

УДК 577.29

CRISPR-СИСТЕМЫ АДАПТИВНОГО ИММУНИТЕТА ПРОКАРИОТ

© 2012 г. К. С. Пугач^{1, 2*}, А. В. Лопатина^{1, 2}, К. В. Северинов^{1, 2}

¹Институт биологии гена Российской академии наук, Москва, 119334

²Институт молекулярной генетики Российской академии наук, Москва, 123182

Поступила в редакцию 25.08.2011 г.

Принята к печати 14.10.2011 г.

Системы CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) прокариот обнаружены в геномах многих эубактерий и подавляющего большинства архей. Уникальная особенность этих систем заключается в том, что они способны обеспечивать *адаптивный* наследуемый иммунитет к мобильным генетическим элементам. В данном обзоре описаны общие принципы функционирования CRISPR-иммунитета и дается характеристика CRISPR-систем различных микроорганизмов.

Ключевые слова: CRISPR-системы, Cas-белки, малые РНК, адаптивный иммунитет.

CRISPR ADAPTIVE IMMUNITY SYSTEMS OF PROCARYOTES, by K. S. Pougach^{1, 2*}, A. V. Lopatina^{1, 2}, K. V. Severinov^{1, 2} (¹Institute of Gene Biology, Russian Academy of Sciences, Moscow, 119334 Russia, *e-mail: xenik1alt@gmail.com; ²Institute of Molecular Genetics, Russian Academy of Sciences, Moscow, 123182 Russia). CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) is a newly identified prokaryotic immunity system against foreign genetic elements. In contrast to other cellular defense mechanisms (e.g. restriction-modification) CRISPR-mediated immunity is adaptive and can be programmed to protect cells against a particular bacteriophage or conjugative plasmid. In this review we describe general principles of CRISPR systems action and summarize known details of CRISPR systems from different microorganisms.

Keywords: CRISPR systems, Cas proteins, bacterial immunity.

ВВЕДЕНИЕ

Значительная часть геномов эубактерий и архей представлена генами, приобретенными в ходе горизонтального переноса генов (ГПГ) [1]. Основные механизмы, обеспечивающие ГПГ, — это поглощение ДНК из внешней среды (*трансформация*) и перенос чужеродного генетического материала мобильными генетическими элементами, такими как плазмиды (*конъюгация*) или бактериофаги (*трансдукция*) [2]. Встраивание чужеродной ДНК может приводить к нарушению функций клеточных генов [3, 4], поэтому у прокариот имеются механизмы, которые ограничивают ГПГ, предотвращая проникновение чужеродной ДНК в клетку или разрушая чужеродную ДНК внутри клетки. Один из таких механизмов обеспечивается системой рестрикции-модификации. Ферменты модификации метилируют рестрикторные сайты в геноме бактерий, предотвращая их раз-

вание нуклеазами рестрикции. Чужеродная ДНК, содержащая неметилированные сайты рестрикции, расщепляется под действием клеточных нуклеаз. В качестве другого механизма, предотвращающего ГПГ, можно рассматривать появление мутаций в генах, кодирующих поверхностные рецепторы, используемые бактериофагами для проникновения в клетку. В результате фаг перестает адсорбироваться на поверхности бактерий [5]. Известны также разнообразные, но мало изученные внутриклеточные механизмы abortивной инфекции, которые активируют гибель зараженной клетки и тем самым препятствуют распространению фагов и плазмид в популяции [3].

Недавно открыт еще один механизм защиты прокариот от бактериофагов и плазмид, осуществляемый системами CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats). Самым неожиданным свойством возникновения

Принятые сокращения: CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) — короткие палиндромные повторы, регулярно разделенные спейсерами; Cas-гены (CRISPR-Associated) — ассоциированные с CRISPR гены; ГПГ — горизонтальный перенос генов; крРНК — короткие фрагменты РНК; пре-крРНК — единый протяженный предшественник крРНК; RAMP (Repair-Associated Mysterious Proteins) — неизвестные белки, ассоциированные с репарацией; tracrRNA малая некодирующая РНК; PAM (Protospacer Adjacent Motif) — последовательность, прилегающая к протоспейсеру.

* Эл. почта: xenik1alt@gmail.com



Рис. 1. Схема CRISPR-кассеты. Черными прямоугольниками изображены повторы, разделенные уникальными спейсерами (шестиугольники серого цвета). Cas-гены обозначены черными пятиугольниками. На схеме указана также лидерная последовательность CRISPR-кассеты (“лидер”).

устойчивости к мобильным генетическим элементам под действием CRISPR-систем оказалось то, что клетка становится защищенной лишь после проникновения чужеродной ДНК. При этом приобретенная защита наследуема в поколениях. Таким образом, в определенном смысле CRISPR-системы являются примером классической адаптивной эволюции по Ламарку (наследуемость благоприобретенных признаков).

Так называемая “CRISPR-кассета” представляет собой набор коротких несовершенных палиндромных повторов ДНК, разделенных спейсерами – участками ДНК одинаковой длины, но различающимися по последовательности, которые внутри одной кассеты практически идентичны (рис. 1). CRISPR-кассете предшествует обогащенная АТ лидерная последовательность, длина которой обычно составляет несколько сот пар нуклеотидов [6]. В непосредственной близости от большинства CRISPR-кассет находятся cas-гены (CRISPR-associated – ассоциированные с CRISPR). Cas-белки (некоторые из них являются хеликазами и нуклеазами) обеспечивают функционирование CRISPR-систем [7].

