

Министерство образования и науки Российской Федерации  
Ярославский государственный университет им. П. Г. Демидова  
Кафедра органической и биологической химии

**КОМПЬЮТЕРНОЕ  
МОДЕЛИРОВАНИЕ  
МАКРОМОЛЕКУЛЯРНЫХ ОБЪЕКТОВ  
И БИОЛОГИЧЕСКИХ СИСТЕМ**

*Методические указания*

*Рекомендовано  
Научно-методическим советом университета для студентов,  
обучающихся по направлению Биология*

Ярославль  
ЯрГУ  
2013

УДК 577.2:004.9(072)  
ББК 3973.2я73+Е070я73  
К 63

*Рекомендовано  
Редакционно-издательским советом университета  
в качестве учебного издания. План 2013 года.*

Рецензент  
кафедра органической и биологической химии ЯрГУ

Составители:  
Д. А. Базлов, А. В. Цивов

**Компьютерное моделирование макромолекулярных объектов и биологических систем:** метод. указания / сост. Д. А. Базлов, А. В. Цивов; Яросл. гос. ун-т им. П. Г. Демидова. — Ярославль : ЯрГУ, 2013. — 48 с.

Методические указания знакомят студентов-биологов с основами такого направления, как биоинформатика, и вычислительными методами для обработки экспериментальных данных, моделирования биологических и биохимических объектов.

Предназначены для студентов, обучающихся по направлению 020400.62 Биология (дисциплины «Молекулярное моделирование механизмов жизнедеятельности»; «Молекулярная биология и компьютерное моделирование биосистем», цикл Б2), очной формы обучения.

УДК 577.2:004.9(072)  
ББК 3973.2я73+Е070я73

© ЯрГУ, 2013

# 1. Биоинформатика

## 1.1. Развитие биоинформатики

Биоинформатика — междисциплинарный предмет. Это наука об использовании информатики в изучении биологии. В биоинформатике биология, информатика и математика сливаются в единую дисциплину. Строго говоря, биоинформатика расширяет предметную область вычислительной биологии, изучающей применение методов количественного анализа в моделировании биологических систем. Таким образом, биоинформатика изучает применение информационных технологий для управления биологическими данными.

На практике иногда это понятие более узкое, под ним понимают использование компьютеров для обработки экспериментальных данных по структуре биологических макромолекул (белков и нуклеиновых кислот) с целью получения биологически значимой информации. В свете изменения шифра научных специальностей (03.00.28 «Биоинформатика» превратилась в 03.01.09 «Математическая биология, биоинформатика») поле термина «биоинформатика» расширилось и включает все реализации математических алгоритмов, связанных с биологическими объектами.

Так когда же появилась биоинформатика, которую сейчас с уверенностью называют наукой третьего тысячелетия? Можно считать, что эта наука зародилась в XIII веке. Тогда молодой итальянец Леонардо из Пизы, вошедший в историю математики под именем Фибоначчи, описал решение задачи о размножении кроликов и таким образом построил первую математическую модель биологического процесса. По сути, этот любопытный случай и положил начало математической биологии. В 20-е годы XX в. другой итальянский математик, Вито Вольтерра, создал модель совместного существования двух биологических популяций типа «хищник — жертва». У математиков появился интерес к теоретической биологии, а у биологов, в свою очередь, возникла потребность к систематизации науки с математических позиций. После Второй мировой войны, с выходом в свет книги Э. Шредингера «Что такое жизнь с точки зрения физики?», в биологию пришли физики и математики. Среди них были такие знаменитые ученые, как Алексей Андреевич Ляпунов и Игорь Андреевич Полетаев. Благодаря их работам

начал развиваться кибернетический подход к биологическим процессам. Вадим Александрович Ратнер, ученик Ляпунова, в начале 1960-х гг. применил математический подход к описанию явлений молекулярного уровня в сложных системах. В 1975 г. была создана первая известная количественная модель управления развитием фага «лямбда». Его геном содержит 48 генов (совсем немного по нынешним меркам), но имеет сложную управляющую систему. Эта и более сложные системы уже требовали использования компьютеров. Из математической биологии стала выделяться отдельная ветвь — биоинформатика, которая предполагает применение информационных технологий для изучения биологических систем. Следует отметить большую роль Института цитологии и генетики Сибирского отделения РАН в развитии биоинформатики. Это область со своей проблематикой и методикой, которые немислимы без компьютеров. Накапливается много информации о первичных последовательностях геномов молекул ДНК.

Биоинформатика — это область науки, разрабатывающая и применяющая вычислительные алгоритмы для систематизации и анализа генетической информации с целью определения молекулярных основ биологических процессов с последующим использованием этих знаний на практике. Ее основная задача — разработка вычислительных алгоритмов для анализа и систематизации данных о структуре и функциях биологических молекул, прежде всего нуклеиновых кислот и белков. Объем генетической информации, накапливаемой в банках данных, начал увеличиваться с возрастающей скоростью после того, как были разработаны быстрые методы секвенирования (расшифровки нуклеотидных последовательностей ДНК).

Биоинформатические методы позволяют не просто обрабатывать этот огромный массив данных, но и выявлять закономерности, которые не всегда можно заметить при обычном эксперименте, предсказывать функции генов и зашифрованных в них белков, строить модели взаимодействия генов в клетке, конструировать лекарства.

Биоинформатика как наука появилась на стыке молекулярной биологии, генетики, математики и компьютерных технологий и обязана своим появлением накоплению обширных экспе-

риментальных данных, особенно заметно это накопление стало проходить начиная с 70-х гг. прошлого века. Главной проблемой, потребовавшей нового подхода, стал анализ последовательностей генов и белков, которые стало возможно «прочитать» в эксперименте. Появление данных о пространственной структуре макромолекул также требовало новых методов работы с ними. Информации, получаемой в эксперименте, было больше, чем возможностей человека к запоминанию фактов и их анализу.

Возникла также необходимость хранения все быстрее увеличивающегося объема информации. Первые несколько сотен расшифрованных последовательностей белков были опубликованы в виде книги-атласа. Отметим, что первая белковая последовательность (инсулин) была расшифрована в 1955 г., а первый «Атлас белковых последовательностей» вышел в 1965 г., т. е. через десять лет. Однако уже в начале 1970-х гг. число расшифрованных последовательностей возросло настолько, что их стало невозможно публиковать в книгах, более того, стало понятно, что нужны специальные программы для сравнения последовательностей, поскольку мозг человека не справляется с анализом такой информации. Если посмотреть на динамику роста информации в крупнейшем банке данных генных последовательностей EMBL Nucleotide Databank (<http://www.ebi.ac.uk/embl/>) за время с 1982 по 2005 г., то с начала XXI в. **наблюдается экспоненциальный рост количества информации.** К концу 2005 г. в банке находилось примерно 115 000 000 000 букв в 64 000 000 последовательностях.

На помощь пришли компьютерные технологии. Первый алгоритм выравнивания последовательностей был предложен ещё в 1970 г. Компьютеры позволили хранить информацию в виртуальных банках данных и оперировать ею уже с другой быстротой. Достижения информатики, лингвистики и теории информации сделали возможным анализ генетических текстов. Напомним, что первый персональный компьютер появился в 1981 г., а мировой Интернет — в 1991 г. Появление Интернета имело колоссальное значение для развития не только предмета науки, но и её организации. Один из главных принципов биоинформатики — принцип единого мирового информационного пространства, объединяющего усилия учёных, работающих по всему земному шару.

Как уже упоминалось, в 1990-е гг. происходит расцвет геномики, обозначивший начало нового этапа в биологии. В настоящее время расшифрованы полные геномные последовательности человека, мыши, цыпленка, лягушки, рыбы, плодовой мухи, круглого червя, нескольких сотен вирусов и бактерий. Прочтение генома бактерии теперь задача, посильная группе из 2–3 исследователей за время, меньшее чем 1 год. Геном человека составляет около 3 млрд букв (что эквивалентно 15 000 книжных томов). Значение факта его «прочтения» для биологов сравнимо с открытием Д. И. Менделеевым периодического закона для химиков. Последовательность находится в открытом доступе в Интернете, с её изучением связаны большие надежды в области практической медицины.

## 1.2. *In silico* или *in vivo*?

Для анализа огромного массива данных требуются высокопроизводительные вычисления, и обычный компьютер не способен справиться с такой задачей — нехватит объема памяти и времени. Какие компьютерные технологии предлагают сегодня биоинформатикам? Сначала в США появилась высокопроизводительная система «Cray» — громадная установка, которая занимает объем комнаты, но если собрать вместе обычные компьютеры, чтобы получить такую же мощность, получится пятиэтажный дом. Сегодня, к примеру, все больше используется суперкомпьютер «Blue» (именно он обыграл Гарри Каспарова в шахматном поединке). Этот компьютер способен проанализировать громадное количество комбинаций за довольно короткий период времени (от нескольких минут до нескольких дней, в зависимости от сложности поставленной задачи). Не менее важны так называемые компьютерные кластеры, то есть несколько последовательно соединенных обычных компьютеров. Задача разделяется на части, работа с которыми идет параллельно. Подобная технология найдет широкое применение в будущем, так как параллельная обработка данных наиболее эффективна. В последнее время возник такой метод, как распределенные вычисления, когда в расчетах участвует множество компьютеров в разных странах и каждый компьютер выполняет одну небольшую часть задачи. Суперкомпьютеры необходимой мощности существуют и в России, например в Пущине. При наличии оптоволоконных се-

тей к нему может подключиться лаборатория, которая находится даже в другом городе.

Сравнительно недавно в науке появился термин «*биология in silico*», буквальный смысл которого — «биология на кремнии», иными словами, проведение биологического эксперимента на компьютере. Сейчас это понятие стало вполне официальным и широко используется. Есть журнал, который так и называется — «*In silico biology*».

Биоинформатика дает возможность быстро и дешево провести модельный эксперимент и зачастую предсказать результат эксперимента. Но модели компьютерной жизни все же пока далеки от реальности. Во всем мире ведутся активные исследования для максимального приближения *in silico* к *in vivo*. Это означает не только привлечение к расчетам данных о структуре и функции исследуемого белка, но и учет взаимодействия между белками в реальной среде. В проект виртуальной клетки сейчас вкладываются огромные средства. В России, например, такие работы ведутся в Пушинском научном центре. Создание «электронной клетки» даст возможность эффективно конструировать лекарства нового поколения. (Афанасьева Г. Биоинформатика: виртуальный эксперимент в шаге от реальности. Наука и жизнь. 2004. № 11. URL: <http://www.nkj.ru/archive/articles/310/>)

### 1.3. Цели и задачи биоинформатики

В настоящее время задачи биоинформатики — это почти исключительно задачи молекулярной биологии. Причина этого в том, что за последние 20–25 лет накоплен поистине колоссальный экспериментальный материал именно о строении и функционировании биологических молекул (белков и нуклеиновых кислот), в качестве примера достаточно привести геном человека. Этот материал требует развитых компьютерных методов для своего анализа.

