

©А.И.Арчаков

ЧТО ЗА ГЕНОМИКОЙ? – ПРОТЕОМИКА.

А.И.АРЧАКОВ

Научно-исследовательский институт биомедицинской химии РАМН, Москва

Восемнадцатый международный конгресс по биохимии и молекулярной биологии, проходивший с 16 по 20 июля 2000 г в Великобритании, закончил свою работу. Его название символично: “Beyond the Genomics”- что мы видим в будущем вслед за эрой геномики? Надо сразу сказать, что первая же пленарная лекция одного из ведущих специалистов в области геномики - J. Craig Venter - директора фирмы «Celera Genomics», задала тон всему конгрессу. Да, действительно в апреле 2000 года было закончено непосредственное секвенирование генома человека. Фантастическая техника, работающая в этой фирме, позволила это сделать даже быстрее, чем ожидалось. Летом этого же года, используя мощные алгоритмы, «Celera» смогла представить в более или менее окончательном варианте непрерывный сиквенс генома человека. Вероятно, лето 2000 года следует считать такой же вехой в истории развития науки о жизни, как отправка человека в космос, создание атомной бомбы и т.д. в других науках.

Геном человека известен. Что дальше? Что сделано? По сути дела, и это самое главное, создана первая мощная структурная база данных, инвентаризированы гены человека и около 50 видов микроорганизмов и растений, что позволяет с определенным оптимизмом смотреть в будущее. Эра структурной геномики началась недавно, в 1991 г. проектом “Геном человека”, но так бурно развивалась, что, по сути дела, прошла пик своего развития именно в этом году. Проект «Геном человека», продолжавшийся 10 лет, в основном, закончен. Вложение примерно трех миллиардов долларов в решение этой проблемы, привело к тому, что прочитано три миллиарда оснований нуклеиновых кислот, то есть стоимость работ один доллар – одно основание. В результате этого картированы все двадцать три хромосомы человека, стали известны десятки, а может быть, даже и сотни генов, которые ответственны за развитие наследственных заболеваний. Нам известны гены, которые контролируют продолжительность жизни клетки, регулируют опухолевый рост и многое другое. Создана новая концепция в медицине – все заболевания можно разделить на

два больших класса: наследственные болезни, которые являются следствием передачи от родителей к потомству дефектных генов, и ненаследственные заболевания, так называемые социально значимые, так как они составляют более 95 % всех болезней человека, которые также связаны с геномом человека. Но, в отличие от первой группы болезней, в данном случае нормальный ген начинает функционировать ненормально в связи с нарушением регуляции его экспрессии. Таким образом, все болезни человека, так или иначе, связаны с геномом, только одна группа является следствием дефекта в структуре гена, а другая является нарушением регуляции экспрессии генов генома человека. Инвентаризация генов человека – это достижение, которое вряд ли можно переоценить, и оно еще много лет будет определять развитие как фундаментальной биохимии и молекулярной биологии, так и многих прикладных наук, так или иначе связанных с геномом. Однако впереди у этой науки, которая получила название «геномика», еще более блестящее будущее. Акцент исследований в настоящее время сместился от человека и животных в сторону исследования геномов микроорганизмов (особенно патогенных) и растений. Эти работы обещают огромные достижения в медицине, особенно в борьбе с инфекционными заболеваниями, а также в сельском хозяйстве. Огромный интерес к данной проблеме в ведущих странах мира неслучаен. Действительно произошло событие, которое трудно переоценить и которое будет определять жизнь всего нашего общества на ближайшее будущее. Дискуссия активно продолжается во всех социальных слоях и, наверное, небезыntenесно будет знать, что нас ждет впереди. А вот здесь самое поразительное заключается в том, что, несмотря на очень горячий интерес, который современная пресса проявляет к проблеме геномики, ученые уже обсуждают другую, проблему: а что же будет тогда, когда проблемы геномики будут решены окончательно, а сейчас это уже вопрос даже и не науки, а современной техники, современных технологий. Что представляет собой постгеномная эра? Где же лежит следующая горячая точка науки о жизни?

