

БИОИНФОРМАТИКА: ВИРТУАЛЬНЫЙ ЭКСПЕРИМЕНТ В ШАГЕ ОТ РЕАЛЬНОСТИ

Биоинформатика как наука появилась на стыке молекулярной биологии, генетики, математики и компьютерных технологий. 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.

В июле 2004 года в новосибирском Академгородке прошла четвертая Международная конференция по биоинформатике регуляции и структуры генома (BGRS'2004), организованная Институтом цитологии и генетики Сибирского отделения Российской академии наук. В течение недели более 200 ученых из разных стран - России, Германии, США, Англии, Италии, Франции, Японии и других - обсуждали проблемы компьютерной геномики, системной и эволюционной биологии, а также протеомики - нового научного направления, связанного с исследованием белков и их взаимодействия в организме. На конференции побывала специальный корреспондент журнала "Наука и жизнь" Г. АФАНАСЬЕВА.



КОГДА БИОЛОГИЯ НАЧАЛА ГОВОРИТЬ ЯЗЫКОМ МАТЕМАТИКИ...

Так когда же появилась биоинформатика, которую сейчас с уверенностью называют наукой третьего тысячелетия? Вот что говорит об этом доктор биологических наук Рустэм Нурович Чураев, заведующий отделом эпигенетики Института биологии Уфимского научного центра и заведующий лабораторией математической и молекулярной генетики:

"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 века 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 в биологию пришли физики и математики. Это событие во многом индуцировал Шредингер выходом в свет своей книги "Что такое жизнь с точки зрения физики?". Среди них были такие знаменитые ученые, как Алексей Андреевич Ляпунов и Игорь Андреевич Полетаев. Благодаря их научным работам начал развиваться кибернетический подход к биологическим процессам. Вадим Александрович Ратнер, ученик Ляпунова, в начале 1960-х годов применил математический подход к описанию явлений молекулярного уровня в сложных системах.

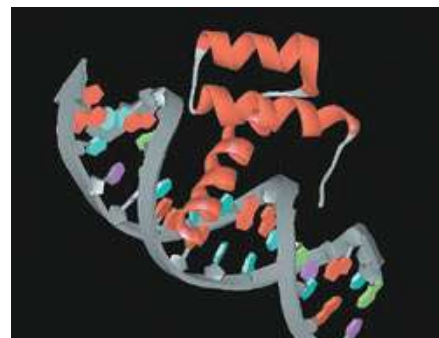


**Член-корреспондент РАН
Н. А. Колчанов открывает четвертую
Международную конференцию по
биоинформатике регуляции
и структуры генома.**

В 1975 году 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 немыслимы без компьютеров. Накапливается много данных, идет вал информации о первичных последовательностях геномов молекул ДНК".

IN SILICO ИЛИ IN VIVO?

Конечно, для анализа огромного массива данных требуются высокопроизводительные вычисления, и обычный компьютер не способен справиться с такой задачей - не хватит 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. Возможно, как раз 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.



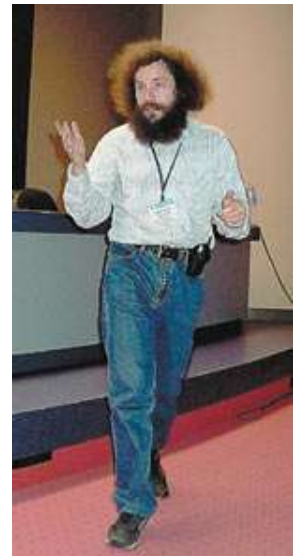
Трехмерная модель контакта регуляторного белка с молекулой ДНК. О молекулярном моделировании пространственной структуры белков и их взаимодействии с нуклеиновыми кислотами рассказали Р. Г. Ефремов (Институт биоорганической химии) и К. В. Шайтан (МГУ).

