



*Практикум по дисциплине  
"Генные конструкции организма  
человека и животных"*

*Макаров Я.М.*

*САНКТ ПЕТЕРБУРГ 2015*

*«В природе все мудро продумано и устроено, всяк должен заниматься своим делом, и в этой мудрости — высшая справедливость жизни»*

*Леонардо да Винчи*

## СОДЕРЖАНИЕ

Введение.....	2
Молекулярное клонирование.....	2
Теория работы с генным материалом .....	3
<i>Практическая часть</i> .....	16
<b>Первый блок для анализа</b> .....	19
<b>Второй блок для анализа</b> .....	20
ЗАДАНИЕ .....	22

## Введение

Клонирование - появление естественным путем или получение нескольких генетически идентичных организмов путем бесполого (в том числе вегетативного) размножения. Термин «клонирование» в том же смысле нередко применяют и по отношению к клеткам многоклеточных организмов. Клонированием называют также получение нескольких идентичных копий наследственных молекул (молекулярное клонирование). Наконец, клонированием также часто называют биотехнологические методы, используемые для искусственного получения клонов организмов, клеток или молекул. Группа генетически идентичных организмов или клеток — клон.

Генетическая инженерия (генная инженерия) — совокупность приёмов, методов и технологий получения рекомбинантных РНК и ДНК, выделения генов из организма (клеток), осуществления манипуляций с генами и введения их в другие организмы.

Генетическая инженерия не является наукой в широком смысле, но является инструментом биотехнологии, используя методы таких биологических наук, как молекулярная и клеточная биология, цитология, генетика, микробиология, вирусология.

Понятие генных конструкций формируется на представлениях о информационных процессах синтеза белковых структур. В качестве основы информационных процессов рассматриваются хромосомы - генные наборы.

## Молекулярное клонирование

Как известно, в генах содержится информация-инструкция для синтеза в организме молекул РНК и белков, в том числе ферментов. Чтобы заставить клетку синтезировать новые, необычные для неё вещества, надо чтобы в ней синтезировались соответствующие наборы ферментов. А для этого необходимо или целенаправленно изменить находящиеся в ней гены, или ввести в неё новые, ранее отсутствовавшие гены. Изменения генов в живых клетках — это мутации. Они происходят под действием, например, мутагенов — химических ядов или излучений. Но такие изменения нельзя контролировать или направлять.

Постулируя возможность создавать необходимую генную конструкцию в первую очередь полагают наличие методологии проведения исследований и методов конструирования.

Огромное количество биологических исследований начинается с одного простого действия — в клетку вносится чужеродный генетический материал. Это действие называется молекулярным клонированием. С его помощью можно получить генетически модифицированные организмы, включить и выключить отдельные гены, определить влияние какого-нибудь белка на какой-нибудь сложный процесс и так далее. Можно сказать, что молекулярное клонирование —

это краеугольный камень, основа основ, фундамент, без которого множество замечательных методик было бы неосуществимо.

Однако поместить в клетку «неродную» ДНК не так просто, как кажется на первый взгляд. Это длинный, трудоемкий и многоэтапный процесс. Молекулярному клонированию посвящены целые книги, и, к сожалению, в этом материале невозможно осветить все его тонкости и нюансы — что-нибудь обязательно останется не раскрыто. Тем не менее, в рамках типового задания постараемся изучить основные приемы конструирования клонов, а фактически основные принципы конструирования разных организмов.

Начнем с простого. Имеется клетка, которую задумали модифицировать. Клетка в последующем, после модификации, будет исполнять программу построения клеточного материала по внесенной нами «правке». Типичная процедура «правки» - вставка неторого фрагмента - части генома.

## **Теория работы с генным материалом**

### **Вставка**

Раз собираемся вставлять в клетки какой-то ген, то самый первый, очевидный, шаг, который нам нужно сделать, — этот ген как-нибудь получить, причем желательно в больших количествах (поскольку — увы! — все методики несовершенны, и нам надо приготовиться к тому, что большая часть копий этого гена бесследно пропадет и так и не достигнет цели).

Чужеродный ген, вносимый в клетку, называется геном-вставкой или просто вставкой. Получить его можно несколькими способами.

Во-первых, мы можем просто выделить его из того генома, к которому он принадлежит. Допустим для простоты, что наша вставка — это какой-нибудь ген слона. Тогда нам нужно:

Получить например, образец тканей слона.

- Извлечь из этого образца ДНК ( DNA ).
- Вычленив из этой ДНК интересующий нас ген и получить его в больших количествах (для этого используется ПЦР).

Заметим в скобках, что получение гена с помощью ПЦР возможно, только если мы знаем его нуклеотидную последовательность или хотя бы последовательность его начала и окончания (для того, чтобы можно было синтезировать праймеры). Если же всё это нам неизвестно, то придется сначала анализировать слоновий геном

Во-вторых, вполне возможно, что нужный нам ген уже был выделен из генома слона и присутствует в библиотеке генов. Тогда нашу вставку можно будет получить оттуда (с этим, на самом деле, тоже придется повозиться, но меньше, чем в первом случае).

И наконец, в-третьих, не обязательно использовать в качестве вставки уже существующий ген. Если исследователь собирается работать с каким-нибудь геном, который является плодом его фантазии и не встречается в природе, то он может синтезировать его искусственно.

## **Следующий этап - поиск и создание вектора**

Запускание в клетку «одиноким» вставки, то есть гена самого по себе, безо всякого сопровождения, — дело совершенно бесперспективное. В клетке плавают множество расщепляющих ДНК ферментов (нуклеаз), которые с радостью набросятся на нашу беззащитную вставку и разрежут ее на кусочки. В результате вставка бесславно исчезнет, не успев совершить ничего полезного, а клонирование не приведет ни к каким результатам.

Поэтому, чтобы защитить вставку, ее встраивают в специальное «транспортное средство», которое называется вектором. В самом элементарном случае вектор — это просто последовательность ДНК, в которую вшивается наша вставка и которая помогает ей не пропасть в клетке и выполнить свое предназначение.

Существует несколько видов векторов, но среди исследователей самой большой (и заслуженной) любовью пользуется один из них — плазмиды. С них-то мы и начнем.

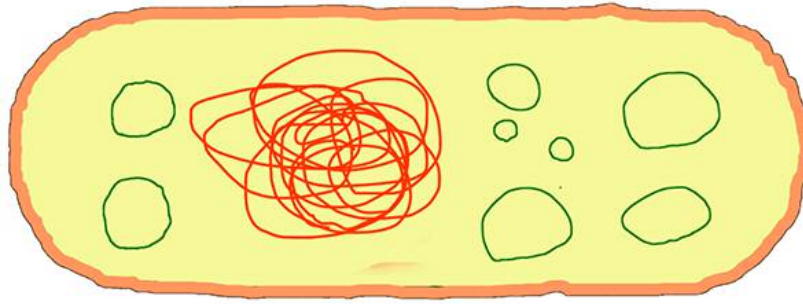


Рис. 1. В бактериальной клетке наряду с бактериальной хромосомой (красная) плавают еще и множество плазмид (зеленые)

### Плазмида

Плазмида — это довольно короткая и обычно кольцевая молекула ДНК, которая плавает в цитоплазме бактериальной клетки. Плазмиды не связаны с бактериальной хромосомой, они могут реплицироваться независимо от нее, могут «выплываться» бактерией в окружающую среду или, наоборот, из этой окружающей среды «проглатываться». С помощью плазмид бактерии обмениваются друг с другом генетической информацией, например передают соседям устойчивость к какому-нибудь антибиотику.

Плазмиды существуют в бактериях в естественных условиях. Но никто не может помешать исследователю искусственно синтезировать плазмиду, которая будет обладать нужными для него свойствами, вшить в нее вставку и запустить в клетку. Плазмида — это, можно сказать, заготовка, болванка; в одну и ту же плазмиду можно вшивать разные вставки. Поэтому плазмиды стараются сделать как можно более универсальными и подходящими для всех случаев жизни.

Для того чтобы из плазмиды получился рабочий вектор, она должна обладать некоторыми важными характеристиками.

### Размножение

Прежде всего, плазмида обязательно должна в клетке размножаться, реплицироваться, потому что иначе она быстро подвергнется деградации, а вместе с ней исчезнет и ген-вставка. Для этого в ней должна быть специальная последовательность под названием точка начала репликации, с которой и начинается удвоение ДНК. У разных видов живых существ эти точки имеют разную нуклеотидную последовательность. Поэтому если мы хотим создать плазмиду, которая бы реплицировалась сразу в двух видах клеток (например, и в дрожжевых, и в бактериальных), то нам надо вставить в нее две точки начала репликации.