Каким образом информационная структура – ген – связана с реально работающей молекулярной машиной – белком? Чтобы правильно ответить на этот вопрос важно помнить следующее положение: в немногочисленных работах, которые выполнены с целью сопоставить карту мРНК с белковой картой в тех же самых клеточных системах, показано, что такой строгой корреляции не существует [1-4]. А так как белки, по сути дела, осуществляют все известные биологические функции, то, следовательно, информационные знания не могут быть прямо переведены в знание реально работающих белковых молекул. Кроме того, совершенно очевидно, что активация или инактивация белков в посттрансляционных механизмах не может быть определена, исходя из уровня мРНК в клетке. Следовательно, геномный анализ, анализ экспрессии генома на уровне мРНК и реально существующие белковые карты биологических систем отражают различные свойства этих систем и несут, по сути дела, или комплементарную или, иногда, может быть, даже исключаящую информацию. Что же нужно делать? И ответ на этот вопрос, казалось бы, предельно прост: нужно провести инвентаризацию существующих белков в клетке. И действительно, возникла такая наука, которая получила название «протеомика», и ответ на вопрос «Что же будет за эрой геномики?» в настоящее время совершенно

прост: будет эра протеомики. Термин этот происходит от двух других хорошо известных в биохимии терминов: "PROTEins" и "genOMe" и впервые был использован в 1995 г [5]. Конечно, геномика не исчезнет, она будет развиваться с той же самой, а может даже и большей скоростью, но ясно, что центр постгеномных исследований будет перенесен в область инвентаризации и выяснения протеомной карты человека. На первый взгляд, задача кажется совершенной не решаемой. Если геномная карта человека одинакова, по сути дела, для всех клеток человека (это 23 хромосомы с одним и тем же набором генов - исключение составляют половые клетки), то в случае протеомной карты человека говорить об общности ее совершенно бессмысленно: каждая клетка, каждая ткань, каждая биологическая жидкость должна иметь собственную протеомную карту. Несмотря на то, что в каждой клетке может быть только около 100000 функционирующих генов, многочисленные реакции модификации могут увеличить число белков в клетке до 10-20 миллионов. Каким же образом будет решаться эта проблема? И здесь надо сказать, что если ее решение где-то лет пять тому назад казалось совершенно не возможным, то в настоящее время создан метод, который позволяет достаточно эффективно решать эту проблему. Здесь мне хотелось бы дать два определения протеомики: узкое, которое можно назвать структурной протеомикой, и более широкое, которое включает и структурную, и функциональную части протеомики. В узком смысле этого слова **протеомикой является наука, занимающаяся инвентаризацией белков с помощью комбинированного использования методов: двумерного электрофореза, масс-спектрометрического анализа молекулярной массы и последовательности разделенных электрофорезом белков биологического материала с последующим анализом полученных результатов методами биоинформатики.** По сути дела, структурная протеомика - это комбинация 2D-электрофореза, масс-спектрометрии и биоинформатики. И если огромные разрешающие возможности двумерного электрофореза известны очень давно, с первой работы О'Farrell в 1975 г [6], то возможности масс-спектрометрического анализа очень быстро определять молекулярную массу и последовательность полипептидных цепей стали ясны только в самое последнее время. Развивались они настолько быстро, что сейчас некоторыми фирмами созданы уже полностью автоматизированные системы для определения молекулярной массы и последовательности белков, работающие на фемтомолярном и атомомолярном уровнях концентраций [7-9]. С помощью комбинации этих методов можно создать протеомную карту любого биологического материала, которая представляет собой фенотипическое проявление генома клетки, ткани или даже целого органа. В более широком смысле термины **протеомный анализ, или протеомика** могут быть использованы не только для инвентаризации белков биологического объекта, но и для контроля обратимой посттрансляционной модификации белков специфическими ферментами, как-то: фосфорилирование, гликозилирование, ацилирование, френилирование, сульфирование и т.д.

В настоящее время уже более 300 различных типов посттрансляционной модификации охарактеризовано с помощью протеомики [10]. Другой аспект функциональной протеомики - это выяснение состава функционально активных комплексов, составляющих различные метаболические цепи, а также выяснение

взаимодействия различных белков или субъединиц в составе олигомерных комплексов также с помощью комбинации методов выделения этих комплексов с последующим масс-спектрометрическим анализом. Структурная протеомика в последнее время еще получила название экспрессионная протеомика, а функциональная протеомика получила название клеточно-картируемой протеомики, так как выясняет она взаимодействие белков в метаболических путях [5].