Суперкомпьютеры необходимой мощности существуют и в России, например в Пущине. При наличии оптоволоконных сетей 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", говоря 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 обычного. 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 Пущинском научном центре. Создание "электронной клетки" даст возможность эффективно конструировать лекарства нового поколения. О перспективах биоинформатики и о доверии экспериментаторов к предсказаниям in silico говорил доктор биологических наук Михаил Сергеевич Гельфанд:

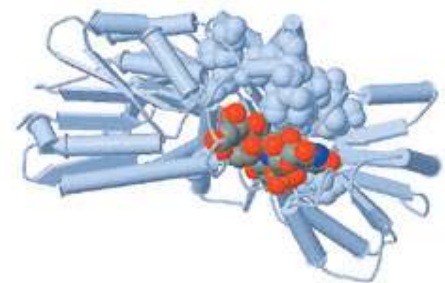
"Биоинформатика - это способ делать биологические утверждения, не наблюдая непосредственно природу (как, например, в зоологии) и не ставя эксперименты (как в молекулярной биологии), а путем анализа уже полученных массовых экспериментальных результатов. Конечно, основой всего является геном: только зная все гены организма, можно делать надежные выводы. А массовые данные получаются из экспериментов, поставленных на индустриальный поток: это расшифровка геномов, изучение синтеза закодированных в генах белков, определение белок-белковых взаимодействий в различных организмах, клетках, условиях роста... Раньше для получения такого рода данных ставились специальные эксперименты, но представьте себе: у средней бактерии примерно 3000 генов, у человека - на порядок больше, 30 000, белков еще больше - сколько же надо провести экспериментов, например, для того, чтобы проверить все пары белков, взаимодействуют ли они. И при этом основное время как раз уйдет на "пустые" эксперименты, с не взаимодействующими белками. Массовый эксперимент - это сотня экспериментов одновременно, которые выполняют несколько десятков роботов. Однако при этом возникает высокий уровень "шума" - случайных ошибок, и задача биоинформатики - извлечь из этих данных достоверные утверждения.



Заседание, посвященное эволюционной биологии, ведет доктор биологических наук М. С. Гельфанд.

То же самое с функциями генов. В принципе, выяснение функции одного гена в эксперименте - это год работы сильного исследователя и хорошая статья. Но генов-то тысячи в каждом организме. И вот оказывается, что биоинформатика позволяет делать надежные предсказания, которые поддаются уже целенаправленной экспериментальной проверке. Это получается намного проще, быстрее и дешевле.

Конечно, бывают и ошибочные выводы, что вызывает недоверие у "традиционных" биологов. Но тут ситуация такая же, как с любой новой техникой исследования. Когда появился микроскоп, тоже сначала мерещилось много всякого, например, некоторые описывали, что видят маленького зародыша в сперматозоиде. Но постепенно экспериментаторы приобретают вкус к биоинформатическим методам и начинают более критично относиться к своей работе. В конечном счете возникает совершенно новое мощное средство для биологических исследований. Собственно, оно уже сейчас есть, надо его развивать и учиться им пользоваться - для этого и проводятся подобные конференции".



Пространственная организация и функционирование активных центров белков.



Чтобы сконструировать лекарство точечного действия, нужно подобрать вещество (лиганд), которое входит в белок-мишень, как ключ в замок. Для этого необходимо знать пространственную структуру молекул.

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

КОМПЬЮТЕР ПРОЕКТИРУЕТ ЛЕКАРСТВА

Одна из самых перспективных и быстро развивающихся областей биоинформатики - конструирование лекарств направленного действия. Действие таких препаратов нацелено на центры связывания конкретного белка в организме возбудителя болезни.

При этом аналогичные белки человека не подвергаются изменениям, а значит, нет и побочных эффектов.

Создание лекарства направленного действия требует знаний о трехмерной структуре белка-мишени, так как точное пространственное соответствие играет ключевую роль.