Основными разделами биоинформатики являются компьютерная геномика, решающая проблемы расшифровки генетических «текстов», хранящихся в последовательностях нуклеотидов ДНК (РНК), и метабономика, исследующая организацию метаболизма клетки и его управления со стороны генома. Важное значение для развития основных разделов биоинформатики имеет

создание компьютерных баз данных по молекулярной биологии, обеспечивающих геномику и метабомику необходимыми для их развития экспериментальными данными в достаточном количестве и удобной для использования форме.

Среди основных задач биоинформатики — описание генных сетей, изобретение новых лекарств с заданными свойствами, разработка компьютерных моделей процессов, происходящих в организме.

К биоинформатике часто относят также информационные службы, обеспечивающую накопление, хранение и использование приобретаемых наукой знаний о биологических системах.

Целью биоинформатики является как накопление биологических знаний в форме, обеспечивающей их наиболее эффективное использование, так и построение и анализ математических моделей биологических систем и их элементов. Информация о строении материальных элементов, обеспечивающих функционирование организма, хранится в последовательности нуклеотидов ДНК (или РНК), образующей его геном. Установление нуклеотидных последовательностей ДНК геномов организмов (секвенирование) стало к началу XXI в. хорошо освоенной и, видимо, достаточно рентабельной технологией. Количество секвенированных геномов быстро увеличивается и определяется в основном только объемом средств, которые можно затратить на эти цели.

По сути биоинформатика включает в себя три компонента: создание баз данных, позволяющих осуществлять хранение крупных наборов биологически данных и управлять ими; разработку алгоритмов и методов статистического анализа для определения отношений между элементами крупных наборов данных; использование этих средств для анализа и интерпретации биологических данных различного типа — в частности, последовательностей ДНК, РНК и белков, белковых структур, профилей экспрессии генов, биохимических путей.

Цели биоинформатики можно свести к следующим:

1. Организовать данные таким образом, чтобы исследователи имели доступ к текущей информации, хранящейся в базах данных, и смогли вносить в нее новые записи по мере получения новых сведений.



2. Развивать программные средства и информационные ресурсы, которые помогают в управлении данными и в их анализе.

3. Применять эти средства для анализа данных и интерпретации полученных результатов таким образом, чтобы они имели биологический смысл.

В целом задачи биоинформатики состоят в анализе информации, закодированной в биологических последовательностях, что предполагает:

- обнаружение генов в последовательности ДНК различных организмов;

- развитие методов изучения структуры и (или) функции новых расшифрованных последовательностей и соответствующих структурных областей ДНК;

- определение семейств родственных последовательностей и построение моделей;

- выравнивание последовательностей и восстановление филогенетических деревьев с целью выявления эволюционных связей;

- обнаружение мишеней для медикаментозного воздействия и отыскивание перспективных опытных соединений.

#### *1.4. Основные направления биоинформатики*

Есть несколько основных направлений этого раздела науки, в зависимости от исследуемых объектов:

- биоинформатика последовательностей;

- структурная биоинформатика;

- компьютерная геномика.

#### *1.5. Биоинформатика последовательностей*

Этот раздел биоинформатики занимается анализом нуклеотидных и белковых последовательностей. В настоящее время разработаны эффективные экспериментальные методы определения нуклеотидных последовательностей. Определение нуклеотидных последовательностей стало рутинной, хорошо автоматизированной процедурой. В результате уже получено огромное количество генетических текстов. Так, в базе данных EMBL на 15.02.2007 г. хранится 87 000 493 документов с описанием нуклеотидных по-

следовательностей, содержащих в целом 157 545 686 001 символов (нуклеотидов), что соответствует примерно библиотеке в 105 толстых томов с убористым шрифтом. Найти нужный ген в EMBL — это все равно что найти цитату в такой библиотеке. Без помощи компьютера сделать это, мягко говоря, очень трудно. А число данных экспоненциально растет.

Представим себе геном небольшой бактерии — это непрерывная строка длиной в 1–10 миллионов символов, и далеко не вся ДНК кодирует белки. Первый тип биоинформатической задачи — это задачи поиска в нуклеотидных последовательностях особых участков, кодирующих белки, участков, кодирующих РНК (например, тРНК), участков связывания с регуляторными белками и др. И это не всегда простые задачи, например, гены эукариотических организмов состоят из чередующихся «осмысленных» и «бессмысленных» фрагментов (экзонов и интронов), и расстояние между «осмысленными» фрагментами может достигать тысяч нуклеотидов.

Пусть ген найден. Что он кодирует? Зачем он нужен? Если речь идет об участке ДНК, кодирующем белок, то с помощью весьма простой операции — трансляции с использованием известного генетического кода — можно получить аминокислотные (белковые) последовательности. Из известных на сегодня 4 273 512 белков около 94 % последовательностей — это именно такие гипотетические трансляты, и больше о них ничего не известно. Скорость поступления информации с автоматических секвенаторов превышает скорость нашего понимания ее смысла!

Но биологические объекты — это объекты, возникшие в процессе эволюции. Сравнительно-эволюционный подход — один из мощнейших подходов в биологии. Например, функция белка из одного организма хорошо экспериментально изучена, в другом организме нашли белок с похожей аминокислотной последовательностью. Можно предположить, что второй (неизвестный) белок выполняет ту же или схожую функцию. И здесь сразу возникает несколько вопросов. Во-первых, что значит похожая последовательность? Как сравнивать последовательности? При какой степени сходства последовательностей можно предполагать, что белки выполняют сходные функции? Сравнение последовательностей (выравнивание) является важнейшей задачей биоинформатики.

Можно привести много примеров того, как сравнительно-эволюционный подход в сочетании с биоинформатическими методами порождает новое биологическое знание.

Генетические тексты — тексты с большой долей шума, сравнивая родственные последовательности, в ряде случаев удается отфильтровать шум и выявить сигнал, например короткую последовательность нуклеотидов, способную связываться с белком-регулятором, или аминокислотные остатки в ферменте, отвечающие за связывание субстрата. Чтобы быть уверенными в результате, биоинформатики используют теорию вероятности и математическую статистику.

Подводя итог, можно сказать, что основные задачи биоинформатики, связанные с анализом отдельных последовательностей, состоят в следующем:

- выравнивание и определение сходства двух последовательностей;
- построение множественных выравниваний;
- распознавание генов;
- предсказание сайтов связывания регуляторных белков;
- предсказание вторичной структуры РНК.

Создание новых экспериментальных технологий ставит перед биоинформатикой целый ряд новых задач. Например, развитие масс-спектрометрии позволяет (пока в принципе) в одном эксперименте проанализировать весь набор белков, присутствующий в клетке. Для решения этой задачи необходим совместный анализ спектров масс и геномов. Открытие новых биологических явлений и механизмов также приводит к появлению новых задач. Хорошим примером служит открытие РНК интерференции, за которую в 2006 г. дали Нобелевскую премию по физиологии. Это открытие породило целый вал биоинформатических работ, посвященных поиску участков связывания микроРНК и новых микроРНК. Многие находки были затем подтверждены экспериментально.

### *1.6 Структурная биоинформатика*

Каждый белок, помимо своей уникальной последовательности аминокислот, из цепочки которых состоит его молекула, обладает ещё и уникальным способом укладки этой цепочки в пространстве.

Задачу предсказания укладки по последовательности можно, в принципе, тоже считать задачей биоинформатики, но это задача в своём общем виде ещё слишком далека от решения. Поэтому структурная биоинформатика занимается анализом пространственных структур, уже определённых экспериментально. Структур белков известно намного меньше, чем последовательностей белков. Это связано с тем, что экспериментальные процедуры для определения структуры намного сложнее, дороже, и к тому же (в отличие от секвенирования) не являются «рутинными», то есть их результат вовсе не гарантирован. Тем не менее к началу 2007 г. для анализа уже были доступны более 30 000 структур, что тоже немало (доступных белковых последовательностей — несколько миллионов). Среди них как структуры отдельных белковых молекул, так и структуры комплексов белков с ДНК, РНК, другими химическими веществами. Например, большинство лекарств представляют собой химические вещества, чьи молекулы способны связываться — образовывать комплексы — с молекулами тех или иных белков (как правило, в результате такого связывания белок оказывается неспособен выполнять свою природную функцию, что и обеспечивает эффект лекарства). Исследование механизма действия лекарств имеет большое практическое значение, поэтому определением структуры комплексов молекул белков с молекулами лекарств занимаются многие экспериментальные группы. Как результат — большое количество доступных для компьютерного анализа структур комплексов.

Примеры задач структурной биоинформатики:

- определение участков белковой молекулы, важных для той или иной функции данного белка (в биоинформатике часто вместо «определение» говорят «предсказание», поскольку компьютерный анализ не может иметь результатом научный факт, а лишь более или менее достоверное предсказание, которое должно быть затем проверено экспериментами);

- сравнительный анализ структур родственных белков, классификация белков на основе их пространственной структуры;

- анализ структур комплексов двух или нескольких молекул белка, комплексов молекул белка с другими молекулами; предсказание воздействия молекул химических веществ (в частности, потенциальных лекарств) на молекулы белков;

- предсказание структуры белка по структуре белка с похожей последовательностью (в такой ситуации задача предсказания укладки часто разрешима).

