

# Секвенирование – методы расшифровки нуклеотидной последовательности ДНК. Сообщение 1

Костюк С.А.

Белорусская медицинская академия последиplomного образования, Минск

Kostiuk S.A.

Belarusian Medical Academy of Post-Graduate Education, Minsk

## Sequencing – methods of DNA nucleotide sequence decryption. Report 1

**Резюме.** Описаны методы, позволяющие идентифицировать нуклеотидную последовательность ДНК. Подробно рассмотрены свойства флуорофоров и полимераз, используемых в секвенировании ДНК, представлены особенности секвенирующего гель-электрофореза.

**Ключевые слова:** секвенирование, ферментативный метод, метод химической деградации, флуорофор, секвенатор.

Медицинские новости. – 2017. – №6. – С. 18–21.

**Summary.** Methods of DNA nucleotide sequence identification are described in this article. Characteristics of phosphors and polymerases using in DNA sequencing are contemplated in detail, gel electrophoresis peculiarities are described.

**Keywords:** sequencing, enzymatic method, method of chemical degradation, fluorophor, sequenator.

Meditsinskie novosti. – 2017. – №6. – P. 18–21.

В 50-е годы прошлого века были разработаны методы, позволяющие определять последовательность аминокислот в полипептидной цепи. Теоретически это несложно, поскольку все аминокислоты, встречающиеся в природных белках, имеют разные свойства. После расшифровки генетического кода появилась возможность восстанавливать нуклеотидную последовательность транскрибируемой ДНК по аминокислотной последовательности соответствующего белка. Однако генетический код является вырожденным, следовательно, первичная структура ДНК, полученная на основе анализа последовательности аминокислот, не является однозначной [5, 14, 16].

Безусловно, определение нуклеотидной последовательности ДНК крайне важно для множества фундаментальных и прикладных задач. Особое место оно занимает в науке: для анализа результатов секвенирования геномов была, фактически, создана новая наука – биоинформатика. Секвенирование сейчас используют молекулярные биологи, генетики, биохимики, микробиологи, ботаники, зоологи, и, конечно же, эволюционисты: практически вся современная систематика основана на его результатах. Секвенирование широко применяется в медицине как метод поиска наследственных заболеваний и изучения инфекций [3, 8].

В настоящее время определение точной нуклеотидной последовательности любого сегмента ДНК умеренной длины – вполне разрешимая задача. Уже определены последовательности нескольких сотен генов про- и эукариот. Зная последовательность гена и генетический

код, легко определить аминокислотную последовательность кодируемого им белка. Раньше для определения структуры белка приходилось делать тщательный и весьма трудоемкий анализ выделенного и очищенного белка, в то время как сейчас бывает проще определить структуру белка через нуклеотидную последовательность, чем с помощью прямого секвенирования. Если секвенирование белка занимает месяцы и даже годы, то ДНК удается секвенировать за несколько недель. Так в 2000 году был секвенирован геном человека, однако в данном случае речь идет только об установлении последовательности нуклеотидов, так как генетическая структура и функции отдельных участков генома еще не идентифицированы, что представляет собой более сложную задачу: прочесть – не значит понять [2, 9, 11, 15].

### С чего все начиналось?

К концу 60-х годов прошлого века Ф. Сэнгером был разработан метод секвенирования РНК, получаемой с ДНК-матрицы при помощи РНК-полимеразы [13]. Применив этот способ, Ш. Вейсман и У. Фирс смогли к концу 1976 года определить последовательность более половины молекулы ДНК SV40, длина которой превышает 5200 нуклеотидных пар [1, 2]. Следующим шагом должна была стать разработка методов прямого секвенирования ДНК.

Первым методом прямого ферментативного секвенирования ДНК стал метод, предложенный Ф. Сэнгером и Д. Коулсоном в 1975 году («плюс–минус» метод). В качестве матрицы в реакции полимеразного копирования использовался одноцепочеч-

ный фрагмент ДНК, в качестве праймеров – синтетические олигонуклеотиды или природные субфрагменты, получаемые при гидролизе рестрикционными эндонуклеазами, а в качестве фермента – фрагмент Кленова ДНК полимеразы I (Poll) из *Escherichia coli* [14].