### Разрезание

Кроме того, в ДНК плазмиды должны быть участки, в которых ее можно будет разрезать, чтобы вшить туда вставку. В качестве «ножниц» используются особые ферменты под названием рестриктазы. Рестриктазы прекрасны тем, что режут ДНК не где попало, а в строго определенных местах, которые называются сайтами рестрикции (каждая рестриктаза распознает только свой сайт и только в нём — или возле него — разрезает ДНК).

Обычно в плазмиду ставят множество разных сайтов рестрикции, расположенных в разных точках, — благодаря этому ее можно будет разрезать в нужном месте нужной рестриктазой. Участок ДНК, на котором собрано несколько сайтов рестрикции, называется полилинкером

### Пояснение

*Сайт рестрикции (участок узнавания) — короткая последовательность нуклеотидов в молекуле ДНК, которая распознаётся ферментом эндонуклеазой рестрикции-модификации (рестрик-*

тазой). Рестриктаза связывается с молекулой ДНК в точке расположения сайта рестрикции и перерезает цепочку нуклеотидов внутри сайта или в непосредственной близости от него.

Ферменты рестрикции выработаны бактериями в процессе эволюции с целью разрушения чужеродной ДНК, способной проникнуть внутрь клетки и вызвать её трансформацию. Размер сайта рестрикции различных рестриктаз составляет, как правило, 4-6 нуклеотидов. Сайты рестрикции в ДНК самой бактерии замаскированы посредством метилирования остатков А и С

Например, фермент рестрикции *EcoRI* распознаёт симметричную последовательность GAATTC и перерезает цепочку между нуклеотидами G и A, оставляя на концах перекрывающиеся участки AATT

Всего у бактерий различных видов найдено несколько сотен ферментов рестрикции. Ниже на рисунках приведены сайты рестрикции и линии разреза цепочки нуклеотидов для нескольких известных рестриктаз

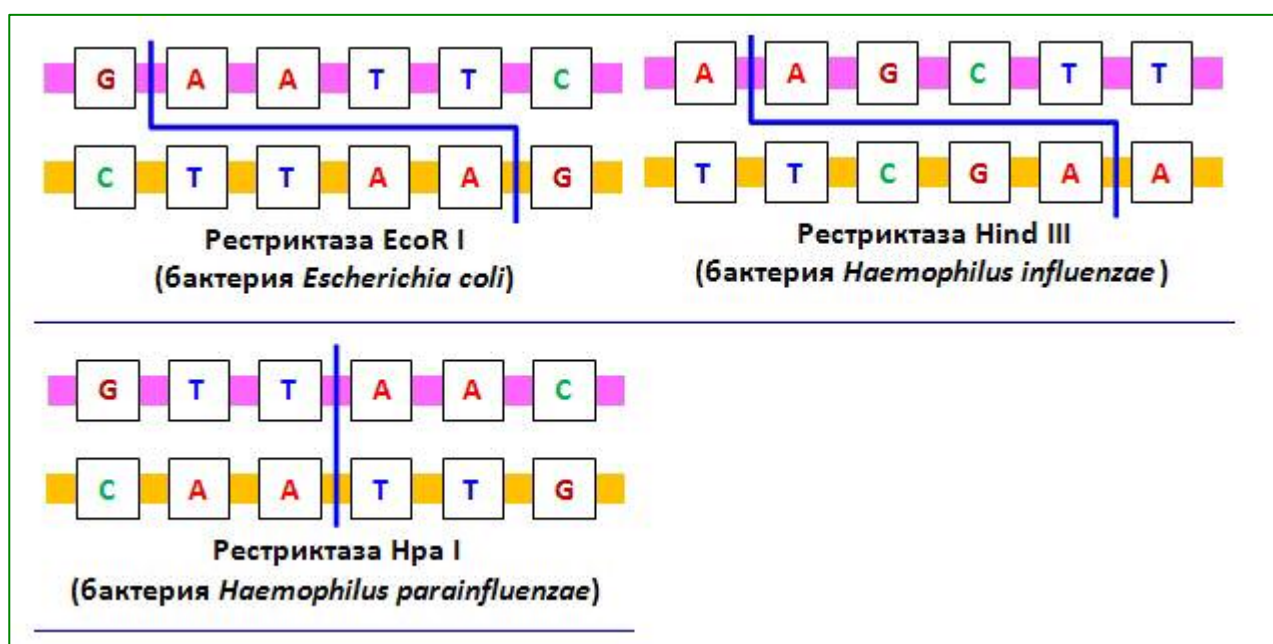


Рисунок 1 Принцип работы рестриктаз

### Селекция

Процесс, при котором бактерия глотает плазмиду, именуется трансформацией. В естественных условиях в каждый момент времени трансформироваться может не вся популяция бактерий, а только ее часть; клетки, которые к этому способны, называются компетентными (см. Competence). Существуют лабораторные методы, с помощью которых можно искусственно увеличить количество компетентных клеток, однако всё равно стопроцентная компетентность для бактериальной культуры — вещь недостижимая.

Так что, добавляя плазмиду к бактериям, мы заранее признаем, что большая часть бактериальных клеток так и останется бесплазмидной, нетрансформированной. Поэтому нам придется отделять зерна от плевел, то есть трансформированные клетки от всех остальных. Для этого используется простой, но остроумный прием.



Допустим, мы встроили в нашу плазмиду ген устойчивости к какому-нибудь антибиотику (такой ген называется селективным маркером). Теперь клетки организма, которые «съели» плазмиду, будут неуязвимы для этого антибиотика и смогут спокойно жить в его присутствии. Это острый способ создавать новые лекарственные формы.

В результате, чтобы выделить из всех бактерий, к которым мы добавили плазмиду, те, которые смогли эту плазмиду использовать по назначению, нам достаточно будет добавить к бактериальной культуре соответствующий антибиотик.

Естественно будем наблюдать за процессом. Те клетки, которые нам нужны, смогут существовать и делиться в присутствии этого антибиотика, а остальные этого делать не смогут.

Но это — только начало селекции. Дело в том, что когда мы вшиваем в плазмиду вставку, то помимо желаемого сочетания (вставка внутри плазмиды) возникает множество побочных продуктов. Поэтому, если клетки устойчивы к антибиотику, это еще не говорит о том, что в них «сидит» нужный нам вектор. Вполне возможно, что они съели просто пустую плазмиду, от которой нам не будет никакого проку.

Поэтому нам нужно провести еще один этап селекции, а для этого в плазмиде должен быть еще один селективный маркер. Это может быть, например, ген устойчивости к какому-нибудь другому антибиотику.

Тонкость тут в том, что наш сайт рестрикции будет расположен посередине этого гена. Если плазида пустая, без вставки, то этот ген будет рабочим, и устойчивость к антибиотику сохранится. Если же в середину этого гена ухитрилась вклиниться вставка, то ген будет испорчен, устойчивость к антибиотику нарушится, и рост бактерий в его присутствии будет угнетаться.

То есть всё, что нам нужно сделать после первого этапа селекции, — просто перенести наши клетки на среду с антибиотиком для второго селективного маркера и в этот раз выбрать те колонии, которые не растут.

Существуют и другие способы провести селекцию. Можно, например, поместить сайт рестрикции не внутрь гена антибиотика, а внутрь какого-нибудь «заметного» гена (скажем, такого, в присутствии которого бактериальные культуры меняют цвет). В результате можно будет отличить нужные колонии от ненужных просто на глаз, безо всяких манипуляций. По такому принципу работает, например, очень модная сейчас система бело-голубой селекции

В принципе, можно обойтись и без второго этапа селекции, а просто проанализировать (например, с помощью ПЦР или электрофореза) плазмидный состав нескольких бактериальных клонов и выбрать тот клон, в котором «сидит» плазида с геном-вставкой. Эти важные «детали» рабочего процесса надо просто понимать, иначе будет много ошибок. Здесь имеется ввиду тот факт, что ошибочно созданная модификация гена, по законам организма приступит к массовому тиражированию внутри организма. Если процесс запуска был хоть немного с ошибкой, придется срочно все исправлять. Это будет сделать сложно.

Если мы собираемся работать только на бактериях, то всем вышесказанным дело и ограничится. Однако если конечная наша цель — поместить вектор в какие-нибудь другие, эукариотические, клетки, например клетки млекопитающих, то нам предстоит еще один этап селекции.

Дело в том, что в большинстве эукариотических клеток плазмиды живут недолго и быстро подвергаются деградации. Поэтому, даже если мы заставили клетку «съесть» плазмиду, не стоит питать надежды на то, что наша вставка теперь останется в этой клетке навсегда.

