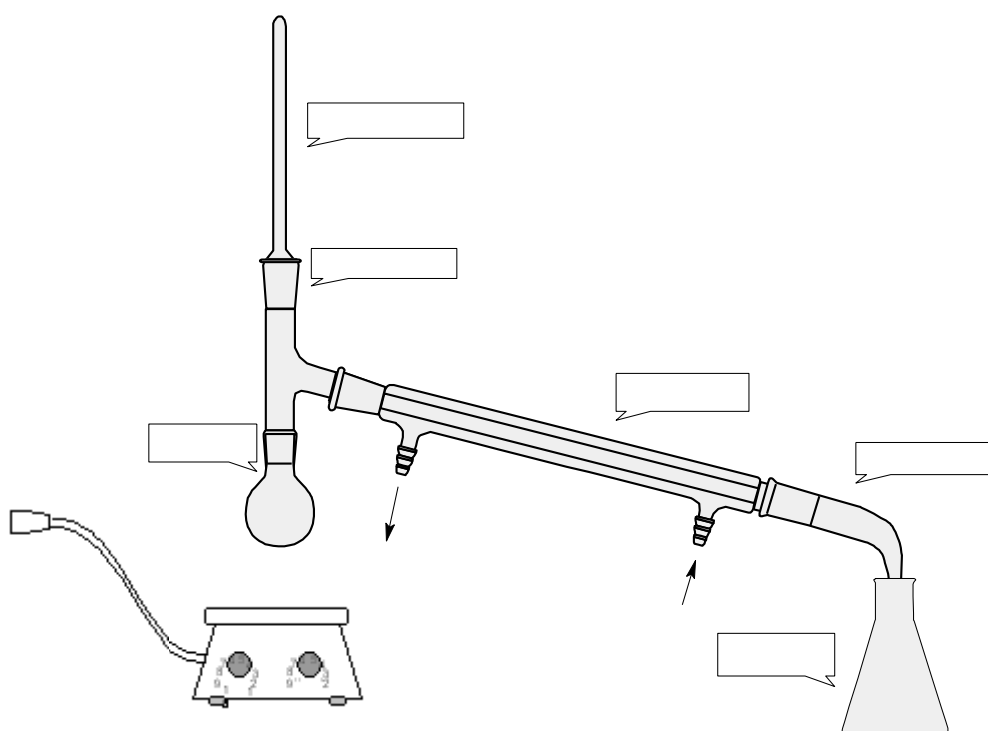
Схема работы:

Structure — набрать формулу,

Draw — дорисовать орбитали (использовать шаблон).

Контрольное задание 8.

Изобразите следующий прибор для перегонки жидкости, подпишите составные части:



11.2. ACD/3D Viewer

Программа *ACD/3D Viewer* может быть загружена двумя способами:

- Если не загружена программа *ChemSketch*:

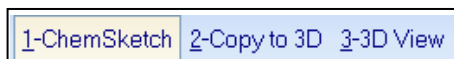
Программы -> ACDLABS 12.0 -> 3D Viewer

- Если загружена программа *ChemSketch*:

на ее верхней панели щелкнуть по значку

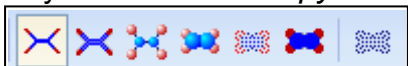


При одновременно работающих программах *ChemSketch* и *ACD/3D Viewer* в левом нижнем углу окна имеются кнопки для **переключения** между обоими окнами:



Для переноса информации из *ChemSketch* в *ACD/3D Viewer* и обратно

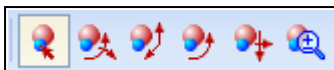
служит кнопка 2-Copy to...



Кнопки **изменения вида** структуры, в том числе, отображения Ван-дер-Ваальсовой поверхности.

Упражнение 1.

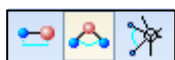
Изобразите в *ChemSketch* структурную формулу этана, проведите 3D-оптимизацию, перенесите информацию в *ACD/3D Viewer* и рассмотрите все возможные варианты отображения структуры.