В 2005 г. три группы ученых независимо друг от друга обнаружили, что последовательности многих CRISPR-спейсеров сходны с участками геномной ДНК бактериофагов и плазмид [8–10]. Предполагали, что такие спейсеры могли бы обеспечивать соответственно устойчивость клеток к заражению бактериофагами и предотвращать трансформацию плазмидами. Это предположение подтверждено экспериментально: в 2007 г. Баррангу (Barrangou) и соавт. [11] показали, что клетки *Streptococcus thermophilus*, несущие в CRISPR-кассете спейсер, идентичный участку геномной ДНК бактериофага, становятся устойчивыми к этому фагу.

В общем виде, механизм защитного действия CRISPR-систем сводится к следующему. CRISPR-кассета транскрибируется как единый протяженный предшественник (пре-крРНК), который процессируется с образованием набора коротких фрагментов (крРНК), каждый из которых содержит последовательность одного из спейсеров CRISPR-кассеты. крРНК в комплексе с Cas-белками связывается с комплементарной чужеродной ДНК или РНК, что приводит к ее расщеплению (рис. 2).

РАСПРОСТРАНЕНИЕ И СТРУКТУРА CRISPR-КАССЕТ

По данным Крисса (Grissa) и соавт. [12] около 48% бактериальных геномов и более 95% геномов архей содержат CRISPR-кассеты. Последовательности, похожие на CRISPR, обнаружены также в митохондриальном геноме кормового боба *Vicia faba* [13]. Количество CRISPR-кассет в одном геноме может варьировать от 1 до 18. Самое большое число кассет (18) найдено в геноме археи *Methanocaldococcus jannaschii*. В среднем, на одну кассету CRISPR приходится около 60 повторов [14]. Рекордсмен на сегодняшний день – термофильная бактерия *Chloroflexus* sp. Y-400-fl, в одной из CRISPR-кассет которой содержится 374 повтора. В известных на сегодняшний день кассетах длина повторов варьирует от 23 до 50 (в среднем 31) п.н., а длина спейсеров – от 17 до 84 (в среднем 36) п.н. Длина CRISPR-повторов и спейсеров внутри одной кассеты постоянна. Последовательности повторов в пределах одной CRISPR-кассеты также практически идентичны [18].

В целом, лишь очень малая часть (около 2%), идентифицированных CRISPR-спейсеров похожа на последовательности ДНК известных мобильных генетических элементов, а в случае наиболее изученной бактерии *Escherichia coli* – никаких соответствий между спейсерами и последовательностями известных бактериофагов вообще не наблюдается. Возможно, это связано с тем, что число таких последовательностей, представленных в базах данных, не превышает 0.0002% от оцениваемого общего числа геномных последовательностей ныне существующих вирусов прокариот [15]. В этой связи представляются интересными данные, полученные в ходе сравнительного анализа CRISPR-кассет ацидофильных бактерий и последовательностей геномов вирусов из затопленной шахты в Калифорнии [16]. В этом случае число спейсеров, гомологичных вирусным геномам, достигает 40% и более. Столь высокий уровень соответствия объясняется, по-видимому, постоянством условий окружающей среды, изолированностью и очень простым видовым составом данного сообщества бактерий и их вирусов.

Cas-БЕЛКИ

cas-Гены имеются только в тех геномах, в которых есть CRISPR-кассеты. cas-Гены расположе-

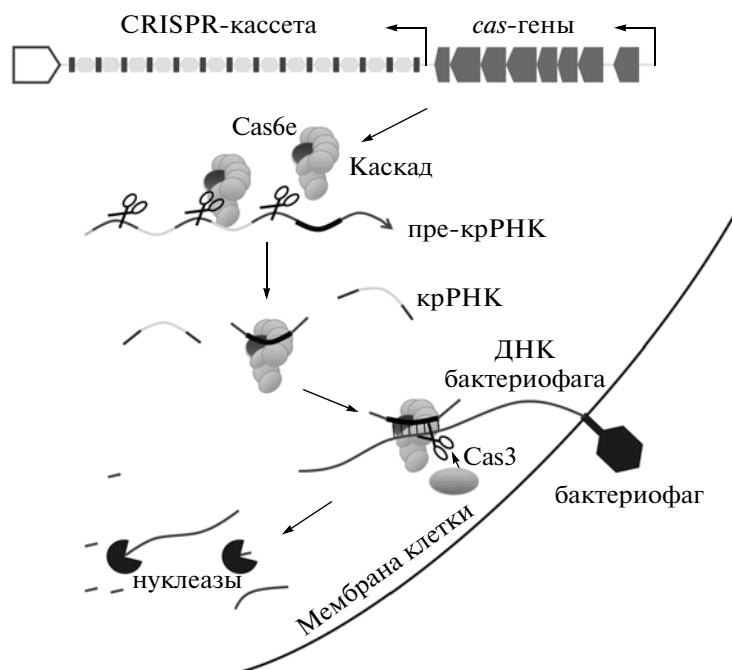


Рис. 2. Гипотетическая схема действия CRISPR-системы *E. coli* K12. CRISPR-кассета транскрибируется в составе одного протяженного РНК транскрипта (пре-крРНК). Продукты *cas*-генов собираются в комплекс под названием Каскад. Белок CasE в составе комплекса Каскад разрезает пре-крРНК по последовательностям повторов. Образующиеся крРНК длиной около 60 н. остаются связанными с Каскадом (эффекторный комплекс). При проникновении в клетку чужеродной ДНК она распознается эффекторным комплексом, содержащим комплементарную крРНК. Это приводит к привлечению нуклеазы/хеликазы Cas3, разрезанию чужеродной ДНК и ее дальнейшей деградации.