Скорее всего, она успеет только немного поэкспрессироваться, прежде чем содержащий ее вектор будет пойман нуклеазой и разрезан на кусочки. Однако если вектор случайно смог встроиться в геном (это событие очень редкое, но не невероятное), то наша вставка, можно сказать, пустит в этой клетке корни — причем не только в этой клетке, но и во всех ее потомках. И для того, чтобы выделить из всех клеток те, которые имеют вектор в своем геноме, нам понадобится еще один селективный маркер — ген устойчивости к какому-нибудь эукариотическому антибиотику (потому что бактериальные антибиотики, как правило, на клетки эукариот не действуют).

Добавив соответствующий антибиотик (например, генетицин) к среде, в которой культивируются клетки, мы через некоторое время получим популяцию только тех клеток, в геноме которых «сидит» наш вектор.

Кроме того, можно вставить в нашу плазмиду в качестве маркера еще и какой-нибудь ген, обладающий яркой индивидуальностью (скажем, ген зеленого флуоресцентного белка GFP). Этот белок светится под флуоресцентным микроскопом, и если клетка съела плазмиду, то она будет видна под микроскопом как светящееся пятно. А в бесплазмидных клетках никакого свечения наблюдаться не будет.

## **Промоторы, энхансеры, сайленсеры**

Перед каждым рабочим геном находится короткий участок ДНК под названием промо-тор. Именно сюда прикрепляется фермент по имени РНК полимераза, который синтезирует РНК на матрице ДНК, что является первым и абсолютно необходимым этапом в экспрессии гена. Если у гена нет промотора, его экспрессию запустить невозможно, он так и остается молчащим, и хотя и присутствует в клетке, но никак себя не проявляет. Можно сказать, что ген без промотора — это всё равно что машина без педали газа. Поэтому в нашей плазмиде обязательно должен быть хотя бы один промоторный участок, под контроль которого можно будет поставить ген-вставку.

### **А промоторы бывают разные.**

Во-первых, они различаются по своей силе. Некоторые вызывают бурную транскрипцию подконтрольного гена, другие — совсем вялую.

Во-вторых, у прокариот и эукариот промоторы отличаются. Прокариотические промоторы не работают в эукариотических клетках и наоборот. Поэтому будет Ужасной Ошибкой поставить тот ген, который должен, например, экспрессироваться в бактериальных клетках, под эукариотический промотор — это будет всё равно что оставить его без промотора вообще.

В-третьих, у эукариот есть несколько типов РНК-полимеразы — они обеспечивают синтез различных видов РНК. И каждый тип РНК-полимеразы распознает только свои промоторы и «не видит» чужие. Поэтому, в зависимости от того, какую именно РНК кодирует наша вставка (например, матричную или, наоборот, шпилечную, а может, и вовсе рибосомальную), нам нужно подбирать и тип промотора, который мы будем ставить в плазмиду.

И, наконец, в-четвертых, разные промоторы включаются по-разному. Некоторые активны постоянно. Другие активизируются только при определенных условиях — например, при повышении окружающей температуры или появлении в клетке каких-то веществ. К тому же, у многоклеточных организмов в каждой ткани включены одни промоторы и выключены другие. Можно, например, подобрать такой промотор, который будет активен только в нейронах. Или только в нейронах головного мозга. Или только в нейронах головного мозга, относящихся к одному из подкорковых ядер. Или только в крохотной субпопуляции нейронов головного мозга, относящейся к одному из подкорковых ядер. И сужать этот круг можно почти до бесконечности.

А еще у эукариот помимо промоторов есть другие регулирующие экспрессию гена участки ДНК — энхансеры (которые усиливают транскрипцию) и сайленсеры (которые ее ослабляют). Они, в отличие от промоторов, могут находиться не вблизи от подконтрольного гена, а на значительном от него удалении.

Знание всего этого дает исследователю удивительную свободу. Подобрал в плазмиду подходящий промотор (а если нужно, еще и какой-нибудь энхансер), он сможет творить с экспрессией гена-вставки почти всё, что ему заблагорассудится. Ну, скажем, сделать так, чтобы он экспрессировался сильно, только в мышечных клетках и только в ответ на повышение температуры.

## **Трансляция белка**

Помещая вектор в клетку, ученый может рассматривать решение двух задач:

- Чтобы происходила только транскрипция гена-вставки (то есть синтез РНК на матрице ДНК — например, этого достаточно, если в клетку вносится какая-нибудь некодирующая РНК).

- Чтобы происходила и транскрипция, и трансляция гена-вставки (то есть экспрессия кодируемого вставкой белка).

В первом случае вектор называется транскрипционным, во втором — экспрессионным. Экспрессионные векторы обычно немного сложнее транскрипционных, потому что в них присутствуют:

1. Консенсусная последовательность Козак. Это длинное имя носит короткий (примерно в 10 нуклеотидов) фрагмент в самом начале молекулы матричной РНК, который через белки-посредники обеспечивает связывание этой мРНК с рибосомой (без чего, как нетрудно догадаться, синтез белка невозможен). Последовательность Козак характерна только для эукариот, причем у представителей разных видов она немного отличается. Поэтому, создавая экспрессионный вектор, надо вставлять в него последовательность, которая характерна именно для того живого существа, в клетки которого мы собираемся вставлять вектор. Кроме того, последовательность Козак бывает сильной и слабой — то есть приводящей к синтезу большого или малого количества белка. У прокариот роль последовательности Козак выполняет последовательность Шайна–Дальгарно, которая непосредственно (в смысле — без посредников, в отличие от последовательности Козак) соединяется с рибосомой, после чего и начинается синтез белка.
2. Последовательность Козак находится *перед* вставляемым геном. А *после* него должны находиться еще несколько коротких участков, к которым присоединяются белки, выполняющие полиаденилирование (Polyadenylation) — пришивание к концу свежесинтезированной РНК полиаденинового хвоста. Этот хвост выполняет несколько функций, в том числе обеспечивает экспорт РНК в цитоплазму и помогает организации трансляции — то есть, если мы хотим обеспечить синтез белка на основе нашей РНК, нам без него не обойтись.
3. И еще один момент. мРНК, которая служит матрицей для синтеза белка, может быть транскрибирована только и исключительно РНК-полимеразой II типа. Поэтому нам нужно вставить в плазмиду именно тот промотор, который работает с этой РНК-полимеразой.

В свое время это была, пожалуй, самая популярная плаزمида во всем научном мире, а потом она стала основой для множества плазмид нового поколения. В этой плазмиде 1 есть участок начала репликации (*ori*), благодаря которому она может размножаться в клетках бактерии *E. coli*; гены устойчивости к двум антибиотикам — ампициллину (*amp*) и тетрациклину (*tet*); а также множество сайтов рестрикции (на самом деле их больше сорока, но здесь представлены только четыре — *Eco RI*, *Sal I*, *Pst I*, *Bam HI*).

Промоторные участки, к сожалению, не показаны, но они, разумеется, тут тоже есть. Некоторые сайты рестрикции находятся в генах устойчивости к ампициллину или тетрациклину, в результате чего и тот и другой сайт можно использовать в качестве второго селективного маркера.

---

<sup>1</sup> Плаزمида, используемая в бактериях *E. coli* как вектор клонирования. Создана в 1977 году мексиканскими биологами Франциско Боливаром и Раймондом Родригесом.

pBR322 содержит 4361 нуклеотидную пару и состоит из участка репликации *ori* (взят из плазмиды pMB1 гена *ampR*, кодирующего белок, обеспечивающий устойчивость к ампициллину (взят из плазмиды RSF2124) и гена *tetR*, обеспечивающего устойчивость к тетрациклину (взят из плазмиды pSC101).

Плазмида содержит уникальные сайты рестрикции для более чем 40 рестриктаз. 11 из этих 40 сайтов находятся внутри гена *tetR*, причём 2 сайта расположены внутри промотора этого гена. 6 сайтов находятся внутри гена *ampR*. Участок репликации *ori* заимствованный из плазмиды pMB1, сходен по структуре с ColE1. Он кодирует две РНК (RNAI и RNAII) и один белок (Rom или Rop).