### *1.7. Компьютерная геномика*

В настоящее время определены полные или почти полные последовательности геномов многих организмов. Прочтение полной нуклеотидной последовательности какого-либо генома не является самоцелью. На самом деле это является первым шагом для исследования того, как функционирует та или иная клетка. Исследование геномов бактерий проводится для того, чтобы исследовать метаболизм бактерий и, в случае патогенных организмов, найти потенциальные мишени для лекарств. С другой стороны, изучение геномов может позволить найти новые метаболические пути или ферменты, которые будут применены в биотехнологическом производстве (например, витаминов). В течение как минимум полувека сотни лабораторий исследовали кишечную палочку (*E.coli*). Но даже такой весьма изученный организм имеет как минимум 25 % абсолютно не охарактеризованных генов. Значительное число секвенированных геномов принадлежат организмам, о которых вообще нет каких-либо других экспериментальных данных. Экспериментальное определение функции только одного гена требует интенсивной работы одной лаборатории как минимум в течение нескольких месяцев. Компьютерный же анализ позволяет с известной степенью точности охарактеризовать несколько тысяч генов силами небольшой группы примерно за неделю. Разумеется, компьютерный анализ не исключает экспериментальную проверку, однако в этом случае экспериментальная работа существенно упрощается.

Компьютерный анализ геномов состоит из следующих основных элементов:

- предсказание генов в последовательностях. При этом в некоторых случаях удается даже найти ошибки в последовательности;
- предварительная аннотация по сходству и другим особенностям белковых последовательностей;
- сравнительный анализ геномов;
- исследование регуляции работы генов;

- поиск «пропущенных» генов. Представим себе, что в клетке есть цепочка реакций, преобразующих вещество А в вещество Б, а затем вещество Б в вещество В. При этом ген, ответственный за первую реакцию известен и в клетке присутствует, а для второй реакции гена не нашли. Это и есть пропущенный ген. На самом деле вторая реакция осуществляется и проблема заключается в том, чтобы найти в геноме подходящую кандидатуру;

- исследование транспортеров (генов, обеспечивающих перенос питательных веществ в клетку и выброс вредных веществ из клетки).

Сравнительная геномика принесла уже несколько значительных открытий и «закрытий». В качестве «закрытия» можно привести триклозан, который считался универсальным антибактериальным препаратом. Он входит в состав широко разрекламированного мыла «Safeguard». Его мишенью является белок, закодированный в гене *fabI*. Этот белок катализирует одну из реакций синтеза жирных кислот — необходимого компонента любой клетки. При этом у животных нет аналога этого белка, поэтому такой препарат безопасен для человека. Компьютерный анализ бактериальных геномов показал, что стрептококки не имеют белка *fabI*, а его функцию выполняет совсем другой белок — *fabR*. Поэтому триклозан не действует на стрептококки. Одним из ярких открытий геномики является открытие принципиально новой системы регуляции — рибопереключателей. Это специфическая структура РНК, которая стабилизируется при непосредственном связывании с низкомолекулярным веществом и блокирует синтез матричной РНК.

Предсказание структуры и механизма действия было блестяще подтверждено экспериментально.

Другой класс исследований, проводимых компьютерной геномикой, — полногеномный анализ и исследование эволюции. В частности, с помощью массового анализа было обнаружено, что альтернативный сплайсинг в генах человека является скорее правилом, чем исключением. Эволюционный взгляд на проблему позволяет выдвинуть гипотезу, что сплайсинг, в частности альтернативный сплайсинг, является эффективным механизмом для эволюции, позволяющим без значительного риска для генома перебирать варианты последовательностей.

Массовый анализ большого количества геномов показал, что, по крайней мере у безъядерных организмов (бактерий и архебактерий), явление горизонтального переноса генов между видами является весьма распространенным явлением — от 10 до 30 % генов в этих геномах горизонтально перенесены из других видов. (Леск А. Введение в биоинформатику. М.: БИНОМ, Лаборатория знаний, 2009. )

## **2. Виды межмолекулярных взаимодействий**

### *2.1. Молекулярное узнавание.*

#### *Механизмы узнавания*

Узнавание (распознавание) сигнала рецептором есть основное свойство регулируемой и регулирующей системы, будь то человеческий мозг или электронная вычислительная машина. Такого рода системы являются узнающими. Они осуществляют классификацию объектов, информация о которых сообщается рецепторам. Эта классификация основывается на некотором принципе, заложенном в системе.

Узнающие системы могут быть не обучающимися и обучающимися. Вторые представляют особый интерес для кибернетики, теории автоматического регулирования, для моделирования деятельности головного мозга животных и человека и т. д. В качестве примера обучаемой узнающей системы можно назвать перцептрон — схему, моделирующую важные черты работы мозга, прежде всего его способность узнавать и классифицировать сигналы, получаемые полем рецепторов.

Очевидно, что возможность обучения узнающей системы определяется ее способностью обучаться, т. е. наличием в ее устройстве элементов памяти. Узнавание сигналов такой системой и является обучением с последующим «экзаменом». Способность обучаться, т. е. узнавать, запрограммирована в устройстве системы.

Обращаясь к глубинным уровням биологической организации, мы встречаемся с обучаемыми клеточными системами и с необучаемыми

мыми молекулярными системами узнавания. Клеточное узнавание имеет принципиальное значение для процессов развития, в частности для возникновения иммунитета. Молекулярное узнавание определяет все важнейшие молекулярно-биологические процессы — ферментативную активность, редупликацию ДНК, все этапы биосинтеза белка, взаимодействие «антиген — антитело» и т. д.

Молекула белка-фермента узнает молекулу субстрата или некоторую ее часть, как, например, в случае протеолитических ферментов, катализирующих гидролиз пептидных связей. Узнавание выражается в образовании реакционного комплекса со специфическим субстратом. Комплексы с ингибиторами и активаторами, с аллостерическими эффекторами также возникают в результате специфического узнавания. В узнавании участвуют непосредственно активный центр фермента, включающий и соответствующий ко фактор, и косвенно вся белковая глобула. Само образование глобулы можно трактовать как результат узнавания, в частности, узнавания гидрофобных остатков гидрофобными же остатками, вследствие чего формируется ядро глобулы.

Более простой случай узнавания реализуется при комплементарном связывании нуклеотидов в двойной спирали ДНК, в гибридной двойной спирали ДНК — РНК, в синтетических полинуклеотидах, при взаимодействии кодон — антикодон.

В биосинтезе белка мы встречаемся с ферментативным узнаванием, происходящем в акте транскрипции, в котором необходимым образом участвует РНК-полимераза, и в акте трансляции, где узнающими системами, наряду с мРНК, служат аминоацил — тРНК-синтетаза, вся рибосома и ряд других факторов.

Из приведенных данных следует определение термина «молекулярное узнавание». Это понятие имеет смысл применительно к системам, в которых узнающее устройство сохраняет свою целостность в акте узнавания и в ряде случаев возвращается в исходное состояние, совершив преобразование молекулярного сигнала. Одна молекула фермента перерабатывает множество молекул субстрата, одна рибосома «читает» весь текст, записанный в цепи мРНК. Можно было бы говорить о молекулярном узнавании и применительно к обычным химическим реакциям, причем



с тем большим основанием, чем они специфичнее. Однако «узнающий» реагент изменяется радикальным образом в ходе реакции и утрачивает способность к дальнейшим актам узнавания.

Таким образом, применяемое здесь определение молекулярного узнавания указывает на специфичность слабых, главным образом нехимических взаимодействий молекул. Для узнавания существенно стерическое соответствие структуры рецептора и структуры сигнальной молекулы, соответствие, фиксированное или индуцируемое. Специфическое понижение свободной энергии происходит вследствие многоточечного взаимодействия, которое и описывается как континуальное соответствие молекулярных поверхностей и находит свое наглядное выражение в атомных моделях. В сущности, старое представление Фишера о соответствии «ключ — замок» и сводится к узнаванию.

Достижение соответствия, как правило, связано с определенной перестройкой взаимодействующих систем, их, следовательно, с конформационными превращениями. Реализуются электронно-конформационные взаимодействия — ЭКВ.

Молекулярное кодирование в биологии основывается в конечном счете на молекулярном узнавании. Генетический код связан с функционированием ряда узнающих систем, перечисленных ранее. Естественно возникает вопрос о ферментном коде, т. е. о классификации соответствий между активными центрами ферментов и субстратами.

Громадное число комбинаций из 20 сортов аминокислотных остатков на поверхности реактивной полости фермента, в его активном центре, обеспечивает практически неограниченное многообразие функциональности ферментов. Можно думать о наличии фиксированных комбинаций, кодирующих узнавание характерных атомных групп субстратов. Точнее, следует говорить о кодовых сорбирующих комбинациях и о кодовых каталитических комбинациях, действующих согласованно, но пространственно разделенных. Имеются некоторые указания на возможность существования такого кода. Так, важный для катализа остаток Сер содержится в активном центре ряда эстераз, протеиназ и фосфому-таз. Для многих из этих ферментов характерна последовательность

— Гли — (Асп) — Сер\* — (Ала) —

Установлено, что Сер «узнает» ацильную группу, будучи ее промежуточным акцептором. Подробно изучены кодовые свойства и других функциональных групп ферментов: имид-азольной группы Гис, е-аминогруппы Лиз, карбоксильных, сульфгидрильных и дисульфидных групп. Тем не менее сегодня можно лишь поставить проблему ферментного кода. Решение этой проблемы требует обширной и разнообразной информации о структуре активных центров, получаемой прежде всего методом рентгеноструктурного анализа.

Взаимодействие, определяющее узнавание субстрата или ингибитора белком, есть процесс передачи информации молекулярным сигналом рецептору. В большинстве реальных случаев передается не вся информация, содержащаяся в данном объекте, но лишь некоторая ее часть, именуемая сигнатурой. Сигнатурой молекулы служат все те ее особенности, благодаря которым она становится участником данной реакции. В случае образования фермент-субстратного комплекса сигнатурой субстрата являются его функциональные группы, взаимодействующие с активным центром. В свою очередь, сигнатура фермента есть его активный центр, т. е. ограниченная совокупность аминокислотных остатков, непосредственно взаимодействующих с субстратом. Узнавание сводится здесь к структурному соответствию молекулярных сигнатур, реализуемому в результате многоточечных слабых взаимодействий.

Если обратиться к обучаемым узнающим системам, возникающим на более высоких уровнях биологической организации, то станет очевидным, что в результате обучения система должна перестать «обращать внимание» на несущественные обстоятельства. Иными словами, система обучается узнаванию сигнатуры.

Совершенство молекулярного узнавания имеет первостепенное значение для молекулярной биологии и биофизики, в частности для процессов развития и эволюции.