Метод включал два этапа. Сначала в ограниченных условиях проводили полимеразную реакцию в присутствии всех четырех типов дезоксирибонуклеозидтрифосфатов (dNTP) (один из них был мечен по  $\alpha$ -положению фосфата), получая на выходе набор продуктов неполного копирования матричного фрагмента. Смесь очищали от несвязавшихся dNTP и делили на восемь частей. После чего в «плюс»-системе проводили четыре реакции в присутствии каждого из четырех типов нуклеотидов, в «минус»-системе – в отсутствие каждого из них. В результате в «минус»-системе терминация происходила перед dNTP данного типа, а в «плюс»-системе – после него (рис. 1).

Полученные таким образом восемь образцов разделяли с помощью электрофореза, «считывали» сигнал и определяли последовательность исходной ДНК. Этим способом была секвенирована короткая ДНК фага фХ174, состоящая из 5386 нуклеотидных пар [13].

В 1977 году Ф. Сэнгер предложил еще один способ ферментативного секвенирования, получивший название метода терминирующих аналогов трифосфатов (метод терминаторов). Этот способ, более мощный и технологичный, несколько модифицированный, применяется до сих пор [13].

Рисунок 1 Секвенирование ДНК ферментативным методом по Ф. Сэнгеру («плюс-минус» метод)

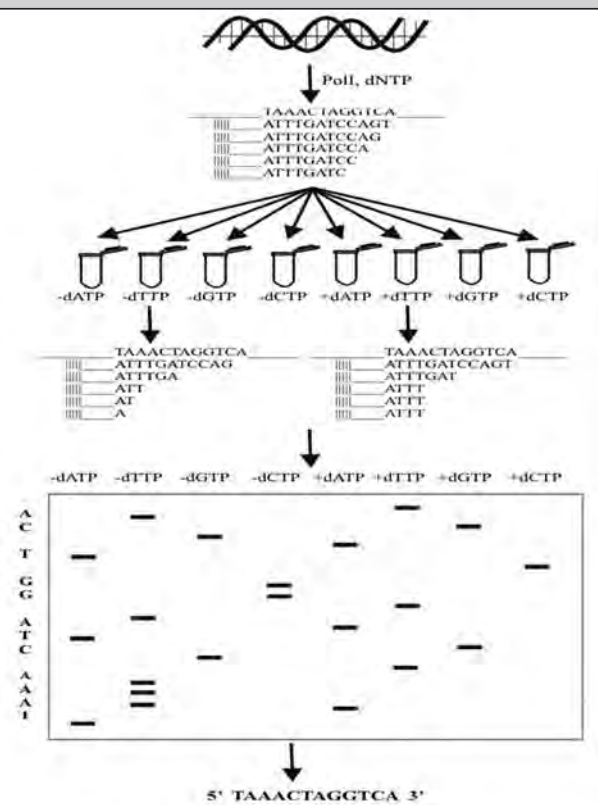
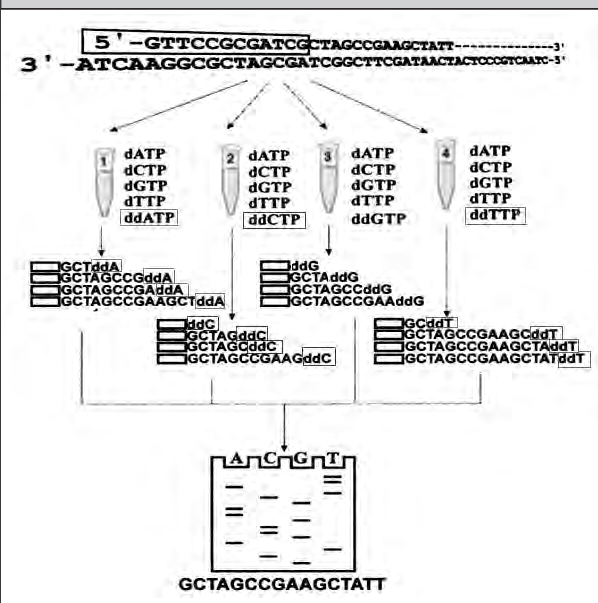