Молекулярная масса плазмиды составляет 2,83·10<sup>6</sup> дальтон. Код в базе данных GenBank — J01749



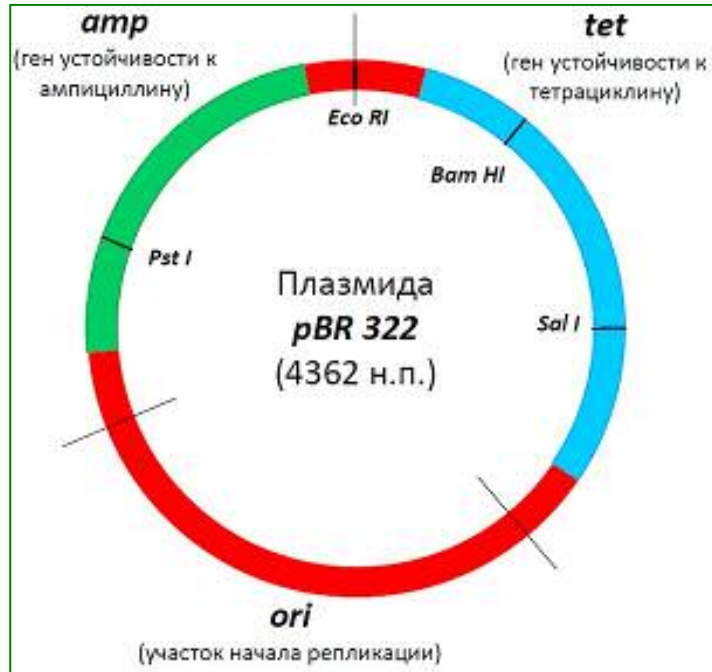


Рис. 2. Структура знаменитой плазмиды PBR322.

Например, если мы разрежем ген устойчивости к ампициллину с помощью рестриктазы *Pst I* и вошьем в это место вставку, то тетрациклин будет первым селективным маркером, ампициллин — вторым, а селекция будет выглядеть так: 1) трансформируем бактерии, выращиваем их на среде с тетрациклином и выбираем только хорошо растущие клоны; 2) переносим эти клоны на среду с ампициллином и выбираем те, рост которых угнетается. Если же мы вошьем вставку внутрь гена устойчивости к тетрациклину (разрезав его с помощью рестриктаз *Bam HI* или *Sal I*), то нам надо будет, наоборот, сначала посадить их на среду с ампициллином, а потом — с тетрациклином.

Итак, мы подобрали все необходимые для плазмиды кусочки. Но мало просто соединить их вместе — огромную роль играет их взаимное расположение. Например, сайты рестрикции должны быть не только многочисленны и разнообразны, но и находиться в «правильных» местах. При этом надо стараться, чтобы итоговая плазида была как можно компактней, поскольку, во-первых, так она будет стабильнее, а во-вторых, охотнее «проглотится» клеткой. Одним словом, вероятно понятно что дизайн хорошей плазмиды — это тонкое и филигранное искусство.

### Плазмидные базы данных

За те несколько десятилетий, что существует методика молекулярного клонирования, были синтезированы тысячи разнообразных плазмид, из которых созданы гигантские базы данных (например, Addgene). В этих базах есть плазмиды на все случаи жизни — с разными типами точек начала репликации, разными полилинкерами, расположенными в разных местах, разными селективными маркерами и промоторами, и так далее. Есть те, в которые можно вшить не одну вставку, а несколько, и есть даже такие, которые уже несут в себе некоторые особенно популярные вставки. Поэтому, как правило, исследователи не синтезируют плазмиду для клонирования самостоятельно, а покупают уже готовую. При необходимости купленную плазмиду можно «довести до ума», вставив или убрав определенные участки (а потом эту модифицированную

плазмиду тоже добавить в базу данных). Иными словами, часто задача ученого сводится просто к тому, чтобы подобрать подходящую плазмиду.

### **Другие векторы**

Плазида — прекрасный вектор для относительно небольших вставок. Если ген-вставка слишком велик, то плазида утрачивает стабильность, потому что ее участки начинают «перетасовываться» друг с другом и теряться при репликации, из-за чего она постепенно укорачивается. Поэтому в качестве вектора для длинных вставок используются более устойчивые конструкции. Например:

- *Космида* — гибрид плазмиды и фага (вируса, который заражает бактерии). По сути дела, это просто плазида, в которую добавлены сайты для связывания с белками оболочки фага (они называются *cos*-сайтами, и именно благодаря им космиды получили свое название). Белковая оболочка делает космиду стабильнее, благодаря чему в нее можно загружать более длинные вставки.
- *Искусственные хромосомы* — это сложные и крупные конструкции, являющиеся, по сути, микрохромосомами. Они относительно стабильные и при этом обладают гигантской емкостью, в них можно вставлять сразу несколько генов. Однако из-за огромных размеров их гораздо труднее поместить в клетку.
- И, наконец, есть еще один вид векторов — *вирусные*. Этот вид настолько важен, что ниже ему будет посвящен целый раздел.

### **Вставляем ген в плазмиду**

Допустим, исследователь подобрал подходящую плазмиду и получил нужную вставку. Теперь нужно соединить одно с другим, чтобы затем засунуть в клетки. Для этого достаточно совершить несколько простых действий.

Как уже говорилось, в плазмиде существует несколько сайтов рестрикции — то есть участков, в которых ее может разрезать нужная рестриктаза. Нам нужно выбрать подходящий сайт, который будет находиться в том месте, куда мы собираемся вшивать вставку, а затем обработать плазмиду соответствующей рестриктазой.

После этого той же рестриктазой нужно обработать вставку, поскольку рестриктазы обычно оставляют выступающие концы на одной из нитей ДНК, и эти концы должны быть совместимы у вставки и плазмиды, чтобы они «согласились» соединиться. Если на кончиках вставки нет нужных сайтов рестрикции, то можно приделать к нему короткие ДНК-фрагменты с нужными сайтами рестрикции на концах.

И наконец, нам нужно соединить в одной пробирке плазмиду и вставку (предварительно очищенные от рестриктаз) и добавить к ним специальный фермент под названием ДНК-лигаза, который умеет лигировать (то есть сшивать воедино) две молекулы ДНК. Конечно, в результате мы получим не только желанный вектор, в котором плазида соединена со вставкой (назовем его чеширским котом с улыбкой), но и целый коктейль побочных продуктов — пустую плазмиду (кота без улыбки), замкнутую вставку (улыбку без кота), несколько сшитых между собой вставок (много улыбок) и так далее. В ходе селекции эти ненужные продукты отсеются, и у нас в руках останется только вектор.

### **Выделяем вектор**

Итак, вначале мы проводим селекцию.

1. Увеличиваем компетентность бактерий, добавляем к ним «коктейль», полученный в результате лигирования, а потом высеем эти бактерии на среду с антибиотиком, который является нашим первым селективным маркером.
2. Выбираем те бактериальные клоны, которые растут на среде с антибиотиком — они смогли съесть плазмиду со вставкой или хотя бы просто плазмиду (кота — с улыбкой или без).

3. Проводим с этими клонами второй этап селекции, в зависимости от того, какой ген мы использовали в качестве второго селективного маркера. Например, если этот ген — устойчивость к другому антибиотику, то мы переносим бактерии на среду с этим антибиотиком и выбираем те клоны, рост которых угнетается, — они ухитрились проглотить не просто плазмиду, а плазмиду, в которую была вшита вставка (кота с улыбкой).
4. Выращиваем полученную культуру бактерий.

И вот мы получили ее — бактериальную культуру, в которой живет созданный нами вектор. Вполне возможно, что это и было нашей конечной целью, и теперь мы, спокойные и счастливые, можем, например, включить в бактериях экспрессию гена-вставки и пожинать урожай синтезированных в результате белков.

Но если нам нужен чистый вектор, который можно будет потом помещать в другие клетки, то у нас появляется проблема, которая кажется неразрешимой. Как выволить вектор из бактерий? Ведь даже если мы выделим из этих бактерий ДНК, то помимо вектора получим еще и совершенно ненужную нам бактериальную хромосому.

Тут можно воспользоваться тем, что плазмидная ДНК имеет важные отличия от хромосомной: она, во-первых, гораздо меньше по размеру, а во-вторых, гораздо больше суперскручена (о том, что такое суперскрученность. Поэтому можно подобрать такие условия, в которых бактериальные хромосомы будут осаждаться, в то время как плазмиды останутся плавать в растворе. Достаточно будет отцентрифугировать получившийся осадок (чтобы вся бактериальная ДНК прочно «упала на дно»), а затем уже из надосадочной жидкости выделить нашу плазмиду (обычно для этого используются специальные колонки, которые очень облегчают и ускоряют работу).

### **Как поместить вектор в клетки**

И вот наступил желанный миг. Исследователь держит в руке пробирку, в которой плещется прозрачная жидкость — столькими трудами полученный вектор. И тут перед ним встает преграда. Клетки, в которые он собирается поместить свой вектор, отказываются его «глотать».

Дело в том, что липидная мембрана, которой окружены клетки, обладает избирательной проницаемостью — то есть она пропускает через себя одни частицы и не пропускает другие.