Вначале кратко оценим возможности структурной протеомики. Комбинация двумерного электрофореза и масс-спектрометрии привела к совершенно непредсказуемому результату. Оба метода развивались настолько быстро, что существующие в них ограничения были очень быстро преодолены. Одним из основных ограничений метода двумерного электрофореза являлась его плохая воспроизводимость, особенно в первом направлении, в котором белки разделялись по их изоэлектрической точке. Однако создание иммобилизованных градиентов pH (сокращенное название нового типа электрофореза IPG-Dalt) позволило преодолеть эту нестабильность и сделать результаты анализа вполне воспроизводимыми, не только в одной, но даже в различных лабораториях [11-13]. Разрешающая способность метода зависит, естественно, от экспериментальных условий, но теоретически вполне возможно разделение полипептидных цепей, различающихся по одному заряду, а по молекулярной массе, вполне возможно различать полипептидные цепи, различающиеся на один аминокислотный остаток. Современные методы масс-спектрометрии позволяют анализировать не отдельно выделенные белки, но даже их смеси. В какой-то мере нерешенной проблемой 2D-электрофореза продолжает оставаться визуализация гелей. Несмотря на то, что в настоящее время окраска серебром в сто раз более чувствительна, чем окраска кумасси, и позволяет определять 0,1 нг белка, этот метод в настоящее время заменяется окрашиванием с помощью особого реактива Sypro Ruby [13]. Лимит метода – 1-2 нг белка, но он позволяет преодолеть недостатки окрашивания серебром: не все белки прокрашиваются серебром, окраска неустойчива, имеет плохую воспроизводимость и т.д. Возможности двумерного электрофореза таковы, что одновременно может быть выявлено до 10000 различных белков. Несмотря на то, что основные успехи протеомики связаны с двумерным электрофорезом, в настоящее время продолжают быстро развиваться многие другие методы разделения белков на первом этапе протеомного анализа, такие, как капиллярный электрофорез, капиллярная хроматография, двумерная и многомерная хроматографии.

В то же время не вызывает сомнения, что решающие достижения протеомики связаны с масс-спектрометрическим анализом. Здесь, естественно, следует выделить несколько подходов. Первый из них, который наиболее широко используется в структурной протеомике, – это комбинация методов MALDI (Matrix Assistant Laser Desorb Ionisation) с TOF (Time-of-Flight). Последовательность анализа такова: белки из неокрашенного 2D-геля электроблоттингом переносятся на мембрану с иммобилизованным трипсином, где они, соответственно, расщепляются до пептидов. Затем эти пептиды улавливаются второй мембраной, которая считывается в MALDI - TOF-масс-спектрометре, шаг за шагом, с определенным интервалом между пятнами, в результате чего определяются молекулярные массы этих пептидов. По сути дела,