Кнопки управления:
выделение, вращение, перемещение структур.

Упражнение 2.

На примере молекулы этана проверьте назначение кнопок управления.



Кнопки **измерения** межъядерных расстояний, валентных и двугранных углов.

Внимание!

Программы, о которых здесь идет речь, *моделируют* химическую структуру. Измеряемые молекулярные параметры — *расчетные*, но не экспериментальные.

Упражнение 3.

На примере молекулы этана проверьте назначение измерительных кнопок.

Конформации молекул

В ходе генерирования трехмерной структуры программа создает и выводит на экран только одну конформацию молекулы.

При наличии нескольких минимумов на кривой потенциальной энергии программа выбирает — в идеальном случае — самый глубокий минимум. Если молекула имеет несколько близких по энергии низкоэнергетических конформаций, пользователь получает одну из них, и не обязательно ту, которую он ожидал увидеть.

Остальные конформации тоже можно сгенерировать, но для этого придется немного поработать вручную.

Для создания второй конформации используем то, что:

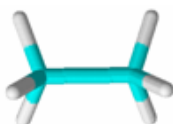
- результат моделирования может зависеть от исходных координат атомов;
- программа позволяет перемещать в трехмерном пространстве не только структуру в целом, но и ее фрагмент.

Для перемещения фрагмента структуры следует:

- выделить перемещаемые атомы;
- в меню *Edit* выбрать параметр *Manipulate Selected*;
- после чего — манипулировать избранным фрагментом.

Упражнение 4.

По умолчанию, программа *ChemSketch* из структурной формулы генерирует заслоненную конформацию этана:



Наша задача будет состоять в превращении заслоненной конформации в заторможенную:



Схема работы:

В *ChemSketch* нарисуйте структурную формулу этана, там же проведите 3D-оптимизацию, скопируйте структуру в *3D Viewer*.

На экране вы увидите модель заслоненной конфигурации этана.

Разверните исходную структуру вокруг оси *y* на 90° (так, чтобы связь C–C оказалось перпендикулярной плоскости экрана).

Выделите три ближайшие к вам атома водорода (внимание! атом C выделять не надо).

В меню *Edit* выберите параметр *Manipulate Selected*.

Поверните в плоскости *xy* выделенные атомы водорода на 60° .

Вращая модель в трехмерном пространстве, убедитесь, что задача выполнена верно.

Примечание.

Предыдущее упражнение мы выполняли в ознакомительных целях.

Если проводить 3D-оптимизацию заслоненной конформации этана в окошке *3D Viewer*, программа сама автоматически генерирует энергетически более выгодную заторможенную конфигурацию.

Контрольное задание 9.

Создайте трехмерную модель молекулы оксида фосфора(V).

Результат предъявите преподавателю.

11.3. Файлы в формате MOL

Формат *MOL* де-факто является стандартным форматом обмена структурной информацией.

Самостоятельная работа 1.

В *ChemSketch* изобразите двумерную структурную формулу уксусной кислоты.

Сохраните структуру в *MOL*-формате (*MDL molfile v. 2.0*).

Откройте сохраненный файл в текстовом редакторе.

Будьте готовы объяснить преподавателю смысл чисел, хранящихся в файле.

Самостоятельная работа 2.

Проведите 3D-оптимизацию изображенной ранее структурной формулы уксусной кислоты.

Сохраните трехмерную структуру в *MOL*-формате (*MDL molfile v. 2.0*).

Откройте сохраненный файл в текстовом редакторе.

Будьте готовы объяснить преподавателю смысл чисел, хранящихся в файле.

12. ПРИЛОЖЕНИЕ. ИССЛЕДОВАТЕЛЬСКАЯ КОНТРОЛЬНАЯ РАБОТА

"Применимость моделей ВС и VSEPR для прогнозирования и объяснения структуры вещества в кристалле".

При качественном прогнозировании структуры изолированной молекулы обычно используют одну из двух моделей:

- **метод валентных связей (ВС)** в его исходной простейшей форме;
- **метод расталкивания валентных электронных пар (VSEPR)**.