Крупные заряженные молекулы (а именно таковой и является ДНК) через эту мембрану самопроизвольно пройти не могут. И если бактерии, например, умеют проглатывать плазмиды из внешней среды (как уже было сказано выше), то, скажем, клетки животных к этому совершенно не склонны. Поэтому для того, чтобы поместить в клетку вектор, исследователю придется прибегать ко множеству хитростей, о которых и будет сейчас рассказано. Но сначала — немного терминов.

Для внесения в клетку вектора есть несколько обозначений в зависимости от того, какой используется вектор и в какие клетки он вносится.

**Трансформация** (о которой уже было немного рассказано) — это внесение плазмид (и других невирусных векторов) в бактерии, а также в клетки растений и грибов.

**Трансфекция** — то же самое, что и трансформация, но только в применении к клеткам животных.

И, наконец, **трансдукция** — это внесение в любые клетки вирусного вектора.

Эти термины, в общем, не очень строгие. Например, даже в некоторых солидных статьях трансдукцию иногда называют *вирусной трансфекцией* (а то и просто трансфекцией).

### **Вещества-проводники**

Самый простой и очевидный путь внесения в клетку генетического материала — соединить вектор с каким-нибудь переносчиком, у которого нет проблем с проникновением через мембрану, и позволить получившемуся комплексу «пролезть» внутрь клетки. Это не отнимает

много времени и не требует дорогостоящего оборудования. Такой способ обычно называют химической трансфекцией. В этом случае события развиваются по следующему сценарию:

*Вектор соединяется с переносчиком.*

Получившийся комплекс проглатывается клеткой и оказывается в цитоплазме.

В цитоплазме комплекс разваливается, и вектор высвобождается.

Вектор проникает в ядро и выполняет свое предназначение (скажем, с него начинается транскрибироваться мРНК).

К сожалению, почти на каждом из этих этапов возникают трудности. Во-первых, клетки захватывают далеко не все плавающие вокруг них комплексы. Во-вторых, не факт, что, оказавшись внутри клетки, вектор отделится от переносчика — вполне возможно, что они так и будут в обнимку плавать в цитоплазме, пока не подвергнутся деградации. В-третьих, даже если какой-то редкий комплекс умудрится проникнуть в клетку и там развалиться, то это означает, что помимо вектора в клетке оказывается еще и переносчик, который может быть токсичен, вызывать побочные эффекты и вообще «замыливать» результаты экспериментов. И, наконец, в-четвертых, только небольшая часть вектора, оказавшегося внутри клетки, сможет проникнуть в ядро. Иными словами, комплексы вектора с переносчиком надо добавлять к клеткам в огромном избытке, чтобы хотя бы маленькая часть из них выполнила свое предназначение.

Утешает то, что среди производителей веществ-переносчиков огромная конкуренция, и поэтому на рынке постоянно появляются новые составы с улучшенными свойствами, которые минимизируют вышеописанные трудности. У каждого из составов есть какая-то своя «фишка», которая дает ему преимущество в конкурентной борьбе — некоторые образуют с вектором такие компактные комплексы, которым гораздо легче пробраться внутрь клетки; другие эффективнее отделяются от вектора, оказавшись в цитоплазме; третьи более универсальны и работают на огромном количестве типов клеток; четвертые, наоборот, славятся своей избирательностью и проникают только в те клетки, которые, например, экспрессируют какой-то специфический рецептор. Одним словом, если исследователь решил засунуть вектор внутрь клетки с помощью химической трансфекции, то ему просто надо выбрать из множества составов, представленных на рынке, тот, который будет лучше работать в данном конкретном случае.

## **Дырки в мембране**

Но некоторые клетки так привередливы и капризны, что в принципе не соглашаются «глотать» комплексы ДНК с переносчиком (таким скверным характером славятся, например, первичные клетки — то есть те, которые не выращивались в культуре, а были получены непосредственно от живого организма). Чтобы ввести в эти клетки генетический материал, ученому приходится прибегать к грубой силе — продырявливать мембрану и помещать ДНК в образовавшиеся отверстия. Этот жестокий подход называется физической трансфекцией; он очень травматичен для клеток, и только некоторые из них переживут столь неделикатное обращение. Поэтому применять данную методику стоит, только если вы точно уверены, что обладаете достаточным количеством клеток и можете пожертвовать большей частью из них. Ну и к тому же, вам потребуется довольно дорогое оборудование.

Наверное, самый распространенный способ продырявливания мембраны называется электропорацией. Дело в том, что у клеток, попавших в электрическое поле, в мембране возникают отверстия (которые получаются тем больше, чем сильнее приложенное к клеткам поле). Если эти отверстия малы, то клетка сможет «залечить» их; если же они слишком велики, то клетка погибнет из-за необратимого нарушения целостности мембраны. Поэтому эмпирическим путем можно подобрать оптимальную величину поля для того, чтобы клетки, с одной стороны, продырявились, а с другой — остались в живых. А когда клетки продырявлены, то добавленный к ним вектор проникает сквозь отверстия и оказывается в цитоплазме.

Кроме электропорации, есть еще несколько способов — экзотических и не очень — сделать в мембране дырки. Например, с помощью:

- ультразвука (это называется сонопорация;

- лазера (оптическая трансфекция);
- нанопроволоки (с пришитым к ней вектором), которая физически прокалывает мембрану (импалефекция);
- магнитных взаимодействий (магнитофекция). (В этом случае вектор присоединяется к магнитным наночастицам, которые транспортируются внутрь клетки с помощью магнитного поля.)

Ну и, наконец, для самых непокорных клеток, которые не поддаются никакой из вышеописанных методик, существует прибор под названием «генная пушка». Генная пушка расстреливает упрямые клетки частичками металла (обычно используется золото) с присоединенным к ним вектором.

Генная пушка подходит практически для всех типов клеток, включая растительные, окруженные твердой клеточной стенкой, которая является практически непреодолимой преградой для большинства других методик.

А вообще, почти все вышеописанные методики дают более или менее похожие результаты на большей части типов клеток, а приборы для них стоят дорого. Поэтому, как правило, лаборатория покупает прибор для какой-то одной методики, и дальше уже по этой методике и доставляет генетический материал в клетки.

### **Объекты - транспорты**

Зачем придумывать новые и трудные пути помещения в клетку нуклеиновых кислот, если можно воспользоваться теми элегантными способами, которые за время долгой эволюции изобрели существа (или, возможно, вещества; нельзя точно сказать, живые они или нет), для которых транспортировка своего генетического материала внутрь клетки является необходимой фазой жизненного цикла? Все, наверное, уже догадались, что речь идет о вирусах.

**Вирусы** — это молекулы ДНК или РНК, упакованные в белковую оболочку (а иногда завернутые в липидный слой со встроенными в него вирусными белками). Именно оболочка играет главную роль в проникновении вируса через клеточную мембрану. Поэтому если засунуть в эту оболочку невирусную нуклеиновую кислоту, то она, будто вирус, тоже сможет попасть в клетку — как овечка, одетая в волчью шкуру. На этом принципе и основано использование вирусных векторов. Пожалуй, вирус — это самое эффективное транспортное средство для доставки в клетку генетического материала. Но приготовление вирусных векторов очень хлопотно, долго и трудоемко. Да вы сейчас и сами увидите.

Итак, чтобы сделать вирусный вектор, нужно для начала подобрать подходящий вирус. Идеальный кандидат:

1. Стабилен — то есть не склонен к спонтанным геномным перестройкам.
2. может вместить в себя даже самую большую вставку.
3. Не влияет на жизнедеятельность клетки.
4. Не вызывает иммунного ответа.
5. Встраивает свой геном не в первое попавшееся место генома хозяина (это может привести к непредсказуемым последствиям и вообще «замылить» результаты экспериментов), а в какую-нибудь определенную точку (а еще лучше — в точку, заданную самим исследователем).
6. И обладает другими симпатичными чертами.

К сожалению, идеал недостижим, и ученым приходится выбирать из того, что есть. А именно:

## Ретро

**Ретровирусы** долгое время были самой популярной основой для векторов. Это РНК-содержащие вирусы, которые, оказавшись в клетке, синтезируют ДНК на основе своей РНК с помощью *ревертазы* (собственно, поэтому они и называются «ретро», ведь синтез ДНК на основе РНК — это, в каком-то смысле, шаг назад). Ретровекторы хорошо выполняют свое предназначение, то есть стабильно доставляют в клетку заключенный в них генетический материал, но у них есть несколько недостатков, из-за которых работать с ними неудобно.