Рисунок 2 Секвенирование ДНК ферментативным методом по Ф. Сэнгеру (метод терминаторов)



В основе метода также лежало ферментативное копирование с помощью фрагмента Кленова ДНК полимеразы I из *Escherichia coli*. В качестве праймеров использовали синтетические олигонуклеотиды. Специфическую терминацию синтеза обеспечивали добавлением в реакционную смесь помимо четырех

типов dNTP (один из которых был радиоактивно мечен по  $\alpha$ -положению фосфата) еще и одного из 2',3'-дидезоксирибонуклеотидтрифосфатов (ddATP, ddTTP, ddCTP или ddGTP), который способен вклучаться в растущую цепь ДНК, но не обеспечивает дальнейшее копирование из-за отсутствия 3'-ОН группы. Отношение концентраций dNTP/ddNTP авторы подбирали экспериментально, так, чтобы в итоге получить набор копий ДНК различной длины.

Таким образом, для определения первичной структуры исследуемого фрагмента ДНК требовалось провести четыре реакции копирования: по одному типу терминаторов в каждой из реакций. После этого полученные продукты разгонялись в полиакриламидном геле на соседних дорожках и по расположению полос определялась последовательность нуклеотидов (рис. 2) [14].

На выходе из сэнгеровского секвенатора получаются короткие участки ДНК, так называемые риды (reads). Для биоинформатики принципиальны две вещи: во-первых, какой длины получаются риды, во-вторых, какие в них могут быть ошибки и как часто. Сэнгеровские риды по этим критериям очень хороши: длина – около тысячи нуклеотидов, причем качество начинает заметно падать только после 700–800 нуклеотидов.

Сам процесс секвенирования по Ф. Сэнгеру предопределяет и эффект падения качества: труднее отличить молекулу массой 700 от таковой с массой 701, чем массой 5 от 6. Другой неприятный эффект: если в геноме встречается длинная последовательность, состоящая из одного вида нуклеотида, то трудно бывает точно определить, какой она длины (все промежуточные массы попадут в одну и ту же пробирку, не-

которые из них могут не встретиться, слиться друг с другом и т. д.). Но все же сэнгеровское секвенирование дает отличные результаты с достаточно длинными ридами.

Дешевизна, точность, а также сравнительная простота автоматизации делает этот метод своеобразным «золотым стандартом» определения последовательности нуклеотидных остатков ДНК. Именно при помощи сэнгеровского секвенирования был впервые расшифрован геном человека. Секвенирование по Ф. Сэнгеру применяется и сегодня, но его все активнее вытесняют другие методы.

В 1976 году А. Максамом и У. Гилбертом был разработан метод секвенирования, основанный на специфической химической деградации фрагмента ДНК, радиоактивно меченого с одного конца – метод химической деградации [6]. Препарат меченой ДНК разделяли на четыре аликвоты и каждую обрабатывали реагентом, модифицирующим одно или два из четырех оснований. Авторы предложили модифицировать пуриновые основания диметилсульфатом. При этом происходит метилирование адениновых остатков по азоту в положении 3, гуаниновых – по азоту в положении 7. Обработка образца ДНК соляной кислотой при 0°C приводит к выщеплению метиладенина. Последующая инкубация при температуре 90°C в щелочной среде вызывает разрыв сахарно-фосфатной цепи ДНК в местах выщепления оснований. Обработка пиперидином приводит к гидролизу образца по остаткам метилгуанина.

Пиримидиновые основания модифицируют гидразином. Если реакцию вести в бессолеватой среде, то модифицируются как цитозин, так и тимидин; если обработку вести в присутствии 2M NaCl, то модифицируется лишь цитозин. Расщепление цепи ДНК на фрагменты и в этом случае осуществляется пиперидином. Условия реакции авторы подбирали таким образом, чтобы в итоге получить полный набор субфрагментов разной длины. Последующий электрофорез в полиакриламидном геле позволяет восстановить полную структуру исследуемого фрагмента (рис. 3) [6].