Специфичность ферментов не абсолютна. Данный фермент зачастую катализирует не определенную реакцию одного строго заданного субстрата, а однотипные реакции группы сходных субстратов. Это определяется двумя причинами. Первая непосредственно связана с общей программой онтогенеза и филогенеза, приводящей к оптимальной экономии числа действующих

белков. В тех ситуациях, в которых биологически существенна одна и та же реакция группы родственных субстратов, она может быть эффективно реализована единственным ферментом. Конечно, вся названная группа должна характеризоваться одной и той же сигнатурой или близкими сигнатурами.

Вторая причина наличия конечного интервала специфичности имеет молекулярно-кинетический характер. Реальная молекулярная узнающая система, фермент, предназначена не только для узнавания сигнала, но и для его достаточно быстрого преобразования. Степень специфичности узнавания, вообще говоря, симбатна степени связывания субстрата, т. е. выражается свободной энергией взаимодействия. Если выигрыш свободной энергии слишком велик, то прочность фермент-субстратного комплекса может быть настолько большой, что число оборотов фермента окажется чрезмерно низким. Необходимо оптимальное соотношение между стабильностью и скоростью преобразования. Эта ситуация с особенной ясностью проявляется в более простых случаях узнавания в полинуклеотидах и нуклеиновых кислотах. Приведем две таблицы.

Табл. 1 характеризует точность узнавания азотистых оснований РНК. Комплементарные пары АУ и ГЦ оказываются действительно наиболее прочными; так, АУ значительно прочнее АА или УУ. Однако возможно образование и некомплементарных пар, что и является одной из важнейших причин мутагенеза.

Константы стабильности достаточно низки, вследствие чего образование пар в полярных средах затруднено.

Таблица 1

**Константы ассоциации  $K_{асс}$  при спаривании оснований  
в полярных растворителях  $C_6H_6$  и  $CCl_4$  при 25 °С  
(2', 3', 5' – О-замещенные рибонуклеозиды)**

$C_6H_6$ \ $CCl_4$	$K_{асс} \cdot \text{моль}^{-1}$			
	У	А	Ц	Г
У	15 45	550	<50	<10 <sup>3</sup>
А	150	8 22	<50	<10 <sup>3</sup>
Ц	<28	<28	28 50	>10 <sup>4</sup>
Г	<1,12·10 <sup>3</sup>	<1,2·10 <sup>3</sup>	3·10 <sup>9</sup>	1,2·10 <sup>3</sup> ~10 <sup>3</sup>

Полужирным шрифтом показаны значения  $K_{acc}$  в растворителе  $C_6H_6$ , курсивом —  $CCl_4$ .

Табл. 2 свидетельствует о преимуществах в узнавании триплетов. Дублеты обладают слишком низкими значениями  $K_{acc}$ . Кодоны, содержащие более трех нуклеотидов, напротив, дают слишком прочное связывание. Время жизни пары «кодон — анти-кодон» не должно превышать нескольких миллисекунд, так как в противном случае оно будет лимитировать скорость работы каталитической системы.

Таблица 2

**Константы ассоциации  $K_{acc}$  при спаривании три- и тетра-нуклеотидов с антикодонами тРНК в водном растворе 1,0 М NaCl, 10 мМ MgCl<sub>2</sub> и 10 мММ фосфата при рН 7 и 0 °С**

Формил- Мет-тРНК AA [УАЦ] УЦ	$K_{acc}$ , моль <sup>-1</sup>	Тип-тРНК AA*[AU*] УЦ	$K_{acc}$ , моль <sup>-1</sup>
<b>АУФ</b>	1200±200	<u>УАЦ</u>	700
АУГА	13 500	УАЦА	90 000
АУГУ	1 400	<u>УАУ</u>	700
АУГЦ	900	УАУА	37 000
АУГГ	1 000	Фен-тРНК	
<u>ГУГ</u>	1 200	АУ[ААГ*]УЦ*	
ГУГА	9 800	<u>УУЦ</u>	900
ГУГУ	1 000	УУЦА	10 000
		<u>УУУ</u>	300
		УУУА	1 000

А\* — N(6)-диметил-А, У\* — псевдо-У, Г\* — 2-О-метил-Г, Ц\* — 2-Ометил-Ц. Дважды зачеркнуты регуляторные кодоны, один раз подчеркнуты — кодоны, соответствующие «влиянию». В квадратных скобках — антикодона.

Константы стабильности для комплементарных триплетов, входящих в состав олигомерных двойных спиралей, содержащих более чем четыре звена, заметно меньше, чем для триплетов и квадруплетов. По-видимому, это объясняется лучшими возмож-

ностями для вертикальных (stacking) взаимодействий в случае коротких экспонированных последовательностей, т. е. возможностями конформационных перемещений, приводящих к оптимальному связыванию. В молекулярном узнавании отчетливо проявляются кооперативные конформационные свойства биополимеров.

Характеристики узнавания, определяемые константами стабильности, относятся, конечно, лишь к термодинамическому равновесию. Биологические процессы редупликации ДНК, транскрипции и трансляции — кинетические процессы, идущие с участием соответствующих ферментов. В их основе лежит узнавание ДНК ДНК-полимеразой, лигазами, РНК-полимеразой, узнавание мРНК и тРНК рибосомой.

Для кинетических процессов существенны не только относительные глубины минимумов свободной энергии, отвечающих «структурам узнавания», но и барьеры, эти минимумы разделяющие. Узнавание непосредственно связано с ускорением соответствующего процесса, т. е. со снижением активационного барьера в результате слабых взаимодействий. Количественная теория этих процессов еще далеко не построена. (Волькенштейн М. В. Молекулярная биофизика. М.: Наука, 1975.)

## *2.2. Гидрофобные взаимодействия*

Неполярные молекулы обладают свойством объединяться друг с другом в полярном растворителе, стремясь таким образом минимизировать площадь контакта с ним. На макроскопическом уровне это наблюдается в виде так называемых гидрофобных (липофильных) взаимодействий, которые играют важнейшую роль в формировании пространственной структуры биологических макромолекул и целых систем, таких как, например, липидная бислойная мембрана. Также в ряде случаев они ответственны и за специфичное узнавание комплекса «рецептор — лиганд». Несмотря на широкую распространенность и фундаментальное значение для структуры биомacroмолекул и их комплексов, природа физических сил, лежащих в основе гидрофобного эффекта, не определена достаточно точно для их количественного описания.

Это объясняется тем, что гидрофобный эффект складывается из множества разных межмолекулярных взаимодействий: электростатического, Ван-дер-ваальсова, энтропийного эффекта и др. В связи с этим до сих пор не предложено универсального метода численной оценки свободной энергии гидрофобных взаимодействий. Тем не менее осознание ключевой роли гидрофобных взаимодействий в организации биомолекулярных систем дало толчок развитию эмпирических подходов к оценке гидрофобности. Самый простой способ — разделить все атомы на гидрофобные (углерод, сера, галогены) и гидрофильные (остальные атомы). Число контактов между гидрофобными атомами может служить приблизительной оценкой величины гидрофобных взаимодействий лиганда с рецептором.

Однако этот подход имеет серьезные ограничения и с его помощью не всегда можно добиться корректного предсказания. Так, гетероциклический фрагмент в составе лиганда (например, аденин в АТР) будет идентифицирован как гидрофильный, хотя известно, что аденин, входящий в состав АТР, предпочтительно связывается в гидрофобных карманах активных сайтов и, следовательно, обладает гидрофобными свойствами. Более точные методы, учитывающие при расчете свойств отдельных частей молекулы не только тип химического элемента, но и влияние окружения, основаны на экспериментальных данных по распределению органических соединений между полярной и неполярной фазами (обычно вода / октанол). Логарифм коэффициента распределения вещества между двумя средами —  $\log P$  — традиционно служит количественной мерой его гидрофобности.

Эмпирические атомные константы гидрофобности используют при изучении пространственного распределения гидрофобных / гидрофильных свойств, например на поверхности молекулы. Согласно концепции молекулярного гидрофобного потенциала (МГП), гидрофобные свойства рассчитывают на поверхности молекулы или в любой другой точке пространства, используя функцию, зависящую от расстояния (рис. 1).

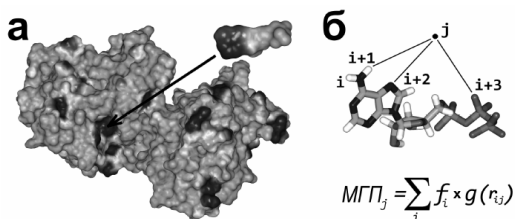


Рис. 1. Расчет молекулярного гидрофобного потенциала

МГП<sub>*j*</sub> — величина МГП в точке пространства *j*, *f<sub>i</sub>* — атомная константа гидрофобности, *g(r<sub>ij</sub>)* — функция затухания, зависящая от расстояния между атомом *i* и точкой *j*.

Необходимо отметить, что, как уже было сказано, строгого описания гидрофобных сил не существует, поэтому используемые функции *g(r<sub>ij</sub>)* носят, по сути, интуитивный и эмпирический характер. При этом в некоторых работах было показано, что одни из них лучше описывают свойства низкомолекулярных соединений, а другие — макромолекул, таких как белки.

Для характеристики вклада гидрофобных взаимодействий при образовании комплексов «лиганд — рецептор» предложены различные критерии, оценивающие соответствие их гидрофобных / гидрофильных свойств. Существуют методы, основанные на сравнении величины и знака констант гидрофобности контактирующих атомов или значений МГП двух молекул на поверхности интерфейса. Другие методы оценивают гидрофобные взаимодействия просто по числу контактов между гидрофобными атомами или химическими группами, либо по доле гидрофобной поверхности, экранированной от растворителя, либо по комплементарности гидрофобных областей на поверхности двух молекул (образованию энергетически выгодного контакта между этими областями). Для количественной оценки комплементарности гидрофобных свойств систем типа АТР-белок используют параметр, учитывающий долю заглубленной гидрофобной поверхности лиганда.

### 2.3. Стэкинг-взаимодействия

Стэкинг-взаимодействия — это гидрофобные связи, которые возникают при таком расположении ароматических молекул, которое напоминает расположение монет в стопке и поддерживается ароматическими взаимодействиями. Наиболее популярный пример такого расположения наблюдается в последовательных парах оснований ДНК. Стэкинг-взаимодействия также часто наблюдаются в РНК и белках.

Ароматическое взаимодействие (или  $\pi$ - $\pi$  взаимодействие) — это нековалентное взаимодействие между органическими соединениями, содержащими ароматические компоненты.  $\pi$ - $\pi$  взаимодействия вызываются межмолекулярным перекрыванием р-орбиталей в  $\pi$ -сопряженных системах, так что они становятся сильнее, поскольку число  $\pi$ -электронов возрастает.