ны в геноме группами и всегда находятся в непосредственной близости от CRISPR-кассет [6]. Тем не менее, *cas*-гены впервые аннотированы не в связи с CRISPR-касетами [17]. В результате анализа *in silico* ряда известных на то время прокариотических геномов было описано около двадцати семейств белков, названных RAMP (Repair-Associated Mysterious Proteins). Предполагалось, что эти белки участвуют в репарации клеточной ДНК. Позже определили набор “коровых” (от слова *core* – сердцевина, базис) *cas*-генов, которые закодированы в геномах большинства организмов, содержащих CRISPR-касеты [6, 7].

В настоящее время в литературе нередко встречаются публикации, в которых для обозначения одного и того же Cas-белка или его гомолога используют различные названия. Наиболее полная классификация CRISPR/Cas-локусов (совокупности CRISPR-кассет и *cas*-генов) предложена Макаровой (Макарова) и соавт. в 2011 г. [18]. Согласно этой классификации выделяют три группы CRISPR/Cas-локусов, различаемых, во-первых, на основании филогенетического сходства последовательности белка Cas1, который кодируется всеми известными *cas*-оперонами; во-вторых, на основании специфических ортологов *cas*-генов, характерных для всех представителей данной группы CRISPR/Cas локусов, а также на основании последовательности CRISPR-повто-

ров. Первые два параметра хорошо коррелируют между собой, в то время как только часть повторов можно соотнести с выделенными по первым двум критериям группам CRISPR/Cas локусов.

Несмотря на некоторую неоднозначность, классификация CRISPR/Cas-локусов, предложенная Макаровой и соавт. [18], на сегодняшний день наиболее полна. Тем не менее, в литературе продолжается использование альтернативных, более ранних, названий. Поэтому, во избежание путаницы при описании различных Cas-белков из различных групп локусов ниже приводятся не только их названия по классификации Макаровой и соавт., но и другие названия. Они приведены также в таблице.

Группа I

Наличие гена *cas3* – характерный признак этой группы, а в группах II и III ортологи Cas3 отсутствуют. В качестве молекулы-мишени таких систем выступает чужеродная ДНК.

Белки Cas6e или Cas6f (называемые также Cse3, Csy4 и CasE) расщепляют полноразмерный транскрипт с образованием крРНК. Комплекс Cas-белков и крРНК узнает молекулу-мишень, а белок Cas3, по-видимому, производит разрезание.

Классификация *cas*-генов

Название гена [18]	Название гена [7]	Название гена [51]	Биохимические свойства/ ферментативная активность	
<i>cas1</i>	<i>cas1</i>	<i>cas1</i>	металл-зависимая оцДНК/дцДНК-эндонуклеаза оцРНКазы, разрезает РНК преимущественно в U-богатых участках, дцДНК/дцРНК-нуклеазная активность	
<i>cas2</i>	<i>cas2</i>	<i>cas2</i>		
<i>cas3</i>	<i>cas3</i>	<i>cas3</i>	нуклеаза, содержит предположительно HD-нуклеазный домен и DEXH-хеликазный домен	
<i>cas4</i>	<i>cas4</i> и <i>csa1</i>	—	RecB нуклеаза	
<i>cas5</i>	<i>cas5a</i> , <i>cas5d</i> , <i>cas5e</i> , <i>cas5h</i> , <i>cas5p</i> , <i>cas5t</i> и <i>cmx5</i>	<i>casD</i>	компонент Каскада	
<i>cas6</i>	<i>cas6</i> , <i>cmx6</i>	—	процессинг пре-крРНК	
<i>cas6e</i>	<i>cse3</i>	<i>case</i>		компонент Каскада, процессинг пре-крРНК и связывание крРНК
<i>cas6f</i>	<i>csy4</i>	—	осуществляет процессинг пре-крРНК и связывание крРНК	
<i>cas7</i>	<i>csa2</i> , <i>csd2</i> , <i>cse4</i> , <i>csH2</i> , <i>csp1</i> и <i>cst2</i>	<i>cas7</i> , <i>cse4/casC</i>	компонент Каскада, встраивание новых спейсеров	
<i>cas9</i>	<i>csn1</i> и <i>csx12</i>	—	процессинг пре-крРНК и разрезание ДНК мишени	
<i>cse1</i>	<i>cse1</i>	<i>casA</i>		компонент Каскада
<i>cse2</i>	<i>cse2</i>	<i>casB</i>		компонент Каскада

В таблице приведены названия *cas*-генов по трем классификациям – Макаровой (Макарова) и соавт. [18], Хафта (Haft) и соавт. [7] и Браунса (Brouns) и соавт. [51]. Кратко перечислены биохимические свойства белков, подробно изложенные в основном тексте.

Группа II

Характерный признак этой группы – наличие гена *cas9* (альтернативные названия *csn1*, *csx12*). В качестве молекулы-мишени выступает чужеродная ДНК.

Процессирование пре-крРНК осуществляет клеточная РНКазы III в присутствии белка Cas9 и малой некодирующей РНК (*tracrRNA*). Cas9 участвует также в связывании с ДНК-мишенью и ее дальнейшей деградации.

Группа III

Характерный признак этой группы – наличие белка Cas10 (альтернативные названия Cmr2, Csm1 and Csx11). В этой группе разрезание пре-крРНК осуществляет белок Cas6 (*Cmx6*), который следует отличать от белков Cas6e и Cas6f, так как они не являются ортологами. CRISPR-системы этой группы способны атаковать не только чужеродную ДНК, но и РНК (показано *in vitro* на примере *Pyrococcus furiosus*) [19].

МЕХАНИЗМ ДЕЙСТВИЯ CRISPR-СИСТЕМ

В работе 2009 г. ван дер Оост (van der Oost) и соавт. [20] предложили выделять три стадии работы CRISPR-систем: стадии адаптации, экспрессии и интерференции. Во время адаптации короткий фрагмент чужеродной ДНК (пре-спейсер)

вырезается и встраивается в CRISPR-кассету. В результате кассета удлиняется на один новый спейсер и один дополнительный повтор. На стадии экспрессии происходит транскрипция CRISPR-кассеты и процессинг пре-крРНК; зрелая крРНК связывается с Cas-белками, образуя эффекторный комплекс. При интерференции эффекторный комплекс узнает фрагмент чужеродной ДНК или РНК за счет комплементарных взаимодействий с участком крРНК, соответствующим спейсеру, и расщепляет его.