Важно отметить, что научное сообщество высоко оценило данные разработки, и в 1980 году Нобелевская премия по химии была присуждена У. Гилберту – за разработку методов секвенирования ДНК путем химической деградации, Ф. Сэнгеру – за разработку методов секвенирования ДНК путем ферментативного построения.

### Автоматическое секвенирование ДНК

В основе автоматического секвенирования лежит уже упоминавшийся выше метод ферментативного секвенирования с использованием терминирующих ddNTP. Как и классический вариант метода Ф. Сэнгера, автоматическое секвенирование включает две стадии: проведение терминирующих реакций и разделение продуктов этих реакций с помощью электрофореза. Как правило, автоматизирована лишь вторая стадия, то есть разделение меченых фрагментов ДНК в ПААГ, получение спектра эмиссии флуорофоров и последующий подсчет собранных данных [4, 10, 12].

Современные автоматизированные секвенаторы разделяют эти фрагменты, пропуская всю смесь через тончайшие капилляры, наполненные гелем. Чем короче фрагмент, тем быстрее он движется в геле по капилляру под действием электрического поля, поскольку фрагменты ДНК – по сути, ионы, движущиеся в электрическом поле от «минуса» к «плюсу». Процесс, называемый капиллярным электрофорезом, настолько эффективен, что фрагмент, только что вышедший из капилляра, оказывается ровно на один нуклеотид длиннее, чем предшествующий ему. По мере того как фрагмент появляется, он освещается лазером, что заставляет светиться меченый нуклеотид на его конце. Компьютер определяет разновидность этих нуклеотидов и регистрирует последовательность их появления, складывая «буквы» (нуклеотиды) в «текст» (последовательность ДНК). В случае расшифровки целого генома так набираются миллиарды коротких «текстов», которые поступают в специальную программу, запускаемую на суперкомпьютерах. Программа находит места перекрытия «текстов» и, располагая их в нужном порядке, выстраивает полную последовательность генома.

Автоматическое секвенирование идеологически отличается от современного ему ручного секвенирования только типом используемой метки. Флуоресцентную метку включают либо в праймер, либо в терминатор транскрипции согласно следующим схемам: меченый праймер (четыре разных красителя) и немеченые терминаторы; меченый праймер (один краситель) и немеченые терминаторы; меченые терминаторы (каждый тип терминатора своим красителем) и немеченый праймер. Использование меченых праймеров предполагает проведение четырех независимых реакций (отдельно с каждым из терминаторов) для каждого секвенируемого образца. Использование

меченых терминаторов позволяет совместить все четыре реакции в одной пробирке. Если используется единственный краситель, то разделение продуктов секвенсированной реакции в геле проводят на четырех разных дорожках. Использование четырех разных красок позволяет разгонять продукты реакции на одной дорожке [7, 17].

Флуоресценция – это излучение, сопровождающее переход электрона из возбужденного состояния в основное без изменения мультиплетности (то есть переход синглет – синглет, либо триплет – триплет). Типичное время излучения для флуоресценции – 10<sup>-8</sup>с. Флуоресценция должна предшествовать обратный процесс – переход электрона в возбужденное состояние в результате поглощения (абсорбции) кванта света. Для флуоресценции доказана справедливость следующих правил: спектр испускания (эмиссии) не зависит от длины волны возбуждения; испускание сдвинуто относительно поглощения в сторону больших длин волн из-за энергетических потерь (сдвиг Стокса); спектр испускания представляет собой зеркальное отражение спектра поглощения. Для измерения флуоресценции используют время затухания флуоресценции и квантовый выход флуоресценции (отношение числа испущенных фотонов к числу поглощенных) [8].

Спектр абсорбции, равно как и спектр эмиссии, зависят от химической структуры флуорофора, а также от условий в которые молекула флуорофора помещена (рН, температура, среда и т.д.). В автоматическом секвенировании используют флуорофоры, абсорбция и излучение у которых происходит в диапазоне длин волн 450–650 нм (видимая область спектра; секвенаторы ABI и Pharmacia) и 650–825 нм (ближняя инфракрасная область спектра; секвенаторы фирмы LI-COR).