Структура белков зашифрована в ДНК, но разобраться в этом шифре не так-то просто. Дело в том, что ген белка помимо кодирующих фрагментов (их называют экзонами) содержит интроны - участки, которые никакой полезной информации о белке не несут. То есть "слова" генетического кода прерываются довольно длинными вставками из случайного набора "букв". Большое число докладов на конференции BGRS'2004 были посвящены поиску новых эффективных методов распознавания экзон-интронной структуры ДНК, а также определению пространственной структуры белков по известной последовательности аминокислот. Об этом говорили в своих выступлениях Ю. Л. Орлов (Институт цитологии и генетики, Новосибирск), М. А. Ройтберг (Институт математических проблем биологии РАН, Пущино) и многие другие. И. И. Абнизова (Великобритания) рассказала о новых вычислительных подходах, которые позволяют судить о функции участков ДНК по тому, каким образом распределены в нем нуклеотиды.



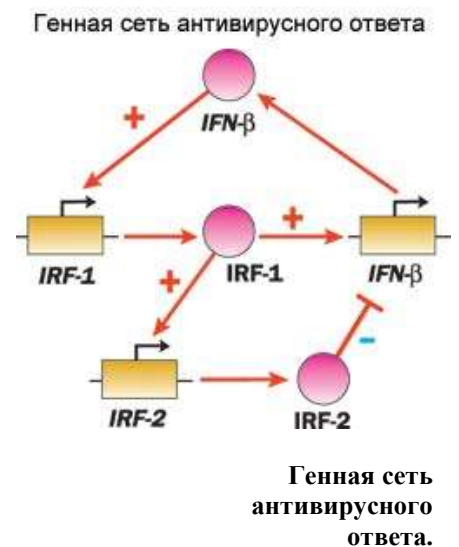
Нуклеотидная последовательность ε-гемо-глобина человека

Сегодня наиболее эффективный метод анализа белков - не постепенная расшифровка структуры для каждого отдельного белка, а сравнение со структурами родственных (гомологичных) белков, которые уже расшифрованы. Наиболее близкая из известных последовательностей берется за "точку отсчета", и дальше проводится ее уточнение. Данные рентгеноструктурного анализа и ядерно-магнитного резонанса позволяют предсказывать пространственную структуру схожих белков и делать вывод об их возможной функции в организме.

Когда пространственная структура белка-мишени установлена, наступает следующий этап: поиск низкомолекулярного вещества (лиганда), которое, соединившись с белком, будет оказывать нужное фармакологическое действие. Такой поиск связан с перебором не одной тысячи вариантов, и без компьютерных технологий тут не обойтись.

ГЕННЫЕ СЕТИ

Гены в клетках организма могут взаимодействовать друг с другом посредством своих продуктов - белков. Например, регуляторные белки способны связываться с определенными участками ДНК, и, таким образом, один ген может включить или выключить другой. Благодаря подобному взаимодействию образуется генная сеть, охватывающая значительное количество генов (от десятков до сотен), которые координируют свою деятельность и контролируют выполнение определенных функций в организме. Выяснение механизмов функционирования генных сетей представляет принципиально важную задачу, ведь именно они определяют внешние признаки организма и наследственные заболевания. Полная и ясная картина взаимодействия генов откроет новые возможности для генной диагностики и генной терапии. Эта тема была одной из центральных на конференции.

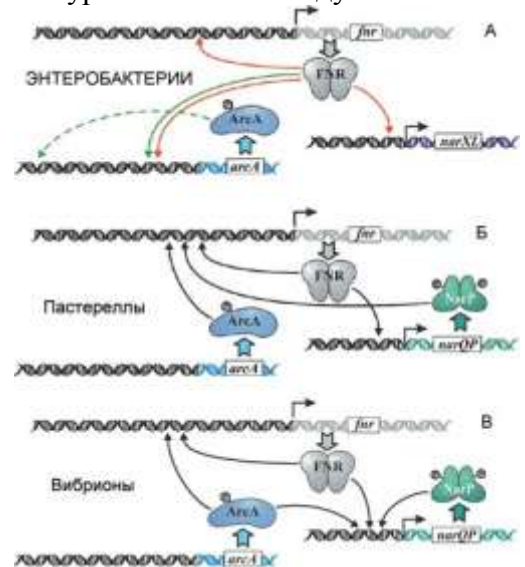


В конструировании искусственных генных сетей используется представление об эпигене как новом классе наследственных единиц, разработанное Р. Н. Чураевым в 1975 году. Эпиген - это "надгенная" система с двумя или более наследуемыми режимами работы генов (например, когда определенный ген "включен" и когда он "выключен"). Эпиген запоминает режим, в котором он находится, и передает информацию об активности генов по наследству. Такая концепция позволяет объяснить наблюдаемые экспериментально случаи наследования приобретенных признаков и ряд других феноменов.