Адаптация

Первые доказательства существования процесса встраивания новых спейсеров в CRISPR кассеты, получены Баррангу (Barrangou) и соавт. [11] при изучении мутантов *S. thermophilus*, приобретших устойчивость к бактериофагу в ходе инфекции. Оказалось, что CRISPR-кассеты мутантных клеток удлиняются за счет встраивания новых спейсеров, идентичных фрагментам генома фага, к которому возникла устойчивость. Введение спейсера (и соответственно повтора) в CRISPR-кассету чувствительных клеток искусственным путем (генноинженерными методами) также приводит к возникновению устойчивости. Таким образом, наличие CRISPR-спейсера, идентичного фрагменту генома бактериофага, явилось достаточным условием для возникнове-

ния устойчивости к этому бактериофагу. Были отобраны также мутантные бактериофаги, которые заражают и чувствительные клетки дикого типа, и устойчивые клетки с удлинённой CRISPR-кассетой. Как оказалось, участок ДНК фага, соответствующий CRISPR-спейсеру устойчивой клетки (такой участок называется “протоспейсер”), содержит однонуклеотидные замены, нарушающие комплементарное соответствие протоспейсера и спейсера. Таким образом, некоторые точечные мутации в протоспейсере снимают защитное действие CRISPR-системы.

Новые спейсеры встраиваются в конец кассеты, примыкающий к лидерной последовательности [21]. Это указывает на то, что лидерная последовательность, скорее всего, активно участвует в адаптации. Безлидерные CRISPR-кассеты, по-видимому, не способны встраивать новые спейсеры [22]. Самые удалённые от лидера спейсеры, возможно, являются самыми древними. Таким образом, CRISPR-кассету можно рассматривать как своеобразную “историю болезни” штамма, в которой порядок следования спейсеров отражает хронологию взаимодействия с фагами/плазмидами. Однако эта история не совершенна, так как рост кассеты ограничен, и многие внутренние спейсеры утрачиваются в результате делеций [16] или рекомбинаций внутри кассеты.

В работах Дево (Deveau) и соавт. [21, 23, 24] показано, что к протоспейсерам, соответствующим спейсерам CRISPR-кассеты *S. thermophilus*, прилежит последовательность PAM (Protospacer Adjacent Motif – последовательность, прилегающая к протоспейсеру) – AGAAG. Как оказалось, мутантные фаги, имеющие точечные замены в этой последовательности, заражают даже те клетки, CRISPR-кассеты которых содержат спейсеры, в точности соответствующие протоспейсеру [21, 23]. Очевидно, что PAM играют важную роль в функционировании CRISPR-кассет.

Белки Cas1 и Cas2 – единственные строго консервативные белки всех CRISPR-систем. В случае *E. coli*, наличие генов *cas1* и *cas2* не влияет на защитное действие CRISPR-системы; это служит основанием для гипотезы о том, что продукты этих генов участвуют в адаптации [25]. К сожалению, прямых экспериментальных данных, подтверждающих это предположение, не имеется ни для одной CRISPR-системы.

В составе Cas1 обнаружен предполагаемый интегральный домен [26]. Cas1 *E. coli* K12 способен расщеплять структуры Холлидея *in vitro* [27]. Cas1 *P. aeruginosa* имеет ДНКазную активность. В присутствии ионов марганца или магния этот белок способен расщеплять как одно-, так и двуцепочечную ДНК на фрагменты длиной около 80 п.н. [28]. Cas1 из *Sulfolobus solfataricus* нуклеазной активностью не обладает, однако этот белок имеет

сильное сродство к ДНК и РНК и способен ускорять гибридизацию ДНК и РНК [29]. Возможно, Cas1 участвует в процессе встраивания новых спейсеров в CRISPR-кассету.

В геномах некоторых микроорганизмов гены *cas1* и *cas4* (*csa1*) слиты в один ген. Этот факт, вероятно, указывает на то, что продукты этих генов вовлечены в один и тот же процесс и работают согласованно [20]. Cas4, по-видимому, является RecB-подобной нуклеазой [26].

Данные о кристаллической структуре и биохимических свойствах Cas2 из *S. solfataricus* получены Белоглазовой (Beloglazova) и соавт. [30]. Cas2 оказался металлозависимой рибонуклеазой, предпочтительно разрезающей одноцепочечные молекулы РНК в районах, обогащенных остатками уридина [30]. В клетках *E. coli* Cas2 накапливается в районах клеточных полюсов [31], которые считаются самыми уязвимыми для проникновения фагов частями клетки [32].

При изучении *S. thermophilus* показано, что клетки, в геноме которых отсутствует ген *cas7*, не способны встраивать новые спейсеры. Однако непонятно, является ли Cas7 единственным белком, задействованным в этом процессе у этого организма, или же он действует в комплексе с другими клеточными белками [44].

Экспрессия

Транскрипция пре-крРНК инициируется с промотора, расположенного внутри лидерной последовательности [6, 33–35]. В некоторых случаях короткие транскрипты могут синтезироваться и с противоположной цепи ДНК CRISPR-кассеты. Возможно, они выполняют регуляторную функцию [36]. На сегодняшний день известны два клеточных белка, которые участвуют в регуляции транскрипции CRISPR-кассет *E. coli* K12, H-NS и LeuO [34, 37]. В клетках с делецией гена *hns* или гиперэкспрессией гена *leuO* количество *cas*-транскриптов и крРНК существенно возрастает.