В ДНК и РНК параллельный стэкинг имеет место между соседними парами нуклеотидов и повышает стабильность молекулярной структуры. Азотистые основания нуклеотидов имеют пуриновые или пиримидиновые группы в своем составе, состоящие, в свою очередь, из ароматических колец. В молекуле ДНК ароматические кольца расположены примерно перпендикулярно оси спирали, поэтому их поверхности расположены параллельно, что способствует перекрыванию р-орбиталей этих оснований.

Стэкинг-взаимодействие между основаниями довольно специфично: полярные заместители одного основания,  $-\text{NH}_2$ ,  $=\text{N}-$ ,  $=\text{O}$ , или галогены, нависают над ароматическим кольцом соседнего основания. Такого рода специфичность стэкинга даже более выражена, чем специфичность при образовании водородных связей.

Интенсивность стэкинг взаимодействия между пуриновыми (а и г, два кольца) и пиримидиновыми (у и с, одно кольцо) основаниями уменьшается в следующем порядке:

Пурин-пурин  $\rightarrow$  Пурин-пиримидин  $\rightarrow$  Пиримидин-пиримидин.

Среди различных типов межмолекулярных контактов особого внимания заслуживает стэкинг ароматических колец, также играющий важную роль в молекулярном узнавании. Хотя многие лиганды, в том числе и лекарственные соединения, содержат ароматические фрагменты и группы, стэкинг часто не учитывается



явно при составлении ОФ. Стэкинг-взаимодействия наблюдаются между двумя ароматическими группами, в результате чего они принимают определенную ориентацию друг относительно друга в пространстве. Самый известный пример — двойная спираль ДНК, где азотистые основания в результате стэкинга располагаются параллельно друг другу. Также возможно и перпендикулярное («Т-образное») взаимное расположение ароматических колец, что показано, например, для бензола. Кроме того, ароматические соединения имеют тенденцию участвовать в  $\pi$ -катионном взаимодействии, при котором образуется контакт между положительно заряженными группами и электронным облаком кольца.

Ароматические контакты, так же как и водородные связи, могут быть описаны с помощью геометрических критериев. В исследованиях взаимодействий АТФ с различными белками было показано, что аденину свойственно образование стэкинга с боковыми цепями аминокислотных остатков белка, преимущественно с фенилаланином, что согласуется и с результатами других работ. Для того чтобы выявить параметры контактов, образованных в результате стэкинга, мы предлагаем анализировать взаимную ориентацию двух ароматических фрагментов соединений в терминах расстояния между центрами циклов и угла между их плоскостями (рис. 2).

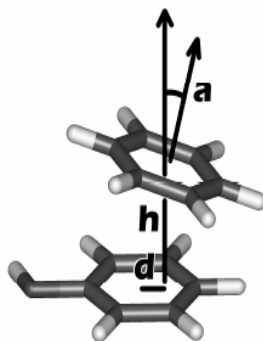


Рис. 2. Стэкинг взаимодействия двух ароматических систем

Диапазон значений параметров, определяющих наличие или отсутствие стэкинга в расчетных алгоритмах, до сих пор остается не выясненным и в оценочных критериях выбирается достаточ-

но произвольно. Картину усугубляет еще и тот факт, что многие ароматические соединения стремятся расположиться не только параллельно, но и перпендикулярно друг другу. Уточнение взаимного расположения ароматических колец для конкретных систем повышает эффективность оценки качества и достоверности структур комплексов «белок — лиганд», предсказываемых методами молекулярного моделирования. Например, в результате анализа пространственных структур комплексов различных белков с лигандами, содержащими в своей структуре гуанин, было установлено, что для таких лигандов характерно образование «параллельного» стэкинга с расстоянием от плоскости кольца гуанина до центра ароматического фрагмента аминокислотного остатка ( $h$ ), равным 3–4 Å, и иногда — «Т-образных» контактов с  $h = 4.5\text{--}5.5$  Å, как в случае с тирозином.

Стоит заметить, что гуанидиновая группа аргинина, так же как и ароматические аминокислоты, склонна образовывать параллельные контакты с аденином. (Молекулярный докинг: роль невалентных взаимодействий в образовании комплексов белков с нуклеотидами и пептидами / Т. В. Пырков, И. В. Озеров, Е. Д. Балицкая, Р. Г. Ефремов. URL: [http://www.rjbc.ru/2010/4/2010\\_36\\_4\(2\).pdf](http://www.rjbc.ru/2010/4/2010_36_4(2).pdf))

## 2.4. Межмолекулярные взаимодействия

Межмолекулярные взаимодействия определяют механизм столкновений молекул и некоторые стадии химических реакций. Они также определяют существование молекулярных жидкостей и твердых тел и их свойства: термодинамические и кинетические характеристики, энергию сцепления, равновесную геометрию, фононные спектры и др. Межмолекулярные взаимодействия обусловлены слабыми дальнедействующими силами и лежат в пределах 1–40 кДж/моль, что по крайней мере на порядок меньше, чем энергия внутримолекулярной связи. Напомним, что уравнение Ван-дер-Ваальса, описывающее состояние реальных газов, содержит константу, учитывающую притяжение молекул. Поэтому силы притяжения между нейтральными атомами и молекулами на расстояниях, превышающих 5 Å, называют Ван-дер-ваальсовыми. Источниками сведений о межмолекулярных

взаимодействиях служат эксперименты по рассеянию атомных и молекулярных пучков, термофизические измерения свойств жидкостей и газов, измерения энергий сублимации, фононных спектров и упругих характеристик кристаллов, спектроскопические измерения колебательно-вращательных спектров, времен спиновой и спиново-решеточной релаксации. Все эти методы дают лишь косвенную информацию и не позволяют установить природу взаимодействий, что является задачей квантовой химии.

Ван-дер-ваальсова связь (молекулярная связь) характерна для молекулярных кристаллов. Наблюдается у ряда веществ между молекулами с ковалентным характером внутримолекулярного взаимодействия. Наличие межмолекулярного притяжения в этом случае возможно при согласованном движении валентных электронов в соседних молекулах. В любой момент времени электроны должны быть максимально удалены друг от друга и максимально приближены к положительным зарядам (ядрам молекулы). Тогда силы притяжения валентных электронов ядром соседней молекулы оказываются сильнее сил взаимного отталкивания электронов оболочек этих молекул. Подобное притяжение между флуктуирующими электрическими диполями получило название дисперсионного взаимодействия. Примером кристалла с молекулярной связью может служить кристаллическая решетка иода. Так, твердый иод имеет молекулярную кристаллическую решетку, в узлах которых находятся молекулы  $I_2$ . Аналогичным образом построена кристаллическая решетка твердого диоксида углерода (сухой лед) — в узлах кристаллической решетки находятся молекулы  $CO_2$ .

Ван-дер-ваальсовое взаимодействие делится на ориентационное, индукционное и дисперсионное.

Ориентационные силы, диполь-дипольное притяжение осуществляется между молекулами, являющимися постоянными диполями. В результате беспорядочного теплового движения молекул при их сближении друг с другом одноименно заряженные концы диполей взаимно отталкиваются, а противоположно заряженные притягиваются. Чем более полярны молекулы, тем сильнее они притягиваются и тем самым больше ориентаци-

онное взаимодействие. Энергия такого взаимодействия обратно пропорциональна кубу расстояния между диполями.

Дисперсионное притяжение (лондоновские силы). Взаимодействие между мгновенным и наведенным диполем. При сближении молекул ориентация микродиполей перестает быть независимой и их появление и исчезновение в разных молекулах происходит в такт друг другу. Синхронное появление и исчезновение микродиполей разных молекул сопровождается их притяжением. Энергия такого взаимодействия обратно пропорциональна шестой степени расстояния между диполями.

Индукционное притяжение. Взаимодействие между постоянным диполем и наведенным (индуцированным). Встречаются полярная и неполярная молекулы. Под действием полярной молекулы неполярная молекула деформируется, и в ней возникает (индуцируется) диполь. Индуцированный диполь притягивается к постоянному диполю полярной молекулы и, в свою очередь, усиливает электрический момент диполя полярной молекулы. Энергия такого взаимодействия обратно пропорциональна шестой степени расстояния между диполями.

Относительный вклад каждого из видов межмолекулярных сил зависит в основном от двух свойств взаимодействующих молекул: полярности и поляризуемости (деформируемости). Чем выше полярность, тем значительнее роль ориентационных сил: чем больше деформируемость, тем значительнее роль дисперсионных сил. Индукционные силы зависят от обоих факторов, но сами обычно играют второстепенную роль.

Ван-дер-ваальсовы силы обуславливают агрегатное состояние вещества.

По сравнению с ковалентной связью Ван-дер-ваальсово взаимодействие очень слабое.

Связь Ван-дер-Ваальса является наиболее универсальной, она возникает между любыми частицами, но это наиболее слабая связь, энергия ее примерно на два порядка ниже энергии ионной и ковалентной связи. Поскольку дисперсионное взаимодействие оказывается очень слабым, молекулярные связи четко проявляются лишь в тех случаях, когда они возникают между атомами

или молекулами. Молекулярная связь легко разрушается тепловым движением. Поэтому молекулярные кристаллы обладают низкими температурами плавления (например, парафин), большими коэффициентами теплового расширения, большой сжимаемостью, малой твердостью. (Степанов Н. Ф. Квантовая механика и квантовая химия. М.: Мир, 2001. 519 с.)

## *2.5 Оценка энергии межмолекулярного взаимодействия*

Существует подход к квантово-химическому описанию межмолекулярных взаимодействий. Он состоит в вариационном расчете комплекса ABC... и составляющих его молекул А, В, С, ... Энергия межмолекулярного взаимодействия

$$E_{\text{межмол}} = E(\text{ABC}...) - [E(\text{A}) + E(\text{B}) + E(\text{C}) + \dots] \quad (1)$$

— маленькая величина, равная ~1 % от полной энергии системы, и для ее расчета необходимо очень точно вычислять энергии каждого из компонентов. По той же причине, что и в случае расчета энергии диссоциации молекул, здесь необходим учет электронной корреляции. Кроме того, чтобы описать потенциальную поверхность вдали от равновесия ( $R > 5 \text{ \AA}$ ), нужен широкий базисный набор, близкий к хартри-фоковскому пределу.