Рисунок 3 Секвенирование ДНК методом химической деградации по Максему – Гилберту

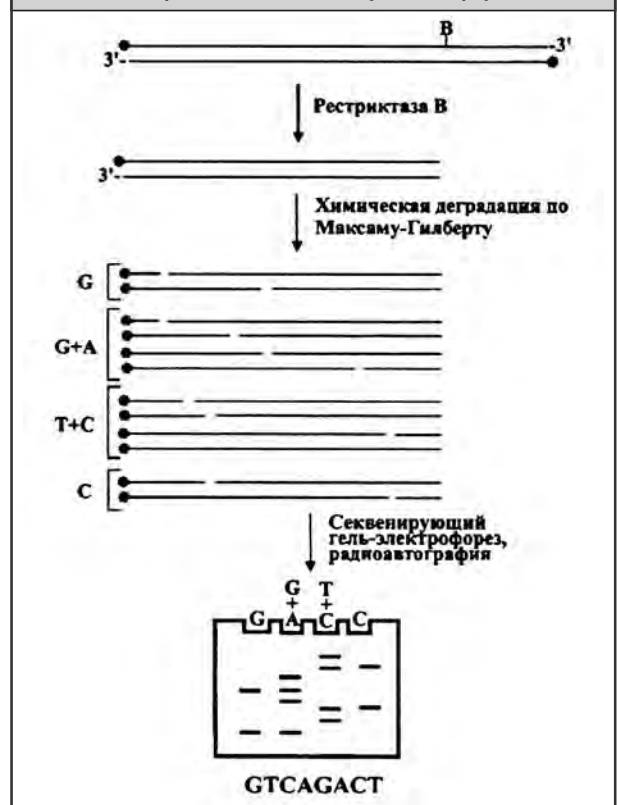
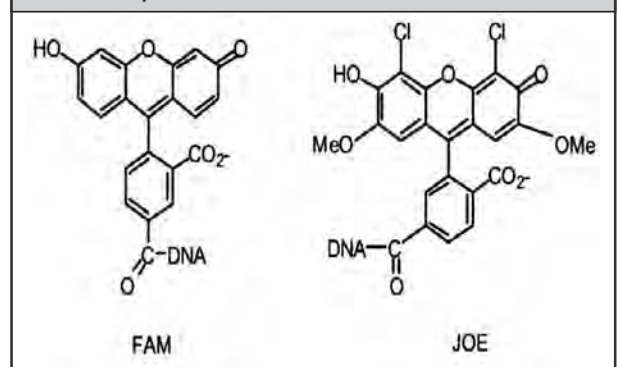
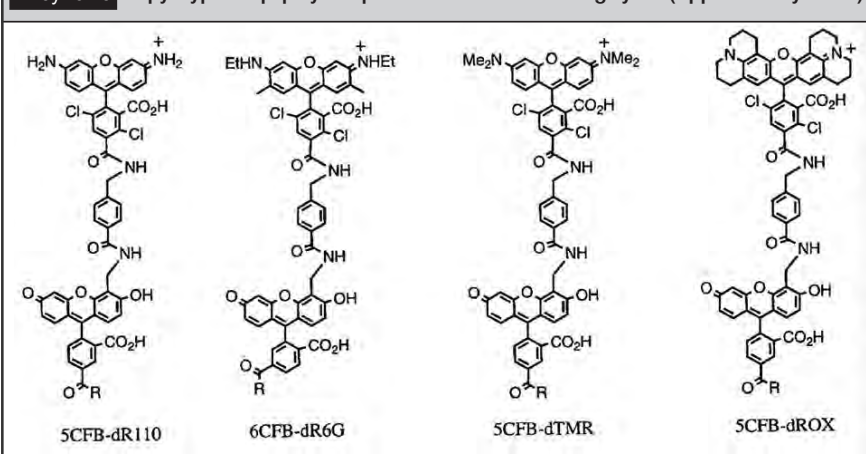