В геноме *Thermus thermophilus* HB8 насчитывается десять CRISPR/Cas-локусов, еще два находятся на плазмиде pTT27. С помощью метода геномных микрочипов показано, что уровень транскриптов большинства генов CRISPR/Cas-локусов этого микроорганизма повышается во время инфекции литическим бактериофагом YS40, несмотря на то, что ни одна из CRISPR-кассет не содержит спейсеров, соответствующих геному фага [38]. Кроме того, экспрессия *cas*-генов этого микроорганизма регулируется белком-рецептором циклического AMP [38].

Полноразмерный CRISPR-транскрипт подвергается процессингу (см. рис. 2) Процессированные транскрипты представляют собой моно- или олигомеры CRISPR-спейсеров и повторов.

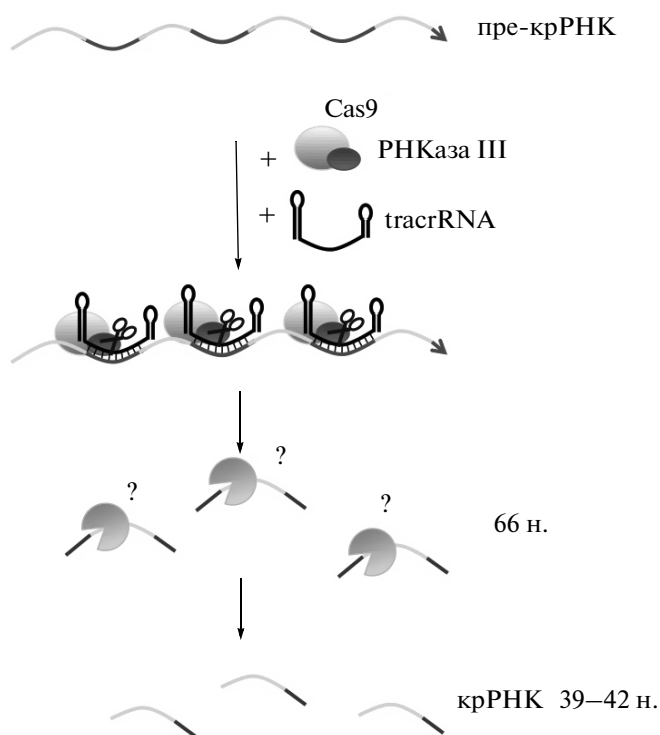


Рис. 3. Модель созревания крРНК *Streptococcus pyogenes*. Светло-серым прорисованы спейсеры, черным — повторы. tracrRNA (черная) связывается с пре-крРНК в составе комплекса с РНКазой III и Cas9. Образующийся дуплекс разрезается с образованием фрагментов РНК длиной 66 н., содержащих последовательности спейсера, которые фланкированы повторами. Эти фрагменты затем процессируются с образованием зрелой крРНК длиной 39–42 н.

При изучении *E. coli* K12, CRISPR/Cas-локус которой относится к группе I по классификации Макаровой и соавт. [20, 26], обнаружено, что белки Cse1 (CasA), Cse2 (CasB), Cas7 (CasC, Csa2, Csd2, Cse4, Csh2, Csp1 и Cst2), Cas5 (CasD, Cas5a, Cas5d, Cas5e, Cas5h, Cas5p, Cas5t и Cmx5) и Cas6e (CasE, Cse3) образуют комплекс, получивший название Каскад (Cascade, CRISPR-Activated Complex for Antiviral Defence; комплекс защиты от вирусов, активируемый CRISPR). Состав Каскада можно описать следующей формулой: Cse1₁Cse2₂Cas7₆Cas5₁Cas6e₁. При помощи электронной микроскопии показано, что этот белковый комплекс имеет форму, напоминающую силуэт морского конька [39]. Предложена модель, согласно которой крРНК связывается вдоль протяженной вогнутой поверхности частицы Каскада. Каскад расщепляет полноразмерный транскрипт CRISPR-кассеты *E. coli* K12 с образованием крРНК длиной 61 н. Для разрезания необходима эндонуклеолитическая активность белка Cas6e (CasE, Cse3). Участок специфического расщепления пре-крРНК располагается внутри последовательности повтора на расстоянии 8 н. от 5'-конца спейсера. 3'-конец крРНК об-

разуется шпильку за счет палиндромной последовательности нуклеотидов, входящей в состав повтора [39]. Зрелая крРНК несет гидроксильную группу на 5'-конце и 2',3'-циклический фосфат на 3'-конце [40]. Сашитал (Sashital) и соавт. [41] получили кристаллическую структуру комплекса крРНК и Cas6e из *T. thermophilus*. Cas6e связывается с пре-крРНК в районе шпильки, образованной палиндромной последовательностью повтора так, что разрезание пре-крРНК всегда происходит по основаниям, следующим за шпилькой.

CRISPR/Cas-локусы *Streptococcus pyogenes* относятся к группе II по классификации Макаровой и соавт. [20, 26]. В геноме этого микроорганизма нет генов *cas6*, *cas6f* (*csy4*) и *cas6e* (*casE*, *cse3*), ответственных за процессинг пре-крРНК в других системах. Группой под руководством Шарпантье (Charpentier) [42] обнаружено, что процессинг пре-крРНК в этой системе осуществляет белок Cas9 (Csn1, Csx12) в комплексе с клеточной РНКазой III и малой некодирующей РНК tracrRNA. Молекулы tracrRNA комплементарно связываются с последовательностями повторов в пре-крРНК, а РНКазы III в присутствии Cas9 разрезают дуплекс с образованием фрагментов CRISPR-РНК длиной 66 н. Эти фрагменты содержат последовательность спейсера, которая фланкирована повторами и впоследствии процессируются с образованием зрелой крРНК длиной 39–42 н. (рис. 3). Механизм этого более глубокого процессинга пока не известен.