## *2.6. Оценка Ван-дер-ваальсовых атомных радиусов*

Когда молекулы в кристалле или жидкости связаны лишь Ван-дер-ваальсовыми силами, то расстояния между ними определяются так называемыми Ван-дер-ваальсовыми радиусами контактирующих атомов. Эти величины являются приближенными, обычно они определяются статистической обработкой данных рентгеноструктурного анализа в предположении, что радиусы либо аддитивны и  $R_{\text{AB}} = R_{\text{A}} + R_{\text{B}}$ , либо для них справедливо среднегеометрическое соотношение  $R_{\text{AB}} = 2(R_{\text{A}} R_{\text{B}})^{1/2}$ .

Квантово-химический способ оценки этих величин также приближенный и состоит в следующем. Результаты многих расчетов показывают, что типичная энергия Ван-дер-ваальсова взаи-

модействия нейтральных молекул примерно равна  $4 \text{ кДж}\cdot\text{моль}^{-1}$ . Тогда расчетом энергии пар атомов можно определить расстояние  $R = 2R_{\text{ВДВ}}$  между ними, при котором воспроизводится указанное значение энергии.

## 2.7. Водородная связь

Водородная связь — это взаимодействие между двумя электроотрицательными атомами одной или разных молекул посредством атома водорода:  $A-H \dots B$  (чертой обозначена ковалентная связь, тремя точками — водородная связь).

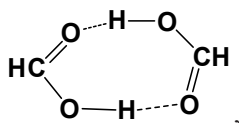
Водородная связь обусловлена электростатическим притяжением атома водорода (несущим положительный заряд  $\delta+$ ) к атому электроотрицательного элемента, имеющего отрицательный заряд  $\delta-$ . В большинстве случаев она слабее ковалентной, но существенно сильнее обычного притяжения молекул друг к другу в твердых и жидких веществах. В отличие от межмолекулярных взаимодействий водородная связь обладает свойствами направленности и насыщенности, поэтому ее нередко считают одной из разновидностей ковалентной химической связи. Она может быть описана с помощью метода молекулярных орбиталей как трехцентровая двухэлектронная связь.

Одним из признаков водородной связи может служить расстояние между атомом водорода и другим атомом, ее образующим. Оно должно быть меньше, чем сумма радиусов этих атомов. Чаще встречаются несимметричные водородные связи, в которых расстояние  $H \dots B$  больше, чем  $A-H$ . Однако в редких случаях (фтороводород, некоторые карбоновые кислоты) водородная связь является симметричной. Угол между атомами во фрагменте  $A-H \dots B$  обычно близок к  $180^\circ$ . Наиболее сильные водородные связи образуются с участием атомов фтора. В симметричном ионе  $[F-H-F]^-$  энергия водородной связи равна  $155 \text{ кДж/моль}$  и сопоставима с энергией ковалентной связи. Энергия водородной связи между молекулами воды уже заметно меньше ( $25 \text{ кДж/моль}$ ).

Водородные связи обнаружены во многих химических соединениях. Они возникают, как правило, между атомами фтора, азота и кислорода (наиболее электроотрицательные элементы),

реже — при участии атомов хлора, серы и других неметаллов. Прочные водородные связи образуются в таких жидких веществах, как вода, фтороводород, кислородсодержащие неорганические кислоты, карбоновые кислоты, фенолы, спирты, аммиак, амины. При кристаллизации водородные связи в этих веществах обычно сохраняются. Поэтому их кристаллические структуры имеют вид цепей (метанол), плоских двухмерных слоев (борная кислота), пространственных трехмерных сеток (лед).

Если водородная связь объединяет части одной молекулы, то говорят о внутримолекулярной водородной связи. Это особенно характерно для многих органических соединений. Если же водородная связь образуется между атомом водорода одной молекулы и атомом неметалла другой молекулы (межмолекулярная водородная связь), то молекулы образуют довольно прочные пары, цепочки, кольца. Так, муравьиная кислота и в жидком, и в газообразном состоянии существует в виде димеров:



а газообразный фтороводород содержат полимерные молекулы, включающие до четырех частиц HF. Прочные связи между молекулами можно найти в воде, жидком аммиаке, спиртах. Необходимые для образования водородных связей атомы кислорода и азота содержат все углеводы, белки, нуклеиновые кислоты. Известно, например, что глюкоза, фруктоза и сахароза прекрасно растворимы в воде. Не последнюю роль в этом играют водородные связи, образующиеся в растворе между молекулами воды и многочисленными OH-группами углеводов.

Атом водорода обладает особой способностью образовывать внутри и между молекулами мостиковые связи X-H...Y, соединяясь одновременно с двумя другими атомами, чаще всего с атомами C, O, N, S, Cl. Одна из этих связей по крайней мере на порядок слабее, чем обычная химическая связь (ее энергия составляет 10–40 кДж/моль) и называется водородной связью. H-связь играет исключительную роль в биологических системах,

где с ее помощью реализуется взаимодействие между различными функциональными группами в жизненно важных молекулах. Водородная связь проявляет себя в снижении частоты валентных колебаний группы X-H и в зависимости химического сдвига протона от окружения или природы растворителя.

Общий подход к описанию водородной связи не отличается от такового в случае межмолекулярных взаимодействий. Здесь также выделяют электростатическое взаимодействие невозмущенных молекул, поляризацию, перенос заряда, обменное и дисперсионное взаимодействия. Однако расстояния между атомами, соединенными атомом H, лежат в пределах 2,4–3,0 Å. Это большой разброс расстояний, поэтому говорят о сильной, промежуточной и слабой водородной связи. В первом случае трехцентровое взаимодействие X-H...Y с делокализацией электронной плотности доминирует и без привлечения квантово-химических представлений объяснить структуру H-связи невозможно. В последнем случае для качественных заключений достаточно использовать электростатический подход, который, впрочем, не объясняет направленный характер водородной связи.

Ясно, что форма потенциальной энергии атома H определяется расстоянием X-H...Y, т. е. силой связи. Симметричные связи наиболее прочны. (Цирельсон В. Г. Химическая связь и межмолекулярные взаимодействия. Конспект лекций по квантовой химии. URL: <http://quant.distant.ru/files/pdf/chbond.pdf>)

### **3. Способы прогнозирования и моделирования структуры и свойств вещества**

#### *3.1. Молекулярное моделирование*

**Молекулярное моделирование (ММ)** — собирательное название методов исследования структуры и свойств молекул вычислительными методами с последующей визуализацией результатов, обеспечивающей их трехмерное представление при заданных в расчете условиях (molecular modeling // IUPAG Gold Book).



Методы молекулярного моделирования используются в компьютерной химии, вычислительной биологии и науке о материалах для изучения как индивидуальных молекул, так и взаимодействия в молекулярных системах.

Расчеты простейших систем при молекулярном моделировании могут быть выполнены вручную, но из-за большого объема вычислений при молекулярном моделировании сколь-либо сложных систем, особенно при исследовании молекулярной динамики, используются компьютерные методы расчета и визуализации, эта техника получила название компьютерного молекулярного моделирования (англ. *computer-assisted molecular modeling, CAMM*) (computer-assisted molecular modeling, CAMM // IUPAC Gold Book).

Общей чертой методов ММ является атомистический уровень описания молекулярных систем — наименьшими частицами являются атомы или небольшие группы атомов. В этом состоит отличие ММ от квантовой химии, где в явном виде учитываются и электроны. Таким образом, преимуществом ММ является меньшая сложность в описании систем, обеспечивающая рассмотрение большего числа частиц при расчётах. (Немухин А. В. Компьютерное моделирование в химии // Соросовский образовательный журнал. 1998. Ж 6. С. 48–52.)

### 3.2. Создание компьютерной модели молекул

Вначале создается компьютерная модель объекта, а также применяется компьютерное моделирование для формирования молекул на месте проведения исследования. Модель может быть как двухмерной, так и трехмерной.

Для создания двухмерной модели используется теория графов. Граф — это абстрактная структура, которая содержит узлы, соединенные ребрами. В молекулярных графах узлам соответствуют атомы, а ребрам — связи между атомами. Часто в молекулярных графах опущены атомы водорода. Узлы и ребра могут иметь свойства, связанные с ними. Например, для узлов это может быть атомное число или тип атома, а для ребер — число связей. Эти свойства могут быть полезны при проведении операций над молекулярным графом. Графы дают только представление о том, в каком порядке связаны между собой атомы в молекуле.

Подграфы — это подмножество узлов и ребер в графе, например бензол является подграфом молекулярного графа аспирина. Таким образом, данный граф может быть выполнен различными способами и не соответствовать стандартному изображению химической структуры. Изображение в виде «дерева» — это особый вид графа, в котором есть только один путь подключения каждой пары вершины, то есть нет циклов и ароматических колец.

Стерические и электронные свойства могут зависеть от того, в каком положении в пространстве находится трехмерная структура или конформация.

Таким образом, это придает особый интерес разработке алгоритмов и систем баз данных, содержащих информацию о трехмерной структуре и конформации. Но использование трехмерной структуры связано с рядом проблем. Большинство молекул, представляющих интерес, может иметь более одной низкоэнергетической конформации, и во многих случаях количество доступных структур очень велико. Поэтому необходимы эффективные алгоритмы, принимающие во внимание конформационную гибкость. Истинные представления о молекулярных свойствах и характеристиках могут быть получены очень сложными вычислительными моделями, разработанными на основе квантовой механики и молекулярной симуляции. Этих вычислительных моделей для работы с большим числом молекул еще нет, и поэтому необходима разработка более эффективных вычислительных методов для представления ключевых характеристик молекулярных конформаций.

Данные о трехмерной структуре вещества получают из экспериментальных данных ядерно-магнитного резонанса или рентген-кристаллографии. После обработки экспериментальных данных представление о трехмерной структуре записывается в специальную базу данных. К примеру, Кембриджская структурная база (CSD) данных содержит данные о структуре полученных методом рентген-кристаллографии более чем 400 000 органических и органометаллических веществ. Протеиновая база данных содержит информацию о более чем 44 000 структурах протеинов, протеин-лигандных комплексов, нуклеиновых кислот, гидрокарбонатных струк-

турах, полученных в результате рентген-кристаллографии и ядерно-магнитного резонанса. Обе базы данных широко используются и постоянно обновляются.