Обе модели в **первом приближении** дают идентичные или близкие результаты.

Напоминаем суть обеих моделей.

ВС

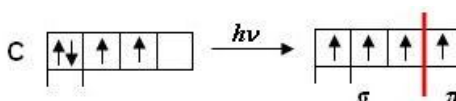
- ковалентная связь образуется как правило по обменному механизму;
- число связей, образуемых атомом, равно числу неспаренных электронов; при необходимости проводят распаривание электронных пар;
- для объяснения пространственного расположения атомов привлекают концепцию гибридизации атомных орбиталей.

VSEPR

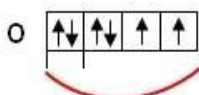
- каждый атом стремится иметь 8-электронную валентную оболочку, т.е. 4 электронные пары (водород — двухэлектронную оболочку, т.е. 1 электронную пару);
- несвязывающие и σ -связывающие электронные пары располагаются в пространстве таким образом, чтобы расталкивание между ними было минимальным.

Пример 1: H_2CO_3

Метод ВС:

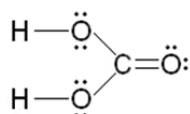


атом С находится в состоянии sp^2 -гибридизации,



два атома О находятся в состоянии sp^3 -гибридизации.

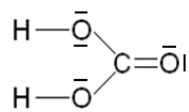
Результат прогнозирования:



$\angle \text{OCO} \approx 120^\circ$
 $\angle \text{HOC} \approx 109^\circ$
 $d(\text{C}=\text{O}) < d(\text{C}-\text{O})$
 $E(\text{C}=\text{O}) > E(\text{C}-\text{O})$

Метод VSEPR:

8-электронные оболочки атомов С, О и двухэлектронная оболочка атома Н получаются при таком распределении электронных пар:

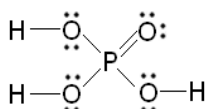


(на схеме каждая пара электронов отображена черточкой)

В примере 1 методы ВС и VSEPR дают **идентичные** результаты.

Пример 2: H_3PO_4

Метод ВС:



Атом фосфора образует три одинарные и одну двойную связи.

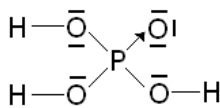
$\angle \text{OPO} \approx 109^\circ$ (sp^3 -гибридизация атома P)

$\angle \text{HOP} \approx 109^\circ$ (sp^3 -гибридизация атома O)

$d(\text{P}=\text{O}) < d(\text{P}-\text{O})$

$E(\text{P}=\text{O}) > E(\text{P}-\text{O})$

Метод VSEPR:



Атом фосфора образует четыре одинарные связи.

Валентные углы такие же, какие следуют из модели ВС:

$\angle \text{OPO} \approx 109^\circ$

$\angle \text{HOP} \approx 109^\circ$

Отличие: **Все связи P–O одинарные, значит, они близки и по энергии, и по длине.**

Результаты прогнозирования по двум методам в этом случае частично совпадают и частично различаются.

Надстройки над простейшими моделями.

а) Почему угол связи Э–О–Р может быть **больше** 109°

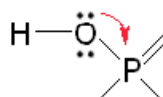
Если один из атомов имеет неподеленную электронную пару, а его сосед — вакантную валентную орбиталь, возникает дополнительное π -взаимодействие:



Энергия связи О–Р (и О=P) увеличивается, длина связи О–Р (и О=P) уменьшается.

О степени дополнительного π -взаимодействия косвенно можно судить по состоянию атома кислорода в цепочке Э–О–Р.

Для чистоты рассуждений предпочтительно, чтобы атом Э не обладал акцепторными свойствами. Будем анализировать вариант Н–О–Р.

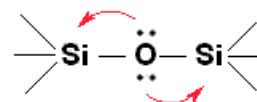


Чем в большей степени неподеленная электронная пара атома кислорода участвует в π -взаимодействии, тем меньше она участвует в гибридизации — следовательно, угол связи Н–О–Р увеличивается от sp^3 -гибридного (109°) в направлении sp^2 -гибридного (120°).