Процессинг молекул пре-крРНК у *P. furiosus*, CRISPR/Cas-локусы которого относятся к группе III по классификации Макаровой и соавт. [20, 26], сходен с таковым у *E. coli* [19]. Разрезание пре-крРНК на короткие фрагменты осуществляется по последовательностям повторов. 5'-концы образовавшихся крРНК всегда содержат “ручку” (handle) из восьми последних нуклеотидов повтора. 3'-концы крРНК варьируют, что приводит к образованию двух классов крРНК длиной ~39 и ~45 н. Карте (Carte) и соавт. [43] показали, что процессинг пре-крРНК *P. furiosus* осуществляется белком Cas6 (Cmx6), входящим в группу “коровых” Cas-белков. Им также удалось выявить участок специфического связывания этого белка на молекуле РНК-предшественника. Этот участок представляет собой первые 12 н. повтора. Далее по направлению к 3'-концу повтора, на расстоянии 10 н. от участка специфического связывания Cas6, находится сайт разрезания пре-крРНК. Связывание нарушается как в случае изменения последовательности этого участка, так и в случае изменения расстояния между участком связывания и участком разрезания. Для разрезания (при условии сохранения связывания) требуется динуклеотид AA на расстоянии восьми нуклеотидов от 3'-конца повтора; введение замен в динуклеотид AA препятствуют расщеплению. Интересно,

что чем дальше спейсер находится от последовательности лидера, тем меньшее количество соответствующей крРНК *P. furiosus* обнаруживается в клетке. Такой градиент, возможно, обусловлен терминацией транскрипции на последовательностях палиндромных повторов, образующих тугоплавкие шпильки.

Интерференция

Специфическое разрезание молекулы-мишени *in vivo* под действием крРНК показано лишь в случае *S. thermophilus*. При инфекции бактериофагом клеток, в CRISPR-касетах которых присутствуют спейсеры, идентичные участкам геномной ДНК бактериофага, внутри последовательностей протоспейсеров накапливаются двуцепочечные разрывы [44]. Естественно, что такая инфекция оказывается непродуктивной. В этой же работе показано, что в клетках, содержащих геномную делецию гена *cas9* (кодируемый им белок является, по-видимому, нуклеазой), расщепления ДНК фага не наблюдается и инфекция продуктивна.

Как описано выше, у *E. coli* K12 не обнаружено спейсеров, соответствующих известным бактериофагам, а экспрессия *cas*-генов подавлена белком H-NS. Поэтому CRISPR/Cas-система этого организма оказывается неактивной. Чтобы проверить, способен ли Каскад *E. coli* K12 в комплексе с крРНК обеспечивать защиту от фаговой инфекции, созданы экспрессионные плазмиды, несущие искусственные CRISPR-касеты со спейсерами, идентичными различным фрагментам генов фага λ [25]. Индукция ко-экспрессии таких пре-крРНК и полного набора Cas-белков обеспечивает клеткам устойчивость к бактериофагу. Устойчивость не возникает при отсутствии любого из компонентов Каскада или гена *cas3*. Отсутствие генов *cas1* и *cas2* не влияет на возникновение устойчивости, что и послужило основанием для гипотезы о том, что продукты этих генов участвуют в адаптации (см. выше). *In vitro* Каскад *E. coli* узнает чужеродную двуцепочечную ДНК. Образующийся при этом комплекс содержит участок локально расплавленной ДНК, образующий РНК • ДНК-гетеродуплекс длиной 33 п.н. Для образования гибрида крРНК • ДНК расходования АТФ не требуется. Для эффективного узнавания мишени эффекторным комплексом необходима РАМ-последовательность (предполагаемый консенсус для *E. coli* ATG/AAG) и полная комплементарность крРНК и первых семи нуклеотидов протоспейсера (так называемая seed-последовательность). Несовпадения по другим позициям допустимы и не препятствуют узнаванию мишени [45].

Как и другие иммунные системы, CRISPR-системы должны обладать способностью отличать

“свою” и чужеродную ДНК. В противном случае эффекторный комплекс Каскада с крРНК мог бы связываться с участком CRISPR-касеты, содержащим соответствующий спейсер, что могло бы приводить к деградации геномной ДНК. Марраффини (Marraffini) и Сонтхаймер (Sontheimer) [46], изучавшие CRISPR-систему *S. epidermidis* RP62A, показали, что для деградации ДНК-мишени эффекторным комплексом необходимо, чтобы не возникали комплементарные взаимодействия между нуклеотидами крРНК и ДНК-мишени в позициях -2 , -3 , -4 (если за $+1$ принять первое основание протоспейсера). По-видимому, комплементарные взаимодействия крРНК и ДНК-мишени по этим позициям нарушают образование эффекторного комплекса, что предотвращает разрезание геномной ДНК и ее последующую деградацию.

Хейл (Hale) и соавт. [47] изучали работу эффекторного комплекса *P. furiosus in vitro*. Оказалось, что мишенью действия этого комплекса является РНК, а не ДНК. Авторы добавляли меченые молекулы РНК-мишени к комплексу очищенных рекомбинантных “коровых” Cas-белков и крРНК, комплементарных меченой РНК. Молекулы РНК-мишени распознаются и разрезаются напротив 14-го нуклеотида с 3'-конца крРНК, т.е. 3'-конец крРНК выступает в роли “молекулярной линейки”, измеряющей расстояние до разрыва. Участвует ли специфическая РНК-расщепляющая активность эффекторного комплекса *P. furiosus* в ограничении ГПГ или является артефактом реакции *in vitro* — не известно.