Рисунок 4 Структурные формулы флуоресцеиновых красителей FAM и JOE



К настоящему времени синтезировано большое число разнообразных флуоресцентных красителей и постоянно продолжается работа над новыми, с улучшенными характеристиками. Помимо высокого квантового выхода флуоресценции красители должны обладать еще рядом свойств. Присоединение молекулы флуорофора способно изменять подвижность меченого фрагмента ДНК в геле. Поэтому, если используются одновременно несколько красителей, то необходимо, чтобы влияние каждого из них на подвижность было либо минимальным, либо одинаковым у всех. Выравнивание электрофоретической подвижности (когда это нужно) проводят

Рисунок 5 Структурные формулы красителей семейства BigDye™ (Applied Biosystems)



с помощью линкерных молекул, встраиваемых между красителем и, например, праймером. Еще одним важным моментом является условие минимального перекрытия спектров эмиссии. К сожалению, полностью избежать перекрытия спектров до сих пор не удалось. В качестве альтернативы синтезу все новых соединений недавно был предложен подход, основанный на определении не спектра, а времени флуоресценции. Правда, какого-либо практического выхода эта идея пока не получила [1, 3, 7, 10, 14].

Первыми флуорофорами, адаптированными к нуждам секвенирования, стали соединения из семейства флуоресцеиновых (FAM, JOE) (рис. 4) и родаминовых (TAMRA, ROX, R110, R6G) красителей.

Следующее поколение флуорофоров этого семейства (dTAMRA, dROX, dR110, dR6G) получило довесок из двух остатков хлора. Это позволило несколько снизить перекрытие спектров испускания и значительно повысить интенсивность флуоресценции, значит, и чувствительность. Еще более высоким выходом флуоресценции характеризуются «трехкомпонентные» красители класса BigDye™ (Applied Biosystems) и DYEamic™ ET (Amersham-Pharmacia-Biotech), при конструировании которых был использован принцип переноса энергии. Под переносом энергии понимают явление безизлучательного переноса энергии возбужденного состояния от донора к акцептору. Донором в BigDye™ является 4'-аминометил-5(или 6)-карбоксифлуоресцеин (5CF или 6CF), который связан с акцептором (представителем семейства d-родаминов) через остаток 4-аминометилбензойной кислоты (рис. 5).

Все перечисленные красители флуоресцируют в видимой области спектра. В секвенаторах, детектирующих флуоресценцию в инфракрасной области спектра, используются красители цианинового

ряда с рабочим диапазоном 650–715 нм и 765–825 нм [1, 2, 9, 12, 16, 17].

#### Полимеразы, используемые при секвенировании ДНК

В оригинальной работе Ф. Сэнгера для проведения сиквенсовых реакций был использован Кленовский фрагмент ДНК-полимеразы I из *E. coli*. В настоящее время для секвенирования используют рекомбинантные ДНК-полимеразы, отвечающие следующим требованиям: отсутствие 3'- и 5'-экзонуклеазной активности, отсутствие дискриминации по включению в растущую цепь как обычных dNTP, так и модифицированных (меченных) ddNTP. Выбор конкретной полимеразы зависит от условий проведения сиквенсовой реакции. Существует два разных подхода. В первом случае коопирование осуществляется при 37°C высокопроцессивными термостабильными полимеразми (например, T7 DNA polymerase - Sequenase v1.0 и v2.0, Amersham Pharmacia Biotech). Во втором – реализуется циклический процесс, который включает денатурацию, отжиг и элонгацию, и предполагает использование термостабильных полимераз (например, ThermoSequenase™, Amersham Pharmacia Biotech и AmpliTaq FS™, PE Biosystems) [1, 2, 5, 8, 10, 15].

В настоящее время нет ни одного метода секвенирования, который работал бы для молекулы ДНК целиком; все они устроены так: сначала готовится большое число небольших участков ДНК (клонировается молекула ДНК многократно и «разрезается» в случайных местах), а потом читается каждый участок по отдельности.