Во-первых, они все (за одним исключением, о котором скоро будет рассказано) способны инфицировать только делящиеся клетки. Поэтому если исследователь, например, собрался изучать нейроны, которые не склонны к делению, ему надо забыть о ретровекторах и начать искать что-нибудь другое.

Во-вторых, ретровирусы встраиваются в самые непредсказуемые участки генома, каждый раз разные, и это приводит к самым непредсказуемым последствиям. Для начала, из-за этого нарушается воспроизводимость экспериментов — но это еще ладно. Беда в том, что ретровектор может вклиниться в середину какого-нибудь важного гена, из-за чего этот ген выключится, а в клетке начнутся патологические изменения, которые могут довести ее до гибели. Или, наоборот, ретровектор может случайно включить какой-нибудь совершенно ненужный ген, например онкоген, что также приведет к очень печальным результатам (особенно если исследования проводятся не на культуре клеток, а на живом организме, и особенно если этот организм — человеческий).

Эти недостатки отвратили сердца ученых от ретровекторов и заставили их искать что-нибудь более подходящее. И найти кое-что замечательное удалось прямо внутри ретровирусного семейства.

## Ленти

**Лентивирусы** — это род ретровирусов, который отличается от прочих представителей своего семейства некоторыми приятными с точки зрения молекулярного клонирования чертами. Прежде всего, лентивирусы умеют заражать не только делящиеся, но и неделящиеся клетки. Эта особенность ужасна с точки зрения врача, который лечит вызванное лентивирусом заболевание, и прекрасна с точки зрения молекулярного биолога, который делает на основе лентивируса лентивектор. Ведь работая с таким вектором, ученый сможет использовать гораздо более широкий ассортимент клеточных типов, а значит, сделать гораздо больше великих открытий.

Плюс к тому, лентивекторы довольно емкие, то есть они способны вместить в себя крупные вставки. Отчасти это связано с тем, что из их генома в целях безопасности выкидывается большая часть, и в результате освобождается куча места.

Ну и кроме того, показано, что лентивирусы встраиваются в чуть менее непредсказуемые участки генома, чем прочие ретровирусы, а это тоже очень здорово.

«Ленти» по латыни значит «медленный». Это слово очень точно отражает характер лентивирусов — они вызывают заболевания с необычайно длинным инкубационным периодом. Вирус СПИДа — это тоже, кстати, лентивирус.

## Адено

**Аденовирусы**, наряду с ретровирусами, долго были самой популярной основой для векторов, но теперь потихоньку сдают свои позиции. Аденовирусы способны заражать не только делящиеся, но и неделящиеся клетки; ассортимент клеточных типов, которые они заражают, довольно широк. Но они не встраиваются в хозяйский геном, и поэтому подходят не для всех экспериментов. Кроме того, аденовирусы часто вызывают сильный иммунный ответ. Поэтому всё чаще они используются не в базовых исследованиях, а для всяких прикладных целей — например, для создания вакцин.



## AAV

И наконец, относительно недавно на сцене появился новый персонаж, который сразу расположил к себе ученых множеством чудесных качеств. Зовут его аденоассоциированный вирус (AAV).

AAV ведет себя настолько тихо, скромно и ненавязчиво, насколько этого вообще можно ждать от вируса. Практически единственное, что он делает, оказавшись в клетке, — это встраивается в хозяйский геном, причем почти всегда не в первое попавшееся, а в строго определенное место. Он, судя по имеющимся сейчас данным, не вызывает никаких заболеваний, поэтому и иммунный ответ на него очень слабый. К тому же, он способен заражать и делящиеся, и неделящиеся клетки. Одним словом, AAV — просто идеальная основа для вектора, хотя и он не лишен некоторых недостатков.

И главный его недостаток — малая емкость. В AAV-вектор могут «влезть» только совсем небольшие вставки, и в этом он очень проигрывает, например, лентивекторам. Кроме того, AAV — дефективный, несамостоятельный вирус. Он может размножаться только в клетках, которые уже заражены аденовирусом (что и отражено в его названии). Это совсем неплохая черта, если мы хотим заразить нашим вектором культуру клеток; но если мы собираемся делать вектор для генной терапии (методики лечения генетических — и не только — заболеваний, при которой организм заражается вирусным вектором, несущим необходимые этому организму гены), то такая дефективность будет нам очень мешать, потому что вирусы не смогут как следует распространяться по организму. Однако сейчас эта проблема вроде бы решена, и разработаны AAV-векторы, которые способны размножаться сами по себе, безо всякой помощи. Ну вот, допустим, подходящий вирус подобран. Теперь начинаются игры с его геномом.

1. Вначале нам нужно освободить в этом геноме место — то есть выкинуть из него какие-то гены. Обязательно нужно оставить те участки, на которые налипает оболочка (чтобы наш вектор был полноценным, «одетым» вирусом), и те гены, которые обеспечивают встраивание вирусного генома в геном хозяйской клетки (чтобы он мог выполнить свое предназначение); при этом от других областей — например, генов белков оболочки — мы можем с чистой совестью избавиться.
2. Из получившегося «огрызка» генома делается плаزمид — вставляются фрагменты, о которых уже было рассказано выше (точки начала репликации, селективные маркеры и так далее). В принципе, такие плазмиды уже есть в плазмидных базах данных, и, как правило, задача исследователя сводится к тому, чтобы подобрать подходящую.
3. В эту плазмиду вшивается необходимая вставка (со всеми прелестями многоэтапной селекции, которые были описаны выше).

*Теперь у нас возникает небольшая проблема. Даже поместив эту плазмиду в клетку, никаких вирусов мы не получим, потому что мы уже выкинули (в пункте 1) те гены, которые нужны для их создания. Поэтому нам придется пойти на маленькую хитрость.*

Мы поместим в клетки не одну плазмиду, а две. Первая, основная (назовем ее Пу), — это та, которую мы получили в пункте 3. А вторая, вспомогательная (назовем ее Ме), будет нести гены, которые мы выкинули в пункте 1. Обе плазмиды начнут размножаться в хозяйской клетке. Плаزمида Ме будет экспрессировать свои белки — например, белки оболочки и белки, необходимые для самосборки вирусов. Поскольку на Пу есть участки для налипания белков оболочки, то эти белки на нее и налипнут, и в результате мы получим вирус с необходимыми генами внутри, чего мы и добивались.

Итак, наш план действий таков:

4. Подбираем какие-нибудь клетки, которые хорошо поддаются трансфекции (такая линия клеток называется «упаковывающей»; обычно это линия эмбриональных клеток человеческой почки HEK293) и помещаем в них сразу две плазмиды — Пу и Ме, основную и вспомогательную.

5. Ждем некоторое время (около двух дней), чтобы успели образоваться вирусы. После этого собираем среду, в которой живут клетки, — вирусы плавают в ней.
6. Очищаем полученные вирусы (как правило, для этого используется центрифугирование и фильтрация) и... другими методами
7. Используем их по назначению, то есть заражаем ту линию клеток, на которой собираемся проводить эксперименты.

Это, конечно, только общая схема, у каждого конкретного вектора есть свои нюансы. Например, бывает, что вместо одной вспомогательной плазмиды используют две или даже три. При создании некоторых AAV-векторов упаковывающие клетки нужно заразить аденовирусом. А если мы создаем вектор для генной терапии, который должен уметь размножаться в хозяйской клетке и заражать ее соседей, то нам придется гораздо аккуратнее обращаться с вирусным геномом и расчищать в нем место с большой осторожностью, чтобы не нарушить способность вирусов к самостоятельному размножению. И так далее.

### **Последний этап**

Нам удалось поместить вектор в клетку. Тем или иным способом мы всё-таки умудрились поместить вектор в клетки. Нам остается последний шаг — нужно выбрать из всех клеток те, которые встроили векторную ДНК в свой геном.

Собственно, для этого мы и добавили в вектор последний селективный маркер — ген устойчивости к антибиотику, работающему на эукариотических клетках. Мы просто будем постоянно добавлять этот антибиотик в среду, в которой находятся наши клетки, — в результате останутся в живых и смогут делиться только те, которые имеют в геноме этот ген и всю нашу векторную ДНК впридачу.

**ФИНИШ!** Клонирование завершено. Мы получили линию генетически модифицированных клеток, в геноме которых присутствует наша вставка. Пришло время проводить с этими клетками необходимые эксперименты

### **Практическая часть**

Наследственность — способность организмов передавать свои признаки и особенности развития потомству. Благодаря этой способности все живые существа сохраняют в своих потомках характерные черты вида. Такая преемственность наследственных свойств обеспечивается передачей генетической информации. У эукариот материальными единицами наследственности являются гены, локализованные в хромосомах ядра и ДНК органелл. Наследственность наряду с изменчивостью обеспечивает постоянство и многообразие форм жизни и лежит в основе эволюции живой природы.