ЗАКЛЮЧЕНИЕ

Исследования последних нескольких лет показали, что CRISPR-системы широко распространены среди прокариот. В отличие от всех других известных механизмов защиты от мобильных генетических элементов, возникновение иммунитета под действием CRISPR-систем адаптивно. Понимание этого факта заставляет задуматься о роли случайности в возникновении основной концепции, которые легли в основу современной молекулярной биологии. Так, если бы знаменитый флуктуационный тест Лурии-Дельбрюка, в котором было показано возникновение мутаций устойчивости к бактериофагам за счет предсуществующей изменчивости, был проведен с *S. thermophilus*, а не с *E. coli*, то авторы с неизбежностью пришли бы к выводу об адаптивном механизме возникновения мутаций у бактерий: ведь подавляющее большинство мутаций устойчивости к бактериофагам у *S. thermophilus* возникает за счет действия CRISPR-систем. С другой стороны, причины, по которым в принципе функциональные CRISPR-системы, присутствующие у *E. coli*, очевидно не играли никакой роли в возникновении

устойчивости ни к одному из известных и широко распространенных фагов этой бактерии, также остаются совершенно неясными.

Очевидно, что CRISPR-системы способны обеспечить защиту клетки от бактериофагов и конъюгативных плазмид и таким образом могли бы ограничивать два важнейших пути ГПГ в природе — трансдукцию и конъюгацию. Некоторые косвенные факты подтверждают важную эволюционную роль CRISPR-систем. Так, например, в случае *Streptococcus pneumoniae* — микроорганизма, при исследовании которого открыта природная трансформация [48], — ни один из 11 штаммов с известной последовательностью генома не несет CRISPR-кассет. Природная трансформация играет огромную роль в развитии пневмонии, так как не только гены вирулентности, но и гены устойчивости к антибиотикам, переносятся между штаммами главным образом с помощью трансформации [49]. Таким образом, можно предположить, что в данном случае CRISPR-система была элиминирована естественным отбором для обеспечения достаточной частоты трансформации. Сходные рассуждения можно применить к микроорганизму *Helicobacter pylori* — патогену, который активно поглощает чужеродную ДНК извне и не имеет CRISPR-систем. Впрочем, стоит отметить, что гонококки, для которых природная трансформация также очень важна, имеют CRISPR-касеты [50].

Многие детали механизма работы CRISPR-систем остаются невыясненными и загадочными. Несомненно, что в ближайшее время в этой быстро развивающейся области будут получены новые фундаментальные результаты. В частности, можно ожидать, что появятся биохимические и структурные данные, которые позволят провести детальный сравнительный анализ действия CRISPR-систем и эукариотических систем РНК-интерференции и выяснить вопрос об общности происхождения этих двух чрезвычайно важных механизмов, каждый из которых основан на малых некодирующих РНК. Другим важным направлением может стать использование CRISPR-систем в биотехнологии и биомедицине для конструирования бактериальных штаммов, предотвращения вирусных инфекций при ферментациях и селективном подавлении генов устойчивости к антибиотикам.

Работа получила финансовую поддержку программы “Молекулярная и клеточная биология” президиума Российской академии наук, Российского фонда фундаментальных исследований (1-04-01373) и благотворительного фонда поддержки научных исследований “Наука за продолжение жизни”.

СПИСОК ЛИТЕРАТУРЫ

1. Nakamura Y., Itoh T., Matsuda H., Gojobori T. 2004. Biased biological functions of horizontally transferred genes in prokaryotic genomes. *Nat. Genet.* **36**, 760–766.
2. Thomas C.M., Nielsen K.M. 2005. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat. Rev. Microbiol.* **3**, 711–721.
3. Jore M.M. 2010. CRISPR-mediated antiviral defence in prokaryotes. *Thesis submitted in fulfillment of the requirements for the degree of doctor*. Wageningen: Wageningen University.
4. Coleman D., Knights J., Russell R., et al. 1991. Insertional inactivation of the *Staphylococcus aureus* beta-toxin by bacteriophage phi 13 occurs by site- and orientation-specific integration of the phi 13 genome. *Mol. Microbiol.* **5**, 933–939.
5. Rocha E.P.C., Danchin A., Viari A. 2001. Evolutionary role of restriction/modification systems as revealed by comparative genome analysis. *Genome Res.* **11**, 946–958.
6. Jansen R., Embden J.D., Gaastra W., et al. 2002. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* **43**, 1565–1575.
7. Haft D.H., Selengut J., Mongodin E.F., Nelson K.E. 2005. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput. Biol.* **1**, e60.
8. Bolotin A., Quinguis B., Sorokin A., Ehrlich S.D. 2005. Clustered regularly interspaced short palindrome repeats (CRISPR) have spacers of extrachromosomal origin. *Microbiol.* **151**, 2551–2561.
9. Mojica F.J., Diez-Villasenor C., Garcia-Martinez J., et al. 2005. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* **60**, 174–182.
10. Pourcel C., Salvignol G., Vergnaud G. 2005. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology.* **151**, 653–663.
11. Barrangou R., Flemaux C., Deveau H., et al. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science.* **315**, 1709–1712.
12. Grissa I., Vergnaud G., Pourcel C. 2007. The CRISPR-Rdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics.* **8**, 172.
13. Mojica F.J., Diez-Villasenor C., Soriano J., Juez G. 2000. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol. Microbiol.* **36**, 244–246.
14. Marraffini L.A., Sontheimer E.J. 2010. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat. Rev. Genetics.* **11**, 181–190.
15. Rohwer F. 2003. Global phage diversity. *Cell.* **113**, 171–182.
16. Andersson A.F., Banfield F. 2008. Virus population dynamics and acquired virus resistance in natural microbial communities. *Science.* **320**, 1047–1049.
17. Makarova K.S., Aravind L., Grishin N.V., et al. 2002. A DNA repair system specific for thermophilic archaea and bacteria predicted by genomic context analysis. *Nucl. Acids Res.* **30**, 482–496.