#### Секвенирующий гель и электрофорез

Полученные в реакции секвенирования радиоактивно/флуоресцентно меченые одноцепочечные фрагменты ДНК разделяются с помощью электрофореза в полиакриламидном геле. Гели, используемые в секвенировании, должны уметь разделять фрагменты, отличаю-

щиеся друг от друга на один нуклеотид в широком диапазоне длин. Разделение должно проходить в денатурирующих условиях, препятствующих ренатурации и возникновению вторичных структур у разделяемых фрагментов. Этим требованиям удовлетворяют 5–8% полиакриламидных гелей, содержащих 7М мочевины. Обычно электрофорез проводится в трис-боратном буфере (89 мМ трис-НСl, 8,9 мМ борной кислоты, 2 мМ ЭДТА, рН 8,0–8,5).

Для эффективного разделения напряжение должно составлять 30–50 В на см длины геля. Важным условием при проведении электрофореза является однородность температуры по всей поверхности геля. Неравномерный нагрев геля и стекол во время электрофореза способен привести к разнице в скоростях движения фрагментов ДНК по ширине геля и, как следствие, к искажению результатов. Для борьбы с этим явлением используют тонкие (обычно 0,4–0,1 мм) гели и принудительный подогрев стекол.

В автоматическом секвенировании помимо электрофореза в пластинах полиакриламида чрезвычайно популярен капиллярный электрофорез в линейном полиакриламиде. Капилляры представляют собой стеклянную трубку длиной 30–100 см, закатанную в полимерный пластификатор. Небольшой диаметр капилляра (50–100 мкм) позволяет проводить разделение значительно быстрее, чем в обычных гелях. Кроме того, капиллярные секвенаторы позволяют обеспечивать гораздо более высокую чувствительность за счет отсутствия горизонтальной диффузии [1, 5, 7, 8, 9, 17].

#### ЛИТЕРАТУРА

1. Костюк С.А. Молекулярно-биологические методы в медицине: Монография. – Минск, 2013. – 327 с.
2. Bergot B.J., Chakerian V., Connell C.R., et al. – Bioconjug Chem. – 1999. – Vol.58. – P.313–327.
3. Fiers W., Contreras R., Haegemann G., et al. – Nature. – 1978. – Vol.273. – P.113–120.
4. Ju J., Ruan C., Fuller C.W., et al. // Anal. Chem. USA. – 1995. – Vol.92, N10. – P.4347–4351.
5. Lee L.G., Spurgeon S.L., Heiner C.R., et al. // Nucleic Acids Res. – 1997. – Vol.25. – P.2816–2822.
6. Maxam A.M., Gilbert W. // Proc. Natl. Acad. Sci. USA. – 1977. – Vol.74. – P.560–564.
7. Menchen S.M., Lee L.G., Connell C.R., et al. // Nucleic Acids Res. – 1993. – Vol.51. – P.216–242.
8. Mujumdar R.B., Ernst L.A., Mujumdar S.R., et al. // Bioconjug Chem. – 1993. – Vol.4. – P.105–121.
9. Narayanan N., Little G., Lugade A., et al. – Nature. – 1998. – Vol.128. – P.141–168.
10. Nunnally B.K., He H., Li L.C., et al. // Anal. Chem. USA. – 1997. – Vol.69, N13. – P.2392–2407.
11. Reddy V.B., Thimmappaya B., Dhar R., et al. // Science. – 1998. – Vol.200, N4341. – P.494–522.
12. Rosenblum B.B., Lee L.G., Spurgeon S.L., et al. // Nucleic Acids Res. – 1997. – Vol.25, N22. – P.4500–4504.
13. Sanger F., Coulson A.R. // J. Mol. Biol. – 1975. – Vol.94. – P.444–448.
14. Sanger F., Niclein S., Coulson A.R. // Proc. Natl. Acad. Sci. USA. – 1977. – Vol.74. – P.5463–5467.
15. Shealy D.B., Lipowska M., Lipowski J., et al. // Anal. Chem. – 1995. – Vol.67. – P.247–251.
16. Tabor S., Richardson C.C. // PNAS USA. – 1995. – Vol.92, N14. – P.6339–6343.
17. Tu O., Knott T., Marsh M., et al. // Nucleic Acids Res. – 1998. – Vol.26, N11. – P.2797–2802.

Поступила 10.03.2017 г.