В ДНК используется четыре азотистых основания — аденин (А), гуанин (G), цитозин (С), тимин (Т), которые в русскоязычной литературе обозначаются буквами А, Г, Ц и Т. Эти буквы составляют алфавит генетического кода. В РНК используются те же нуклеотиды, за исключением нуклеотида, содержащего тимин, который заменён похожим нуклеотидом, содержащим урацил, который обозначается буквой U (У в русскоязычной литературе). В молекулах ДНК и РНК нуклеотиды выстраиваются в цепочки и, таким образом, получают последовательности генетических букв. В этом отношении генетический код построения организма, в рамках модели, можно рассматривать как текст.

Рассмотрим имитационную модель конструкции гипотетического организма. В такой модели постараемся сохранить «игровую среду», которая позволит работая с литературным текстом, обладающим некоторым смыслом, создавать информационные оценки для геновых конструкций. Другими словами заменяя генную конструкцию литературным текстом, для сохранения

интереса к исследованию, проводим вычисления, результаты которых можно перенести на конкретные ДНК структуры реальных организмов.

Исходные аксиомы:

1. Фрагмент ДНК - законченная фраза, представленная набором словоформ
2. Каждая фраза отделяется от последующей «стоп сигналом» (пробел)
3. Набор словоформ - типовое высказывание
4. Набор высказываний - генная конструкция организма,
5. Отдельная словоформа обладает информационным показателем, вычисляемым по выражению:

$$E_{\text{словоформы}} = \delta_0 + \delta_1 + \dots + \delta_i + \dots + \delta_n$$

$\delta_0 = 0,0294$  вероятность появления пробела между словами;  $\delta_i$  - вероятность появления буквы слова.

Где  $0 < i < 1.0$

Для русского алфавита имеем:

1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.	21.	22.	23.	24.	25.	26.	27.	28.	29.	30.	31.
<b>а</b>	<b>б</b>	<b>в</b>	<b>г</b>	<b>д</b>	<b>е, ё</b>	<b>ж</b>	<b>з</b>	<b>и</b>	<b>й</b>	<b>к</b>	<b>л</b>	<b>м</b>	<b>н</b>	<b>о</b>	<b>п</b>	<b>р</b>	<b>с</b>	<b>т</b>	<b>у</b>	<b>ф</b>	<b>х</b>	<b>ц</b>	<b>ч</b>	<b>ш</b>	<b>щ</b>	<b>ъ, ь</b>	<b>ы</b>	<b>э</b>	<b>ю</b>	<b>я</b>
0,0 75	0,0 17	0,0 46	0,0 16	0,0 3	0,0 87	0,0 09	0,0 18	0,0 75	0,0 12	0,0 34	0,0 42	0,0 31	0,0 65	0,1 1	0,0 28	0,0 48	0,0 55	0,0 65	0,0 25	0,0 02	0,0 11	0,0 05	0,0 15	0,0 07	0,0 04	0,0 17	0,0 19	0,0 03	0,0 07	0,0 22

### Общие вводные замечания

Биоинформатика является частью многих областей биологии. В экспериментальной молекулярной биологии методы биоинформатики, такие как создание изображений и обработка сигналов позволяют получать полезные результаты из большого количества исходных данных. В области генетики и геномики, биоинформатика помогает в упорядочивании и аннотировании геномов и наблюдаемых мутаций. Она играет роль в анализе данных из биологической литературы и развитии биологических и генетических онтологий по организации и запросу биологических данных. Она играет роль в анализе гена, экспрессии белка и регуляции. Инструменты биоинформатики помогают в сравнении генетических и геномных данных и, в целом, в понимании эволюционных аспектов молекулярной биологии. В общем виде, она помогает анализировать и каталогизировать биологические пути и сети, которые являются важной частью системной биологии. В структурной биологии, она помогает в симуляции и моделировании ДНК, РНК и белковых структур а также молекулярных взаимодействий.

В настоящее время проводятся исследования первичных структур различных молекул ДНК. Известны полные нуклеотидные последовательности ДНК ряда вирусов и плазмид.

Ген, участок молекулы ДНК (у некоторых вирусов — РНК), в котором закодирована информация, обеспечивающая развитие определенного признака (свойства) у данного организма и его передачу в ряду поколений. Участки нуклеиновой кислоты, кодирующие аминокислотную последовательность белков или последовательность оснований транспортных и рибосомных РНК, называют структурными генами. Последние вместе с необходимыми для их функционирования выражения регуляторными участками объединяются в более сложные генетические единицы — опероны. Многие гены высших организмов имеют прерывистое строение кодирующие части гена (экзоны) чередуются с некодирующими вставками (интронами).

Мы рассматриваем ген как молекулярную структуру. Ген эукариот-это совокупность сегментов ДНК, образующих экспрессирующуюся единицу. В результате экспрессии образуются один или несколько функциональных генных продуктов- РНК или полипептидов. Каждый ген содержит один или несколько сегментов ДНК, ответственных за регуляцию транскрипции и, следовательно, регуляцию экспрессии гена. Кодирующие области- это сегменты ДНК, которые кодируют какой-либо полипептид или функциональную РНК либо их составляющие. Те сегменты ДНК, с которых не транскрибируется никакой генный продукт, называются некодирующими. Одни некодирующие области (например, регуляторные сигналы, фланкирующие кодоны, или вставочные последовательности, прерывающие ген) являются составными частями генов. Другие участки, имеющие отношение к репликации ДНК или выполняющие какие-то пока неизвестные функции, находятся между генами.

С тех пор как в 1977 году был секвенирован фаг Phi-X174, последовательности ДНК всё большего числа организмов были дешифрованы и сохранены в базах данных. Эти данные используются для определения последовательностей белков и регуляторных участков. Сравнение генов в рамках одного или разных видов может продемонстрировать сходство функций белков или отношения между видами (таким образом могут быть составлены Филогенетические деревья). С возрастанием количества данных уже давно стало невозможным вручную анализировать последовательности. В наши дни для поиска по геномам тысяч организмов, состоящих из миллиардов пар нуклеотидов используются компьютерные программы. Программы могут однозначно сопоставить (выровнять) похожие последовательности ДНК в геномах разных видов; часто такие последовательности несут сходные функции, а различия возникают в результате мелких мутаций, таких как замены отдельных нуклеотидов, вставки нуклеотидов, и их «выпадения» (делеции). Один из вариантов такого выравнивания применяется при самом процессе секвенирования.

Полагаем, что между молекулярной структурой гена и тестом есть аналогия. Поскольку нам понятней работать с текстом (некоторым смысловым массивом, определенным образом записанных символов), приступим к анализу нескольких предложений. Предложения, по смыслу, содержат известные сведения о геноме. Это дает возможность провести информационный анализ предложения, где полученные выводы будут полезны в интерпретации последовательностей аминокислотной последовательности.

Энтропия достигает своего максимума, если все символы алфавита  $\{S_1, \dots, S_n\}$  независимы и генерируются источником с одинаковой вероятностью. Можно сказать, что в равновероятном алфавите все символы несут максимальную информационную нагрузку. Если же

алфавит не равновероятен, то некоторые его символы будут иметь меньшую информационную нагрузку, чем другие. Так, если считать, что в русском алфавите 31 буква (отождествляем еи ё, а также ь и ъ) и все они равновероятны, то

$$H_0 = - \sum_{j=1}^{31} 1/31 \log_2(1/31) = \log_2 31 \approx 4.954 \text{ бит}$$

Если учитывать частоты букв в соответствии с приведенной таблицей, то получим:

$$H_1 = - \sum_{j=1}^{31} P_j \log_2(P_j) \approx 4.460 \text{ бит}$$

т.е. средняя информация, приходящаяся на одну букву русского текста, заметно понижается.

Рассмотрим пример 1

Пример 1	Простые	слова	ложатся	на	лист	бумаги,	оставленный	на	столе
Значения Н [бит]	0,531	0,528	0,434	0,508	0,509	0,387	0,434	0,530	0,531
Сумма вероятностей	0,357	0,407	0,169	0,266	0,268	0,670	0,169	0,388	0,376
Полное значение Н	4,391 бит								

### Алгоритм количественного описания текста.

Полагаем, что появление любой буквы из алфавита, в тексте предложения, имеет определенную вероятность. Значения вероятности для каждой буквы алфавита представлены в таблице.

Набор словоформ - это некоторая часть генома. Оперирруя данными словоформами, в рамках гипотетической модели, рассуждаем о структуре геном или его фрагменте. Осуществляем вычисление информационной емкости сообщения (предложения).