18. Makarova K.S., Haft D.H., Barrangou R., et al. 2011. Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.* **9**, 467–477.
19. Hale C., Kleppe K., Terns R.M., et al. 2008. Prokaryotic silencing (psi)RNAs in *Pyrococcus furiosus*. *RNA* **14**, 2572–2579.
20. van der Oost, J., Jore M.M., Westra E.R., et al. 2009. CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends Biochem. Sci.* **34**, 401–407.
21. Horwath P., Romero D.A., Coute-Monvoisin A.-C., et al. 2008. Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J. Bacteriol.* **190**, 1401–1412.
22. Lillestøl R.K. 2006. A putative viral defence mechanism in archaeal cells. *Archaea*. **2**, 59–72.
23. Deveau H., Barrangou R., Garneau J., et al. 2008. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.* **190**, 1390–1400.
24. Mojica F.J., Diez-Villasenor C., Garcia-Martinez J., et al. 2009. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiol. J.* **155**, 733–740.
25. Brouns S.J., Jore M.M., Lundgren M., et al. 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*. **321**, 960–964.
26. Makarova K.S., Grishin N.V., Shabalina S.A., et al. 2006. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol. Direct*. **1**, 7.
27. Babu M., Beloglazova N., Flick R., et al. 2011. A dual function of the CRISPR-Cas system in bacterial antiviral immunity and DNA repair. *Mol. Microbiol.* **79**, 484–502.
28. Wiedenheft B., Zhou K., Jinek M., et al. 2009. Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. *Structure*. **17**, 904–912.
29. Han D., Krauss G. 2009. Characterization of the endonuclease SSO2001 from *Sulfolobus solfataricus* P2. *FEBS Lett.* **583**, 771–776.
30. Beloglazova N., Brown G., Zimmerman M.D., et al. 2008. A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats. *J. Biol. Chem.* **283**, 20361–20371.
31. Tang J., Akerboom J., Vaziri A., et al. 2010. Near-isotropic 3D optical nanoscopy with photon-limited chromophores. *Proc. Natl. Acad. Sci. USA*. **107**, 10068–10073.
32. Edgar R., Rokney A., Feeney M., et al. 2008. Bacteriophage infection is targeted to cellular poles. *Mol. Microbiol.* **68**, 1107–1116.
33. Marraffini L.A., Sontheimer E.J. 2008. CRISPR interference limits horizontal gene transfer in *Staphylococci* by targeting DNA. *Science*. **322**, 1843–1845.
34. Pul U., Wurm R., Arsian Z., et al. 2010. Identification and characterization of *E. coli* CRISPR-cas promoters and their silencing by H-NS. *Mol. Microbiol.* **75**, 1495–1512.
35. Pougach K., Semenova E., Bogdanova E., et al. 2010. Transcription, processing and function of CRISPR cas-
setes in *Escherichia coli*. *Mol. Microbiol.* **77**, 1367–1379.
36. Lillestøl R.K., Shah S.A., Brugger K., et al. 2009. CRISPR families of the crenarchaeal genus *Sulfolobus*: bidirectional transcription and dynamic properties. *Mol. Microbiol.* **72**, 259–272.
37. Westra E.R., Pul U., Heidrich N., et al. 2010. H-NS-mediated repression of CRISPR-based immunity in *Escherichia coli* K12 can be relieved by the transcription activator LeuO. *Mol. Microbiol.* **77**, 1380–1393.
38. Agari Y., Sakamoto K., Tamakoshi M., et al. 2010. Transcription profile of *Thermus thermophilus* CRISPR systems after phage infection. *J. Mol. Biol.* **395**, 270–281.
39. Jore M.M., Lundgren M., van Duijn E. et al., 2011. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nature Struct. Biol.* **18**, 529–537
40. Wiedenheft B., van Duijn E., Bultema J. 2011. RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc. Natl. Acad. Sci. USA*. **108**, 10092–10097.
41. Sashital D.G., Jinek M., Doudna J.A. 2011. An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3. *Nature Struct. Biol.* **18**, 680–688.
42. Deltcheva E., Chylinski K., Sharma C.M., et al. 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor Rnase III. *Nature*. **471**, 588–589.
43. Carte J., Wang R., Li H., et al. 2008. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev.* **22**, 3489–3496.
44. Garneau J., Dupuis M.-E., Villion M., et al. 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*. **468**, 67–72.
45. Semenova E., Jore M.M. 2011. Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl. Acad. Sci. USA*. **108**, 10098–10103.
46. Marraffini L.A., Sontheimer E.J. 2010. Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature*. **463**, 568–572.
47. Hale C., Zhao P., Olson S., et al. 2009. RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell*. **139**, 945–956.
48. Griffith F. 1928. The significance of pneumococcal types. *J. Hyg.* **27**, 113–159.
49. Johnsborg O., Havarstein L.S. 2009. Regulation of natural genetic transformation and acquisition of transforming DNA in *Streptococcus pneumoniae*. *FEMS Microbiol. Rev.* **33**, 627–642.
50. Dorer M.S., Talarico S., Salama N.R. 2009. Helicobacter pylori's unconventional role in health and disease. *PLoS Pathog.* **5**, e1000544
51. Al-Attar S., Westra E.R., van der Oost J., Brouns S.J.J. 2011. Clustered regularly interspaced short palindromic repeats (CRISPRs): the hallmark of an ingenious antiviral defense mechanism in prokaryotes. *Biol. Chem.* **392**, 277–289.