содержимое пятна второй мембраны является в геле отпечатком белка в пептидной форме. В результате два типа данных доступны после масс-спектрометрического анализа. Простейшие масс – спектрометры, такие, как MALDI-TOF анализаторы измеряют массу полученных пептидов. Тандемные масс-спектрометры позволяют также прочитать последовательность этих пептидов [5, 8]. На второй стадии этого анализа ионизированные пептиды, выбранные по массе в первом масс-спектрометре, разрушаются дальше при столкновении с молекулами газа и фрагментированные остатки пептидов анализируются во втором спектрометре. Конечно, последовательность плюс масса более информативный анализ, чем одна масса. Однако для многих случаев уже достаточно бывает и определения массы пептида, так как существующие методы биоинформатики (на основании имеющихся баз данных и мощных алгоритмов их анализа) позволяют установить принадлежность того или иного пептида тому или иному белку и идентифицировать этот белок [5, 8, 14]. Такой подход особенно эффективен для тех объектов, где геномная карта уже известна. Существующие базы данных позволяют проводить этот анализ достаточно быстро и эффективно. При этом, однако, следует помнить, что для идентификации пептидов используются специальные базы данных, содержащие не целые белки как таковые, а полученные из них компьютерным путем (*in silico*) пептидные фрагменты. Это прежде всего SWISS-2D PAGE, SEQUEST, PROWL и др. [14]. При этом, если вспомнить, что для представления результатов 2D-электрофореза используется собственное программное обеспечение и базы данных, такие как ProteographTM, ROSETTATM, то становится очевидным, что методы биоинформатики в данном случае играют столь же важную роль, что и сам масс-спектрометрический анализ [14, 19]. В настоящее время прилагаются огромные усилия для того, чтобы увеличить эффективность алгоритмов, идентифицирующих пептиды, и алгоритмов, которые конструируют соответствующие белки из этих пептидов, и эта наука развивается очень быстро. В тех же случаях, когда такой простой путь не реализуется, можно использовать тандемную масс-спектрометрию или комбинацию MALDI с квадрупольным TOF-спектрометром [8]. Другой метод подготовки биологического материала для анализа, получивший название электроспрейной ионизации (или в наилучшем варианте наноэлектроспрейной ионизации). Метод широко используется не только для установления аминокислотной последовательности белков, но и для выявления модифицированных белков [5, 8]. В этом случае в отличие от MALDI, полученный спрей пригоден для анализа с помощью HPLC и капиллярного электрофореза. Методы электроспрейной ионизации в плане секвенирования гораздо более эффективны, но в силу своей медлительности (25 нл/мин) они не могут использоваться эффективно для выяснения последовательности огромного количества белков, получаемых при 2D-электрофорезе. Таким образом, в настоящее время следует различать два подхода: первый - комбинацию методов 2D-электрофореза с MALDI-TOF-анализом и, в будущем, тандемной спектрометрии, что позволяет весьма эффективно секвенировать полученные белковые зоны электрофореграм. При этом необходимо помнить, что для анализа многих белков вполне достаточно определения их молекулярной массы и координат в 2D-электрофорезе. Если же речь идет о характеристике какой-то

определенной группы белков, в том числе и модифицированных, а также белков, находящихся в метаболических комплексах, или белков, взаимодействующих друг с другом, то метод электроспрейной ионизации, особенно наноэлектроспрейной ионизации, является методом выбора [5, 8]. Метод наноэлектроспрейной ионизации позволяет вводить в анализируемую камеру масс-спектрометра биологический материал очень медленно до 1-2 нл/мин, и поэтому здесь отдельные пептиды могут считываться в режиме реального времени. В настоящее время для анализа последовательности модифицированных белков этот метод является предпочтительным. В ближайшее время следует ожидать новых впечатляющих успехов в этой области. По заявлению руководителей фирм "Celera" и "PE Biosystems" следует ожидать появления на порядок более быстрых масс-спектрометров, которые позволят идентифицировать до 1 миллиона белков в день [15].

Несмотря на огромную мощь и еще не реализованные возможности комбинации методов разделения и очистки белков, с одной стороны, и масс-спектрометрии - с другой, начали создаваться и развиваться так называемые протеомные микрополя. ДНК-овые и белковые микрополя высокой плотности - это небольшие плоские поверхности, которые позволяют одновременно анализировать тысячи молекулярных параметров в одном эксперименте. Если в области геномики этот метод, основывающийся на PCR-анализе, является одним из доминирующих в настоящее время [16], то в области протеомики они лишь начинают появляться [17, 18]. Обычно они основаны на роботизированном создании весьма компактных чипов с иммобилизованными белками, которые после взаимодействия их с биологическим материалом проявляются с помощью иммунохимического анализа или какого-либо другого окрашивания [18]. К сожалению, ситуация здесь гораздо сложнее, чем при анализе ДНК, так как основной метод полимеразной цепной реакции, который используется при создании ДНК-овых микрополей, является, естественно, неприемлемым. Другое ограничение связано, конечно, с высокой ценой производства антител к соответствующим белкам, поэтому создание белковых микрополей и их использование - это более дорогой процесс. Однако следует помнить, что в диагностическом плане эти поля имеют, конечно, наибольшие перспективы в будущем [17].