### *3.3. Описание модели квантово-химическими расчетами. Молекулярная механика и динамика*

**Молекулярная механика** — один из подходов в молекулярном моделировании, использующий классическую механику для описания физических основ модели. Атомы (ядра с электронами) представляются точечными массами с соответствующими зарядами. Взаимодействия между соседними атомами включают упругие взаимодействия (соответствующие химическим связям) и силы Ван-дер-Ваальса, описываемые традиционно потенциалом Леннарда — Джонса. Электростатические взаимодействия вычисляются по закону Кулона. Атомам в пространстве присваиваются Декартовы, или внутренние, координаты; в динамических расчётах атомам также могут быть присвоены скорости, соответствующие температуре. Поверхность потенциальной энергии, которая в квантово-химических моделях подлежит прямому расчету, здесь аппроксимируется определенными эмпирическими функциями разной степени сложности, представляющими собой, например, суммы парных потенциалов взаимодействия атомов. Эти потенциальные функции, определяющие так называемое силовое поле молекулы, содержат некоторые параметры, численное значение которых выбирается оптимальным образом так, чтобы получить согласие рассчитанных и экспериментальных характеристик молекулы. В простейшем случае параметрами являются равновесные межъядерные расстояния и валентные углы, а также силовые постоянные, то есть коэффициенты жесткости упругих сил, связывающих пары атомов. Метод основан на допущении возможности переноса этих параметров из одной молекулы в другую, так что численные значения параметров, подобранные для некоторых простых молекул, используются далее при прогнозировании свойств других более сложных соединений.

Простейшие модели молекулярной механики учитывают растяжения связей, деформацию валентных и двугранных углов, взаимодействие валентно несвязанных атомов, называемое также Ван-дер-ваальсовым взаимодействием, электростатические вклады и т. д.:

$$U = \sum U_{\text{раст}} + \sum U_{\text{деф}} + \sum U_{\text{торс}} + \sum U_{\text{вдв}} + \sum U_{\text{эл-стат}}$$

где  $U_{\text{раст}}$  — энергия растяжения связей;  $U_{\text{деф}}$  — энергия деформации валентных углов;  $U_{\text{торс}}$  — энергия деформации двугранных углов;  $U_{\text{вдв}}$  — энергия Ван-дер-ваальсового взаимодействия;  $U_{\text{эл-стат}}$  — энергия электростатических вкладов.

Набор параметров, состоящий из равновесных значений длин связей, валентных углов, величин парциальных зарядов, силовых констант и Ван-дер-ваальсовских параметров, называется силовым полем. Различные реализации молекулярной механики используют слегка отличающиеся математические выражения и, следовательно, различные константы в потенциальной функции. Распространенные силовые поля, используемые в настоящее время, были разработаны с использованием точных *квантовых расчетов* и / или подгонкой под экспериментальные данные.

Для каждого слагаемого записывается определенное аналитическое выражение, и параметры соответствующих функций подгоняются по каким-либо свойствам базовых молекул. Например, для описания потенциальной функции предельных углеводородов при не очень высоких требованиях к точности расчета достаточно около десяти параметров.

Сумма всех перечисленных вкладов определяет энергию  $U$  молекулы как функцию геометрической конфигурации ядер, и для нахождения равновесной геометрической конфигурации исследуемой молекулы необходимо определить минимум  $U$  с помощью компьютерной программы поиска стационарных точек на многомерных потенциальных поверхностях. Таким образом, практические действия исследователя чаще всего сводятся лишь к заданию стартовой геометрии и вызову программы оптимизации геометрических параметров из условия минимума энергии. На выдаче просматривается полученная структура и, если необходимо, анализируются энергия и ее составляющие.

Молекулярная механика играет роль молекулярного конструктора: для первичной оценки строения интересующей нас молекулы зачастую проще собрать молекулу на компьютере, чем тратить время на поиск необходимой информации в справочной литературе. При расчетах молекулярной структуры на более высоком уровне методами квантовой химии полезно использовать координаты ядер молекулы, найденные с помощью молекулярной механики, в качестве начального приближения. Для многих задач, например для конформационного анализа, уровень моделирования методами молекулярной механики оказывается вполне достаточным для качественных и даже количественных заключений.

В каждом конкретном случае необходимо интересоваться, для каких классов соединений параметризована та версия программы, которую предполагается применять при моделировании свойств нового соединения. Особенно осторожно следует относиться к оценкам энергий, хотя и для геометрических конфигураций возможны грубые ошибки.

**Метод молекулярной динамики**, изначально разработанный в теоретической физике, получил большое распространение в химии и, начиная с 1970-х гг., в биохимии и биофизике. Взаимодействие между объектами может быть описано силовым полем (классическая молекулярная динамика), квантово-химической моделью или смешанной теорией, содержащей элементы двух предыдущих (QM/MM (quantum mechanics /molecular mechanics, QMMM (*англ.*))).

В основе методов молекулярной динамики лежит модельное представление о многоатомной молекулярной системе, в которой все атомы представляют собой материальные точки. Причём поведение отдельного атома описывается классическими уравнениями движения.

Для поиска локального минимума потенциальной энергии используются соответствующие методы минимизации (например, метод наискорейшего спуска и метод сопряженных градиентов), а для изучения поведения систем с течением времени используются *методы молекулярной динамики*. Низшие энергетические состояния более стабильны и имеют более важное значение из-за своей роли в химических и биологических процессах. Молекулярно-динамические расчеты, с другой сторо-

ны, показывают поведение системы как функцию от времени. И для минимизации, и для молекулярной динамики главным образом используется второй закон Ньютона —  $F = ma$  (или, что равносильно,  $a = F/m$ ). Интегрирование этого закона движения с помощью различных алгоритмов приводит к получению траекторий атомов в пространстве и времени. Сила, действующая на атом, определяется как отрицательная производная функции потенциальной энергии.

Молекулы могут быть смоделированы как в вакууме, так и в присутствии растворителя, например воды. Расчёты систем в вакууме называются расчётами «в газовой фазе», в то время как расчёты, включающие молекулы растворителя, называются расчётами «с явно заданным растворителем». Другая группа расчётов учитывает наличие растворителя оценочно, с помощью дополнительных членов в потенциальной функции — так называемые расчёты «с неявным растворителем».

При моделировании методами молекулярной динамики или Монте-Карло интересующее нас свойство системы большого числа молекул вычисляется через статистические средние по положениям и движениям молекул. Как и в методах молекулярной механики, здесь также необходимо перечислить все частицы системы и задать потенциалы межчастичных взаимодействий. Однако, в отличие от молекулярной механики, в данных подходах области задания межчастичных потенциалов взаимодействия должны быть достаточно протяженными и не должны ограничиваться малыми смещениями от положений равновесия. Это накладывает существенные более высокие требования на способы расчета потенциалов.

Как уже упоминалось, число частиц при моделировании методами Монте-Карло и молекулярной динамики с помощью современных суперкомпьютеров может достигать колоссальных величин. Даже без суперкомпьютеров достаточно типичны численные эксперименты для значений  $N$  порядка десятков и сотен тысяч. Примеры успешного применения методов Монте-Карло и молекулярной динамики для моделирования равновесных составов смесей при постоянном давлении, фазовых равновесий, адсорбции на поверхности твердых тел, свойств жидкостей в ми-

кропорах и т. д. достаточно многочисленны. Этими же методами решаются задачи поиска устойчивых конформаций (поворотных изомеров) полимерных молекул, чрезвычайно важные для биохимических приложений.

В настоящее время методы молекулярного моделирования стали обыденными при изучении структуры, динамики и термодинамики неорганических, биологических и полимерных систем.

Наиболее часто используемые силовые поля при расчётах био-макромолекулярных структур:

- AMBER (Assisted Model Building with Energy Refinement) используется для белков, нуклеиновых кислот и ряда других классов молекул. Не рекомендуется использовать для расчётов свойств материалов.

- CHARMM (Chemistry at HARvard Macromolecular mechanics) используется для различных систем, от небольших молекул до сольватированных комплексов биологических макромолекул.

- CVFF (Consistent Valence Force Field) включает уточняющие вклады ангармоничности и взаимодействия составляющих силового поля. Поле параметризовано для расчётов пептидов и белков.

В программной реализации молекулярной динамики внутренние координаты системы пересчитываются в Декартовы координаты атомов, и наоборот. (Степанов Н. Ф., Пупышев В. И. Квантовая механика молекул и квантовая химия. М.: Изд-во МГУ, 1991. 384 с.)

### *3.4. Квантово-химические методы расчета*

Реальные многоатомные системы содержат большое количество взаимодействующих электронов, а для таких систем не существует аналитического решения уравнений, и, по всей видимости, оно не будет найдено и в дальнейшем. По этой причине в квантовой химии приходится строить различные приближённые решения. Решение уравнения Шредингера часто строится на уравнении Хартри — Фока — Рутана итерационным методом (SCF-self consistent field — самосогласованное поле) и состоит в нахождении вида волновой функции. Приближения, используемые в квантовой химии:

1. Приближение Борна — Оппенгеймера (адиабатическое): движение электронов и движение ядер разделено (ядра движутся настолько медленно, что при расчёте движения электронов ядра можно принять за неподвижные объекты). В связи с этим приближением существует так называемый эффект Яна-Теллера. Данное приближение позволяет представить волновую функцию системы как произведение волновой функции ядер и волновой функции электронов.

2. Одноэлектронное приближение (или приближение Хартри): считается, что движение электрона не зависит от движения других электронов системы. В связи с этим в уравнения, используемые в квантовой химии, вносятся поправки на взаимное отталкивание электронов. Это позволяет волновую функцию электронов представить в виде суммы волновых функций отдельных электронов.

3. Приближение МО ЛКАО (Молекулярная Орбиталь как Линейная Комбинация Атомных Орбиталей): в данном подходе волновая функция молекулы представляется как сумма атомных орбиталей с коэффициентами. Решение задачи состоит в нахождении коэффициентов  $C$ .

Методы квантовой химии делятся на эмпирические (молекулярная механика), неэмпирические (*ab initio*) и полуэмпирические.

Неэмпирические методы в основном используют

- для сравнения с экспериментом;
- расчета тех свойств молекул, которые затруднительно определить экспериментально;
- развития полуэмпирических методов;
- расчета гипотетических соединений.

Главное и основное в неэмпирических методах — это базисный набор орбиталей и функций и их выбор.

Типы базисов (базисный ряды):

- STO-NG (минимальный);
- 3-21 G (расщепленный тройной-двойной);
- 6-31 G (расширенный);
- 6-31 \*G (расширенный + поляризационные АО);
- 6-31 \*\*G (расширенный + поляр + диффузн АО).