ЭВОЛЮЦИЯ ЧЕРЕЗ ПРИЗМУ БИОИНФОРМАТИКИ

Сравнение геномов различных организмов дает возможность проследить эволюцию живых организмов. В секции компьютерной эволюционной биологии М. С. Гельфанд представил доклад, посвященный эволюции регуляторных систем бактерий. Анализ сотен бактериальных геномов позволил описать поведение регуляторных систем в ходе эволюции на самых разных уровнях: эволюцию отдельных регуляторных сигналов, изменения в структуре белков-регуляторов, поведение наборов регулируемых генов и, наконец, полную перестройку регуляторных систем. Возможно, самая древняя регуляторная система - так называемые РНК-переключатели, о которых рассказал доктор биологических наук Андрей Александрович Миронов. РНК-переключатель представляет собой молекулу РНК, которая способна не только переносить информацию, но и управлять активностью гена самостоятельно, без помощи белков. Регулирующие свойства РНК-переключателей были вначале предсказаны методами биоинформатики, а затем обнаружены экспериментально. В пользу предположения о древности системы РНК-переключателей говорят два факта: присутствие регуляторных сигналов такого типа практически во всех группах организмов от бактерий и археобактерий до растений и грибов и то, что регуляция осуществляется непосредственно при связывании регуляторной структуры РНК с малыми молекулами (витаминами и аминокислотами) без участия какого-либо посредника. Тем самым такие регуляторные системы могли существовать в "РНК-мире" еще до появления белков.

С практической точки зрения интересно исследование, проведенное в группе Гельфанда А. Е. Казаковым, О. В. Калининой и Е. А. Перминой. Они исследовали устойчивость бактерий к тяжелым металлам. Во многих случаях эта устойчивость обеспечивается белками-транспортерами, выбрасывающими токсичные катионы металла из клетки. Однако различных транспортеров много, они весьма разнообразны, и предсказать специфичность белка-транспортера к тому или иному металлу удастся далеко не всегда. Исследователи пошли другим путем: они изучили, как регулируется работа генов, кодирующих белки-транспортеры. Оказалось, что в ней участвует определенное семейство белков-репрессоров, узнающих о присутствии катионов



Регуляция системы дыхания



Оживленные дискуссии продолжались в перерывах между заседаниями. Сложную науку объясняет "на пальцах" гость из Индии доктор Сринивасан Рамачандран (Институт геномики, Дели).

металлов при помощи специального функционального участка, в который входят не менее трех остатков аминокислоты цистеина. Связавшись с катионом, белок-репрессор освобождает сигнальный участок ДНК и тем самым запускает работу гена, кодирующего белок-транспортер. Авторы показали, что по последовательности сигнального участка в ДНК можно предсказать, к какой группе относится соответствующий белок-репрессор, а эти белки высокоспецифичны к катионам металлов. Полученные результаты могут быть использованы для создания надежных индикаторов на загрязнение. А по схеме, построенной учеными, можно предсказать, к каким металлам будет нечувствительна та или иная бактерия, что особенно важно при создании штаммов, применяемых для биологической очистки промышленных стоков.

БИОИНФОРМАТИКА - НАУКА ИЛИ МЕТОД ИССЛЕДОВАНИЯ?