Естественно, что как и геномика, протеомика ориентирована на создание новых лекарственных препаратов, в которых те или иные белки будут служить молекулярными мишенями [5, 20]. Процесс нахождения новых мишеней для действия лекарств сложен, однако, следует помнить, что в случае геномики эта проблема эффективно решается при помощи биоинформатики, причем объектом анализа является геном. Однако, после анализа генома нам необходимо получать дополнительные доказательства того, что данный белок эффективно экспрессируется и находится в клетке в рабочем состоянии. Протеомика этот барьер преодолевает, т.е. в данном случае совместное использование геномики и протеомики абсолютно необходимо. Анализируя геном, мы находим молекулярную генетическую мишень для лекарства. Затем с помощью протеомного анализа убеждаемся, что данный ген экспрессируется, и только в этом случае белок уже используется в качестве молекулярной мишени. Однако,

протеомика может и сама по себе решать проблемы нахождения мишеней. Если получены протеомные карты нормальных и патологически измененных тканей, то по их разнице можно установить, какие белки важны для развития того или иного патологического состояния и выбрать их в качестве мишеней или использовать эти знания для диагностики [20]. Общее же в обоих подходах заключается в том, что с помощью методов молекулярного моделирования мы должны пройти путь от последовательности (гена) белка к молекулярной модели белка, к строению его активного центра и затем, основываясь на этих знаниях, компьютерным путем сконструировать наилучший лиганд, который будет являться, чаще всего ингибитором, в некоторых случаях, активатором. Таким образом, этот путь геномика и протеомика проходят одновременно. Что еще общего в обеих науках? Общее, на мой взгляд, заключается сейчас в том, что лимитирующей стадией в их развитии является не собственно метод(ы) идентификации генов, идентификации белков, их секвенирования – в этом достигнуты огромные методические успехи. Лимитирующим звеном является все-таки биоинформатика. Действительно, без соответствующих алгоритмов анализа генома, он является просто непрерывной цепью нуклеиновых оснований, точно также как в отсутствие соответствующего анализа последовательностей белковых молекул, без их декодирования, сиквенс является просто цепочкой аминокислотных остатков. Создание же методов, позволяющих анализировать как геном человека, так и последовательность белковых молекул, т.е. перевод тетрамерного кода нуклеиновых кислот и 20-мерного кода белков в бинарный код вычислительных машин, эффективный анализ существующих баз данных, создание новых баз данных, содержащих информацию о структурно-функциональных доменах белков, об экспрессируемых участках генома, об экзон-интронной структуре генов, о молекулярной трехмерной организации белковых молекул и молекул нуклеиновых кислот, по сути дела, является сейчас решающим в достижении конечной цели этих наук – выяснении механизма функционирования биологических текстов. В информационном плане – это молекулы нуклеиновых кислот, а в реально работающих молекулярных машинах – белки. Получив общее представление о протеомике как науке структурной, функциональной биохимик может задаться вопросом: «А какова же конечная цель, где горизонты и какие основные направления развития этой науки?». Здесь мне хотелось бы остановить ваше внимание на следующем: что если в случае геномики ориентиры уже видны – это, естественно, инвентаризация генов и, прежде всего, человека, исследование в сравнительном плане геномов животных, и прежде всего, близких к человеку приматов, а также исследование геномов сельскохозяйственных важных растений, патогенных микробов, важных в медицинском аспекте, то в случае протеомики горизонты более размыты. Ясно, что в ближайшем будущем цель инвентаризации и установления абсолютно всех метаболических цепей внутри клеток даже только человека, является пока лишь мечтой. Задачи протеомики, в этом плане, гораздо сложнее и реализация их будет проходить медленнее, чем в геномике. Поэтому на первое место выступает так называемое медицинское приложение протеомики. Сравнение опухолевых и нормальных клеток по их белковым профилям, сравнение клеток до и после определенных воздействий, например, после холодового и теплового шоков, использование биологических