### Первый блок для анализа

Это набор предложений текста, которые внесены в первый программный модуль анализа.

Гены человека, как правило, представляют собой функционально прерывистую последовательность нуклеотидов .

Относительно короткие кодирующие последовательности оснований чередуются в них с длинными декодирующими последовательностями.

Последовательности гена, представленные в молекуле зрелой иРНК, получили название экзонов.

Именно экзоны являются кодирующими участками гена, контролирующими аминокислотную последовательность белков.

Экзоны разделены не кодирующими участками — нитронами, которые вырезаются (сплайсинг).

В процессе созревания иРНК и не участвуют в процессе трансляции.

В настоящее время в понятие ген включаются не только транскрибируемые области (экзоны и интроны), но и фланкирующие ген последовательности.

Фланкирующие области гена, как правило, высоко консервативны.

Они характеризуются постоянством нуклеотидной последовательности, наблюдаемым даже при сравнении представителей различных видов.

Фланкирующие области гена содержат последовательности, необходимые для его правильной работы.

Например, промоторная область в начале 5 -области или хвостовая нетранслируемая область поли -А, расположенная на 3 -конце гена.

## Второй блок для анализа

### Рассказ А.П. Чехова - Мыслитель

Этот текст надо самостоятельно внести в модуль анализа, в том виде как он представлен ниже

Знойный полдень. В воздухе ни звуков, ни движений...

Вся природа похожа на одну очень большую, забытую богом и людьми, усадьбу.

Под опустившейся листвой старой липы, стоящей около квартиры тюремного зрителя Яншина, за маленьким треногим столом сидят сам Яншин и его гость, штатный зритель уездного училища Пимфов.

Оба без сюртуков; жилетки их расстегнуты; лица потны, красны, неподвижны; способность их выражать что-нибудь парализована зноем...

Лицо Пимфова совсем скисло и заплыло ленью, глаза его посоловели, нижняя губа отвисла.

В глазах же и на лбу у Яшкина еще заметна кое-какая деятельность; по-видимому, он о чем-то думает...

Оба глядят друг на друга, молчат и выражают свои мучения пыхтеньем и хлопаньем ладонями по мухам.

На столе графин с водкой, мочалистая вареная говядина и коробка из-под сардин с серой солью.

Выпиты уже первая, вторая, третья...

-- Да-с! -- издает вдруг Яшкин, и так неожиданно, что собака, дремлющая недалеко от стола, вздрагивает и, поджав хвост, бежит в сторону. --

Да-с! Что ни говорите, Филипп Максимыч, а в русском языке очень много лишних знаков препинания!

-- То есть, почему же-с? -- скромно вопрошает Пимфов, вынимая из рюмки крылышко мухи. --

Хотя и много знаков, но каждый из них имеет свое значение и место.

-- Уж это вы оставьте! Никакого значения не имеют ваши знаки.

Одно только мудрование... Наставит десяток запятых в одной строчке и думает, что он умный.

Например, товарищ прокурора Меринов после каждого слова запятую ставит.

Для чего это? Милостивый государь -- запятая, посетив тюрьму такого-то числа -- запятая, я заметил -- запятая, что арестанты -- запятая... тьфу!

В глазах рябит! Да и в книгах то же самое... Точка с запятой, двоеточие, кавычки разные.

Противно читать даже. А иной франт, мало ему одной точки, возьмет и натыкает их целый ряд... Для чего это?

-- Наука того требует... -- вздыхает Пимфов.

-- Наука... Умопомрачение, а не наука... Для форсу выдумали... пыль в глаза пушать...

Например, ни в одном иностранном языке нет этого ять, а в России есть... Для чего он, спрашивается?

Напиши ты хлеб с ятем или без ятя, нешто не всё равно?

-- Бог знает что вы говорите, Илья Мартыныч! -- обижается Пимфов. --

Как же это можно хлеб через е писать? Такое говорят, что слушать даже неприятно.

Пимфов выпивает рюмку и, обиженно моргая глазами, отворачивает лицо в сторону.

-- Да и секли же меня за этот ять! -- продолжает Яшкин. --

Помню это, вызывает меня раз учитель к черной доске и диктует: "Лекарь уехал в город". Я взял и написал лекарь с е.



Выпорол. Через неделю опять к доске, опять пиши: "Лекарь уехал в город". Пишу на этот раз с ятем. Опять пороть. За что же, Иван Фомич? Помилуйте, сами же вы говорили, что тут ять нужно! "Тогда, говорит, я заблуждался, прочитав же вчера сочинение некоего академика о ять в слове лекарь, соглашаюсь с академией наук.

Порю же я тебя по долгу присяги"... Ну, и порол. Да и у моего Васютки всегда ухо вспухши от этого ять...

Будь я министром, запретил бы я вашему брату ятем людей морочить.

-- Прощайте, -- вздыхает Пимфов, моргая глазами и надевая сюртук. --

Не могу я слышать, ежели про науки...

-- Ну, ну, ну... уж и обиделся! -- говорит Яшкин, хватая Пимфова за рукав. --

Я ведь это так, для разговора только... Ну, сядем, выпьем!

Оскорбленный Пимфов садится, выпивает и отворачивает лицо в сторону.

Наступает тишина. Мимо пьющих кухарка Феона пронесет лохань с помоями.

Слышится помойный плеск и визг облитой собаки. Безжизненное лицо Пимфова раскисает еще больше; вот-вот растает от жары и потечет вниз на жилетку.

На лбу Яншина собираются морщинки. Он сосредоточенно глядит на мочалистую говядину и думает...

Подходит к столу инвалид, угрюмо косится на графин и, увидев, что он пуст, приносит новую порцию... Еще выпивают.

-- Да-с! -- говорит вдруг Яншин.

Пимфов вздрагивает и с испугом глядит на Яншина. Он ждет от него новых ересей.

-- Да-с! -- повторяет Яшкин, задумчиво глядя на графин. -- По моему мнению, и наук много лишних!

-- То есть, как же это-с? -- тихо спрашивает Пимфов. -- Какие науки вы находите лишними?

-- Всякие... Чем больше наук знает человек, тем больше он мечтает о себе. Гордости больше...

Я бы перевешал все эти... науки... Ну, ну... уж и обиделся!

Экий какой, ей-богу, обидчивый, слова сказать нельзя! Сядем, выпьем!

Подходит Феона и, сердито тыкая в стороны своими пухлыми локтями, ставит перед приятелями зеленые щи в миске.

Начинается громкое хлебание и чавканье.словно из земли вырастают три собаки и кошка.

Они стоят перед столом и умильно поглядывают на жующие рты.

За щами следует молочная каша, которую Феона ставит с такой злобой, что со стола сыплются ложки и корки. Перед кашей приятели молча выпивают.

-- Всё на этом свете лишнее! -- замечает вдруг Яшкин.

Пимфов роняет на колени ложку, испуганно глядит на Яншина, хочет протестовать, но язык ослабел от хмеля и запутался в густой каше...

Вместо обычного "то есть, как же это-с?" получается одно только мычание.

-- Всё лишнее... -- продолжает Яшкин. -- И науки, и люди... и тюремные заведения, и мухи... и каша...

И вы лишний... Хоть вы и хороший человек, и в бога веруете, но и вы лишний...

-- Прощайте, Илья Мартыныч! -- лепечет Пимфов, силясь надеть сюртук и никак не попадая в рукава.

-- Сейчас вот мы натрескались, налопались, -- а для чего это? Так... Всё это лишнее...

Едим и сами не знаем, для чего... Ну, ну... уж и обиделся! Я ведь это так только... для разговора! И куда вам идти? Посидим, потолкуем... выпьем!

Наступает тишина, изредка только прерываемая звяканьем рюмок да пьяным побрякиваньем...

Солнце начинает уже клониться к западу, и тень липы всё растет и растет.

Приходит Феона и, фыркая, резко махая руками, расстилает около стола коврик.

Приятели молча выпивают по последней, располагаются на ковре и, повернувшись друг к другу спинами, начинают засыпать...

"Слава богу, -- думает Пимфов, -- сегодня не дошел до сотворения мира и иерархии, а то бы волосы дыбом, хоть святых выноси..."

## **ЗАДАНИЕ**

1. Провести анализ текстовых блоков представленных ниже.
2. Для каждого блока вычислить информационную оценку ВСЕХ предложений
3. Информационную оценку всего блока
4. Найти подобные (близкие по информационному показателю) фрагменты блока.
5. Необходимые расчеты провести используя программный модуль
6. Полученные результаты описать в среде Word.
7. Представить графические иллюстрации полученных результатов