Если обе неподеленные электронные пары атома кислорода полностью принимают участие в π -взаимодействии, атом кислорода переходит в состояние sp -гибридизации (угол 180°).

Экспериментальный факт:

в силикатах величина угла Si–O–Si
в пределах 130° – 180° — обычное явление:



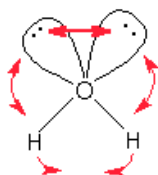
б) Почему угол связи Э–О–Р может быть **меньше** 109°

Объяснение с позиций модели VSEPR.

Электронные пары расталкиваются между собой, и сила расталкивания уменьшается в ряду:

несвязывающая–несвязывающая > несвяз.–связ. > связ.–связ.

Несвязывающие пары "сжимают" связывающие пары, например:



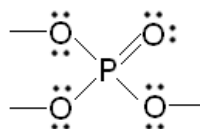
В результате угол Э–О–Э становится меньше 109° .

Умозрительный эксперимент.

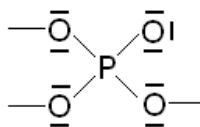
Некое вещество содержит PO_4 -фрагмент.

Предположим, мы нашли экспериментальные данные о структуре этого вещества. Проанализируем, как эти данные могут быть интерпретированы с позиций двух моделей: ВС и VSEPR.

а) Если три связи фосфор–кислород имеют одинаковую длину, а четвертая заметно короче, структура PO_4 -группировки корректно отображается простейшей моделью по методу ВС:



Если все четыре связи фосфор–кислород имеют примерно одинаковую длину, структура PO_4 -группировки корректно отображается простейшей моделью по методу VSEPR:



б) Если угол связи Э–О–Р близок к 109° , связь О–Р строго одинарна.

Если угол связи Э–О–Р больше 109° , порядок связи О–Р может быть больше единицы (он без сомнений больше единицы, если атом Э не обладает акцепторными свойствами).

Угол связи Э–О–Р, меньший 109° , объясняем с позиций расталкивания электронных пар.

в) Если экспериментальные данные не укладываются в эту простейшую схему, следует проанализировать, каким образом окружение PO_4 -группировки может влиять на ее структуру, а также то, как в самой PO_4 -группировке могут накладываться друг на друга разные тенденции.

Переходим к онлайн-эксперименту.

Будем решать следующую задачу:

Проанализировать, в какой степени каждая из моделей — ВС и VSEPR — пригодна для описания структуры вещества, находящегося в кристаллическом состоянии.

Объекты исследования — соединения, содержащие в своем составе группировку PO_4 .

Источники информации — кристаллографические базы данных.

Схема исследования:

1. В кристаллографических базах данных найдите пять веществ, содержащих группировку PO_4 .

Учтите, что наличие водородной связи или ионного взаимодействия усложнит дальнейший анализ.

Поэтому извлекайте вещества явно ковалентного строения, да такие, чтобы в цепочке Э–О–Р атом Э не был акцептором электронной пары атома кислорода (идеальный вариант — это атом С, несколько хуже — Н).

2. Извлеките и сохраните на своем диске CIF-файлы этих веществ.
3. Определите геометрические параметры группировок PO_4 в каждом из этих веществ:
 - длины связей Р–О;
 - углы О–Р–О;
 - углы Э–О–Р.
4. Проанализируйте численные значения величин, полученных в п. 3. Обратите внимание на следующие аспекты:
 - различаются ли длины связей Р–О в пределах одной структуры;
 - если различаются, то можно ли связи Р–О разделить на две группы: "условно одинарные" и "условно двойные";
 - имеется ли сходство в длинах связей Р–О для разных веществ;
 - можно ли с той или иной степенью точности все наблюдаемые длины связей Р–О разделить на группы и соотнести их с одинарными, двойными связями;
 - есть ли среди ваших структур такая, которая однозначно (или почти однозначно) соответствует модели ВС;
 - есть ли среди ваших структур такая, которая однозначно (или почти однозначно) соответствует модели VSEPR;
 - в какой степени модели ВС и VSEPR были бы способны предсказать строение PO_4 –группировок в данном комплекте веществ;
 - как с точки зрения метода гибридизации атомных орбиталей можно было бы охарактеризовать состояние атомов кислорода в цепочке Э–О–Р.