На этот вопрос, вызывающий немало споров среди ученых, ответил председатель оргкомитета конференции член-корреспондент РАН Николай Александрович Колчанов:

"И наука, и метод, и стратегия исследования. Метод - если рассматривать биоинформатику как набор технологий, позволяющий работать с огромным массивом данных и получать с их помощью новые результаты. Научная сторона биоинформатики связана с развитием информационной биологии. Как любая наука, она имеет собственные задачи и занимается изучением механизмов хранения, передачи и реализации биологической информации на разных уровнях. Таких уровней несколько: геном, клетка, взаимодействие между клетками, организм, популяция в целом. Биоинформатика появилась на стыке наук - молекулярной биологии, физиологии, математики, информатики, физики и химии, а такого рода объединения имеют стратегический характер. Самые крупные достижения возникают тогда, когда происходит синтез наук. Биология и информатика - классическое отражение этой ситуации".

Первая конференция по биоинформатике регуляции и структуры генома состоялась в Академгородке по инициативе Института цитологии и генетики Сибирского отделения РАН в 1998 году. Что изменилось за прошедшие шесть лет? Вновь слово Н. А. Колчанову:

"В 2002 году закончена расшифровка генома человека. Следующая важнейшая задача - расшифровать протеом. Этот термин образован от слова "протеин" (по аналогии с геномом) и означает совокупность белков, которые функционируют в организме. Конечно, получение "белкового портрета" организма потребует времени, но в принципе эта задача вполне решаема. Общий объем накопленной информации сейчас таков, что на первый план выходит системная биология, цель которой - не просто объединить достижения, полученные различными методами, но интегрировать знания и перевести их на качественно новый уровень. Помимо глобальных задач появилось много интересных прикладных направлений, например технология ДНК-микрочипов. Изменения в направлениях исследований отражаются и в тематике конференции".

Иллюстрация "Пространственная организация и функционирование активных центров белков".

Конфигурация активного центра белка (выделен цветом) обеспечивает избирательность биохимических взаимодействий. Новейшие достижения в исследованиях пространственной организации и функционирования активных центров белков представили М. Ондрехен (США), С. Рамачандран (Индия), Д. А. Афонников, В. А. Иванисенко (Институт цитологии и генетики, г. Новосибирск) и другие участники конференции.

Иллюстрация "Нуклеотидная последовательность ϵ -гемо-глобина человека".

Нуклеотидная последовательность ϵ -гемо-глобина человека. Синим цветом выделены информативные участки - экзоны, красным - "бессмысленные" интроны. Одна из задач биоинформатики - определить экзон-интронную структуру и распознать те области в геномах, которые кодируют белки.

Иллюстрация "Генная сеть антивирусного ответа".

Генная сеть, приведенная на рисунке в качестве примера, включает комбинацию положительных и отрицательных обратных связей. При попадании вируса в клетку происходит активация гена интерферона- β (IFN- β). Белок (фактор транскрипции) IRF-1 вначале усиливает работу гена, продуцирующего интерферон, а позднее запускает синтез другого белка IRF-2, который подавляет производство интерферона, обеспечивая возвращение системы в исходное состояние. Прямоугольниками на рисунке обозначены гены, а кружками - белки.

Иллюстрация "Регуляция системы дыхания".

В регуляции системы дыхания у бактерий участвуют различные белки-регуляторы (факторы транскрипции), в частности FNR (переключатель аэробно-анаэробного метаболизма), ArcA (датчик наличия кислорода) и NarP (регулятор азотного дыхания). Д. А. Равчеев и А. В. Герасимова из исследовательской группы М. С. Гельфанда провели компьютерное сравнение бактериальных геномов и установили, что у микроорганизмов разных групп эти белки могут формировать различные регуляторные каскады. На рисунке белки представлены цветными фигурами, кодирующие их гены - фрагментами ДНК соответствующего цвета, а регуляторные взаимодействия показаны как тонкие стрелки. Из экспериментальных работ было известно, что у кишечной палочки (А) основным фактором является FNR, который регулирует свой собственный ген *fng*, а также гены, кодирующие другие регуляторы (*arcA* и *narXL*). Оказалось, что у пастерелл (Б) факторы FNR и NarP образуют петлю (регулируют гены друг друга) и, кроме того, ген *fng* регулируется факторами FNR (авторегуляторная петля) и ArcA. У вибрионов (В) обнаружена авторегуляция генов *fng* и *narP*, а ArcA регулирует оба этих гена и, по-видимому, является основным регулятором.