В зависимости от тех свойств молекулярной системы, которые нам нужно узнать, мы выбираем тот или иной базис. Например, для молекулярной геометрии мы выбираем минимальный базис, за исключением расчетов диэдральных углов и геометрии пирамидальных структур, где необходимо использовать поляризационные функции.

Полуэмпирические методы — это методы в которых пренебрегают или оценивают (параметризуют) из каких-либо соображений (например, симметрии) часть молекулярных интегралов кулоновского отталкивания в уравнении Рутаана. Полуэмпирические методы объясняют многие главные закономерности без привлечения сложных расчетов.

Прежде всего, используя приближение Борна — Оппенгеймера, задают структуру молекулы в виде координат ядер. Затем прибегают к приближению МО ЛКАО и выбирают аналитические функции, которыми будут аппроксимироваться АО. Эти функции называются базисом.

В настоящее время существует большое количество вариантов упрощения решения уравнений метода Хартри — Фока — Рутаана, основанные на параметризации матричных элементов и игнорировании ряда из них, упрощения вида обменного члена и многих других. Самое частое упрощение — валентное приближение, при котором в разложении МО в ЛКАО учитываются только электроны валентной оболочки. Внутренние электроны (1s, например) считаются локализованными, экранирующими ядра, и они образуют неполяризованный атомный остов. Другое приближение — заменяют энергию отталкивания ядер энергией отталкивания «остов — остов».

В табл. 3 приведена сравнительная характеристика полуэмпирических методов.

В настоящее время имеется большое количество квантово-химических программ, реализующих те или иные расчетные методы. Наиболее полное их собрание существует в Фонде квантово-химических программ при Химическом университете Индиана (США).

Таблица 3

*Сравнительная характеристика полумпирических методов*

<i>Метод (приближение)</i>	<i>Параметризуемое свойство</i>	<i>Хорошо воспроизводимые свойства</i>	<i>Плохо воспроизводимые свойства</i>
CNDO/2	Электронная плотность	Дипольные моменты, длины связей, валентные углы, силовые константы	Теплоты образования, потенциал ионизации, сродство к электрону, спектры, реакции
CNDO/S INDO/S, ZINDO	Спектр	Спектр	Теплоты образования, геометрия молекул, реакции
INDO	Спиновые плотности	Спиновые плотности, константы сверхтонкого взаимодействия, геометрия молекул	Теплоты образования, потенциалы ионизации, сродство к электрону, спектры,
MINDO/3	Потенциал атом-атомного взаимодействия	Теплоты образования, потенциалы ионизации, длины связей	Спектры, водородная связь
MNDO	Теплоты образования	Теплоты образования, геометрия молекул	Спектры, водородная связь
AM1	Теплоты образования	Теплоты образования, геометрия молекул	Спектры
PM3	Теплоты образования, параметры межмолекулярного взаимодействия	Теплоты образования, геометрия молекул, водородная связь, межмолекулярные взаимодействия	Спектры

RM1	Теплоты образования, дипольных моментов, потенциалов ионизации и геометрических величин (длин связей и углов)	Является параметризацией метода AM1. Параметры метода определяются по 1736 молекулам, содержащим атомы C, H, N, O, P, S, F, Cl, Br, I и др.	
PM6	Теплоты образования, геометрия молекулы (оптимизация аминов, соединенных с ароматическим кольцом в правильную плоскую геометрию)	Является усовершенствованной версией метода PM3. Параметризованы все основные группы элементов и группы переходных металлов	

Неэмпирический метод Хартри — Фока и его расширения за счет учета электронной корреляции реализован в нескольких компьютерных программах (прежде всего GAUSSIAN, GAMESS и CADPAC).

А одними из самых популярных и доступных программ, реализующих полуэмпирические методы, являются HyperChem (кроме того, добавлены методы молекулярной механики), ChemOffice ( ChemLab, ChemDraw, ChemCad и т. д.).

## Контрольные вопросы

1. Установление первичной структуры ДНК, значение.
2. Характеристика и значение рестриктаз. Механизмы действия, номенклатура.
3. Высокоповторяющиеся последовательности ДНК (ВПП) и их значение.
4. Умеренно повторяющиеся последовательности ДНК (УПП) и их значение.
5. Уникальные последовательности ДНК (УП), их значение.
6. Особенности генома про- и эукариот.
7. Методы гибридизации нуклеиновых кислот. Мозаичная структура генов эукариот, роль интронов и экзонов.
8. Понятие о процессинге продуктов транскрипции генов.
9. Процессинг предшественников р-РНК.
10. Процессинг предшественников т-РНК. Эндонуклеаза. Каталитическая роль ДНК.
11. Процессинг предшественников и-РНК эукариот. Роль малых ядерных РНК.
12. Репликация ДНК: значение, механизм.
13. Транскрипция ( синтез РНК). Особенности. Характеристика РНК-полимеразы.
14. Обратная транскрипция, ее суть и значение.
15. Биосинтез белка на рибосоме. Значение, основные этапы.
16. Процесс активирования аминокислот: значение, механизм.
17. Трансляция (механизм сборки полипептидной цепи на рибосоме ). Основные этапы.
18. Генетический код и его особенности.
19. Изучение свойств биологических молекул при помощи методов молекулярного моделирования.
20. Метод молекулярной механики.
21. Полуэмпирические методы квантовой химии. Методы, использующие нулевое дифференциальное перекрытие. Расширенный метод Хюккеля. Метод Хюккеля для электронных

систем. Возможности и ограничения применения полуэмпирических методов квантовой химии.

22. Неэмпирические методы расчета.

23. Методы теории функционалов плотности.

24. Межмолекулярное взаимодействие. Ориентационная и индукционная составляющие. Дисперсионное взаимодействие. Ван-дер-ваальсовы комплексы. Водородная связь.

## Литература

1. Бирштейн, Т. М. Конформации макромолекул / Т. М. Бирштейн, О. Б. Птицын. — М. : Наука, 1964.
2. Волькенштейн, М. В. Молекулярная биофизика / М. В. Волькенштейн. — М. : Наука, 1975.
3. Волькенштейн, М. В. Конфигурационная статистика полимерных цепей / М. В. Волькенштейн. — М. : Изд. АН СССР, 1959.
4. Дэн, Г. Строки, деревья и последовательности в алгоритмах: Информатика и вычислительная биология / Г. Дэн; пер. с англ. И. В. Романовского. — СПб. : Невский Диалект; БХВ-Петербург, 2003. — 654 с.
5. Игнасиоту, С. Основы биоинформатики / С. Игнасиоту. — М.; Ижевск : НИЦ «Регулярная и хаотическая динамика», Институт компьютерных исследований, 2007. — 320 с.
6. Леск, А. Введение в биоинформатику / А. Леск. — М. : БИНОМ, Лаборатория знаний, 2009. — 318 с.
7. Немухин, А. В. Компьютерное моделирование в химии / А. В. Немухин // Соросовский образовательный журнал. —1998. — Ж 6. — С. 48–52.
8. Степанов, Н. Ф. Квантовая механика и квантовая химия / Н. Ф. Степанов. — М. : Мир, 2001. — 519 с.
9. Степанов, Н. Ф. Квантовая механика молекул и квантовая химия / Н. Ф. Степанов, В. И. Пупышев. — М. : Изд-во МГУ, 1991. — 384 с.
10. Denissiouk, K. A. X-ray crystal structure of MTH938 from *Methanobacterium thermoautotrophicum* at 2.2 Å resolution reveals a novel tertiary protein fold / K. A. Denissiouk, V. V. Rantanen, M. S. Johnson // PROTEINS. — 2001. — V. 44. — P. 282–291.
11. Cappello, V. Structure of an anti-HIV monoclonal Fab antibody fragment specific to a gp120 C-4 region peptide / V. Cappello, A. Tramontano, U. Kock // PROTEINS. — 2002. — V. 47. — P. 106–115.
12. Lipophilicity in Molecular Modeling / B. Testa, P. A. Carrupt, P. Gaillard, F. Billois, P. Weber // Pharm. Res. —1996. — V. 13. — P. 335–343.
13. Torshin, I. Y. Bioinformatics in the Post-Genomic Era: The Role of Biophysics / I. Y. Torshin. — Novapublishers, 2006.

## Оглавление

1. <i>Биоинформатика</i> .....	3
1.1. Развитие биоинформатики.....	3
1.2. In silico или in vivo?.....	6
1.3. Цели и задачи биоинформатики.....	7
1.4. Основные направления биоинформатики.....	9
1.5. Биоинформатика последовательностей.....	9
1.6 Структурная биоинформатика.....	11
1.7. Компьютерная геномика.....	13
2. <i>Виды межмолекулярных взаимодействий</i> .....	15
2.1. Молекулярное узнавание. Механизмы узнавания.....	15
2.2. Гидрофобные взаимодействия.....	21
2.3. Стэкинг-взаимодействия.....	24
2.4. Межмолекулярные взаимодействия.....	26
2.5 Оценка энергии межмолекулярного взаимодействия.....	29
2.6. Оценка Ван-дер-ваальсовых атомных радиусов.....	29
2.7. Водородная связь.....	30
3. <i>Способы прогнозирования и моделирования структуры     и свойств вещества</i> .....	32
3.1. Молекулярное моделирование .....	32
3.2. Создание компьютерной модели молекул.....	33
3.3. Описание модели квантово-химическими расчетами. Молекулярная механика и динамика.....	35
3.4. Квантово-химические методы расчета.....	39
<i>Контрольные вопросы</i> .....	44
<i>Литература</i> .....	46

Учебное издание

**КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ  
МАКРОМОЛЕКУЛЯРНЫХ ОБЪЕКТОВ  
И БИОЛОГИЧЕСКИХ СИСТЕМ**

*Методические указания*

Составители:

**Базлов** Дмитрий Александрович  
**Цивов** Алексей Владимирович

Редактор, корректор М. Э. Левакова  
Верстка Е. Б. Половковой

Подписано в печать 25.06.13. Формат 60×84 1/16.

Усл. печ. л. 2,79. Уч.-изд. л. 2,0.

Тираж 50 экз. Заказ

Оригинал-макет подготовлен  
в редакционно-издательском отделе ЯрГУ

Ярославский государственный университет им. П. Г. Демидова.  
150000, Ярославль, ул. Советская, 14.