В ходе анализа учтите, что, согласно методу ВС, в ионах PO_4^{3-} , HPO_4^{2-} , H_2PO_4^- есть связи О–Р, у которых порядок связи больше единицы и меньше двух (кстати, какие именно и почему?).

Письменный отчет сдайте преподавателю.

Формат письменного отчета

	Название базы данных	Брутто- формула вещества	2D-структура PO_4 -группировки и ее ближайшего окружения	Геометрические параметры фрагмента структуры	Выводы (в произвольной форме, на основе п.4)

ЛИТЕРАТУРА

1. Leach, A.R. An Introduction to Chemoinformatics / A.R. Leach, V.J. Gillet. – Springer, 2007.
2. Chemoinformatics : A Textbook / Edited by Johann Gasteiger and Thomas Engel. – Wiley-VCH, 2003.
3. Chemoinformatics: Theory, Practice, & Products / B.A. Bunin [et al.]. – Springer, 2007.
4. Потапов, В.М. Химическая информация. Где и как искать химику нужные сведения / В.М. Потапов, Э.К. Кочетова. – Москва : Химия, 1988.
5. Рагойша, А.А. Поиск химической информации в Интернете. Поисковые системы и тематические каталоги : учеб. пособие для студ. хим. фак. / А.А. Рагойша. – Минск : БГУ, 2003.
6. Рагойша, А.А. Поиск химической информации в Интернете : научные публикации : учеб. пособие для студ. хим. фак. спец. 1-31 05 01 / А.А. Рагойша. – Минск : БГУ, 2007.
7. Рагойша, А.А. Текстовый поиск научной химической информации в Интернете [Электронный ресурс] : (практикум для студентов) / А.А. Рагойша. – Минск : БГУ, 2012. — 64 с. — Рус. — Деп. в ГУ «БелИСА» 21.03.2012 г., № Д201211. <http://elib.bsu.by/handle/123456789/14599> .

BIBLIOGRAPHIC RECORD

Type: Book (Electronic)

Title: Retrieval of Chemical Information from Online Structural Databases : Manual

Author: Rahoisha, Aliaksandr A.

Year: 2013

Description: 101 p. : ill.

Abstract: The manual guides students in designing queries for the effective retrieval of chemical information from structural, crystallographic, spectral, and reaction databases.

Subject: Chemistry

Institution: Belarusian State University

Country: Belarus

Language: Russian

Address for correspondence: Aliaksandr Rahoisha, Chemistry Department, BSU, pr. Nezalezhnasci 4, Minsk 220030, Belarus.

E-mail: arahoisha@gmail.com

Индекс УДК: 54:004.65(076.5)(075.8)

Основное заглавие: Поиск информации о структуре химического вещества в онлайн-базах данных [Электронный ресурс]

Сведения, относящиеся к заглавию: Практикум по курсу "Информационные технологии в химии" для студентов спец. 1-31 05 01 Химия (по направлениям)

Автор: Рагойша, Александр Антонович

Год: 2013

Количество страниц: 101 с.

Реферат: Пособие предназначено для обучения студентов-химиков начальным навыкам работы со структурными, кристаллографическими, спектральными и реакционными базами данных. Рассматриваются правила формулирования структурного запроса, алгоритмы извлечения информации и ее анализа. Каждый раздел пособия содержит описательную обучающую часть, рекомендации по ведению самостоятельной работы, проверочные вопросы и задания, предназначенные для промежуточного контроля.

Место издания: Минск : Белорусский государственный университет

Адрес для переписки: А. А. Рагойша, химический факультет, БГУ, пр. Независимости 4, Минск 220030, Беларусь.

E-mail: arahoisha@gmail.com