жидкостей в диагностических целях, представляет само по себе огромный интерес и открывает совершенно новые перспективы как для диагностической медицины, так и для фармацевтической индустрии в плане создания новых диагностических тестов и лекарственных препаратов. Однако, в свою очередь, задача эта даже более сложна, чем поиск иголки в стоге сена. В последнем случае поиск облегчен тем, что вы разыскиваете предмет, сделанный из другого материала, но сравнимого с его окружающим. В случае поиска интересующих нас белков они сделаны из тех же 20 аминокислот, что и все остальные белки, составляющие 99,999% протеомной карты. И, несмотря на то, что существует лишь один метод – метод масс – спектрометрии, позволяющий решать эту задачу в научном плане, она представляет наибольшую важность. В 2000 году в ее решение включилась хорошо известная фирма “Celera Genomics”, вложившая 1 млрд. долларов в организацию исследовательского центра протеомики. Конкуренция в этой области, где только в настоящее время уже работают 10 ключевых хорошо оснащенных фирм (табл.), определяющих уровень протеомных исследований, настолько высока, что не приходится сомневаться в бурном прогрессе и

Таблица. Ключевые фирмы в области протеомики

Компания	Местонахождение	Область исследования
Celera	Роквилл, Мэриленд	база данных
Incyte Pharmaceuticals	Пало Альто, Калифорния	база данных
GeneBio	Женева, Швейцария	база данных
Proteome Inc.	Бeverly, Массачусетс	база данных
PE Biosystems	Фрамингем, Массачусетс	приборы
Ciphergen Biosystems	Пало Альто, Калифорния	белковые массивы
Oxford GlycoSciences	Оксфорд, Англия	2D гель-МС*
Protana	Оденза, Дания	2D гель-МС
Genomic Solutions	Ann Arbor, Мичиган	2D гель-МС
Large Scale Proteomics Corp.	Роквилл, Мэриленд	2D гель-МС

Примечание: * 2D гелиевый электрофорез и масс-спектрометрия

блестящих достижениях протеомики в ближайшем будущем. И, наконец, последнее, что показал конгресс и, может быть это одна из наиболее важных примет времени, заключается в том, что произошло существенное перераспределение сил в науке между фирмами и государственными научными учреждениями. В наиболее бурно развивающихся областях науки о жизни, таких как геномика и протеомика, ключевую роль стали играть коммерческие структуры. Лишь биоинформатика по-прежнему сохранила свои ведущие позиции в рамках государственных структур.

ЛИТЕРАТУРА

1. *Anderson L., Seihamer J.* (1997) Electrophoresis **18**, 533-537
2. *Gygi S.P. et al.* (1999) Mol. Cell Biol. **19**, 1720-1730
3. *Futcher B. et al.* (1999) Mol. Cell Biol. **19**, 7357-7368
4. *Gygi S.P., Aebersold R.* (2000) Proteomics: A Trends Guide. 31-36
5. *Blackstock W.P., Weir M.P.* (1999) Trends Biotech. **17**, 121-126
6. *O'Farrell P.H.* (1975) Journal Biol. Chem. **250**, 4007 – 4021
7. *Lopez M.E.* (2000) Electrophoresis. **21**, 1082-1093
8. *Blackstock W.P.* (2000) Proteomics: A Trends Guide. 12-17
9. *Field F.D., Caldwell J.A., Shabanovitz J., Hunt D.F.* (2000) The Biochemist. **22**, 16-19
10. *Jenson O.N.* (2000) Proteomics: A Trends Guide. 36-42
11. *Gorg A. et al.* (1988) Electrophoresis. **9**, 531-546
12. *Gorg A.* (1991) Nature. **349**, 545-546
13. *Gorg A.* (2000) Proteomics: A Trends Guide. 3-6
14. *Grouford M.E., Cusick M.E., Garrels J.I.* (2000) Proteomics: A Trends Guide. 17-21
15. *Service R.F.* (2000) Science. **287**, 2136-2138
16. *Clark M.D. et al.* (1999) Methods Enzymol. **303**, 205-233
17. *Cahill D.J.* (2000) Proteomics: A Trends Guide. 47-51
18. *Brem R.* (2000) Modern Drug Discovery. 87-88
19. *Lueking A. et al.* (1999) Anal. Biochem. **270**, 103-111
20. *Page M.J., Amess B., Rohlf Ch., Stubberfield C., Parekh R.* (1999) Drug Discovery Today. **4**, 55-62

Поступила 28. 07. 2000.