

УДК 577.214.5

КАНОНИЧЕСКОЕ И НЕКАНОНИЧЕСКОЕ ПОЛИАДЕНИЛИРОВАНИЕ РНК

© 2017 г. И. Г. Устьянцев, Ю. С. Голубчикова, О. Р. Бородулина, Д. А. Крамеров*

Институт молекулярной биологии им. В.А. Энгельгардта Российской академии наук, Москва, 119991

*e-mail: kramеров@eimb.ru

Поступила в редакцию 29.06.2016 г.

Принята к печати 03.08.2016 г.

Полиаденилирование – это процесс нематричного присоединения адениловых нуклеотидов к 3'-концу РНК, происходящий по завершении транскрипции. В результате полиаденилирования образуется так называемый поли(А)-хвост, длина которого обычно составляет 250–300 н. В первой части обзора, посвященного полиаденилированию, рассмотрен классический процесс формирования 3'-конца мРНК путем разрезания транскрипта, синтезированного РНК-полимеразой II, и сопряженного с этим процессом полиаденилирования, осуществляемого канонической поли(А)-полимеразой. Подробно описаны нуклеотидные последовательности (в частности, сигнал полиаденилирования AAUAAA), необходимые для разрезания мРНК и синтеза поли(А)-хвоста, а также многочисленные белки и образуемые ими комплексы, которые осуществляют разрезание и полиаденилирование мРНК. Рассмотрено значение поли(А)-хвоста для увеличения времени жизни мРНК и стимуляции ее трансляции. Во второй части приведены данные, свидетельствующие о способности к AAUAAA-зависимому полиаденилированию РНК, транскрибируемых РНК-полимеразой III с некоторых видов SINE (Short Interspersed Elements). Особенности структуры РНК-полимеразы III и осуществляемой ею транскрипции делают полиаденилирование РНК, синтезированных этим ферментом, весьма необычным процессом, требующим дальнейшей расшифровки. Вторая, наиболее подробная, часть обзора отражает историю развития данного направления исследований. Это обусловлено тем, что AAUAAA-зависимое полиаденилирование РНК, синтезируемых РНК-полимеразой III, было открыто авторами данной статьи и ранее не освещалось в научных обзорах. В третьей части представлены сведения о полиаденилировании РНК, не требующем AAUAAA и осуществляемом неканоническими поли(А)-полимеразами (Trf4 и Trf5). Эти ферменты входят в состав белкового комплекса TRAMP и добавляют короткие поли(А)-хвосты на молекулы РНК, которые способствуют быстрой деградации этих РНК. Этот механизм необходим для выявления поли(А)-маркирования и элиминации неправильно свернутых некодирующих транскриптов (например, рибосомных и транспортных РНК).

Ключевые слова: канонические и неканонические поли(А)-полимеразы, сигналы и элементы полиаденилирования, белковые факторы полиаденилирования, РНК-полимеразы II и III, SINE

DOI: 10.7868/S0026898417010189

КАНОНИЧЕСКОЕ ПОЛИАДЕНИЛИРОВАНИЕ

Полиаденилирование – это процесс нематричного добавления адениловых нуклеотидов на 3'-конец РНК, чаще всего осуществляемый как посттранскрипционная модификация. Пре-мРНК, синтезируемая эукариотической РНК-полимеразой II (RNAP II), еще не может служить полноцен-

ной матрицей для синтеза белка. Прежде она должна пройти несколько модификаций, важнейшими из которых являются кепирование, сплайсинг и полиаденилирование. Кепирование представляет собой присоединение 7-метилгуанозина на 5'-конец молекулы пре-мРНК, чему обычно сопутствует метилирование остатков рибозы первых двух нуклеотидов. Сплайсинг заключается в удалении из

Сокращения: RNAP II и RNAP III (RNA polymerase II and III) – РНК-полимераза II и III соответственно; CTD (C-terminal domain) – С-концевой домен; USE (upstream sequence element) и DSE (downstream sequence element) – нуклеотидные последовательности (элементы), участвующие в 3'-процессинге пре-мРНК и расположенные до или после сайта полиаденилирования соответственно; PAS (polyadenylation signal) – сигнал полиаденилирования; RRM (RNA recognition motif) – РНК-опознающий мотив; CPSF (cleavage and polyadenylation specificity factor) – фактор специфичности разрезания и полиаденилирования; CstF (cleavage stimulation factor) – фактор стимуляции разрезания; CFIm и CFIIIm (mammalian cleavage factor Im and IIIm) – фактор разрезания I и II соответственно; PAP (poly(A) polymerase) – поли(А)-полимераза; SINE (Short Interspersed Elements) – короткие рассеянные по геному элементы; TRAMP (Trf4/Air2/Mtr4p Polyadenylation complex) – полиаденилирующий комплекс, состоящий из белков Trf4, Air2 и Mtr4p.

пре-мРНК некодирующих последовательностей (интронов) и последовательном соединении между собой остальных участков РНК (экзонов). Полиаденилированию подвергается 3'-конец молекулы пре-мРНК, а структура, образующаяся в результате этого процесса, называется поли(А)-хвост. Обычно длина поли(А)-хвоста составляет 250–300 н. Полиаденилированию предшествует разрезание цепи пре-мРНК, в результате которого и формируется ее 3'-конец. Разрезание РНК и полиаденилирование сопряжены между собой, вместе эти два события носят название 3'-процессинга [1, 2]. Поли(А)-хвост защищает мРНК от действия клеточных нуклеаз и тем самым удлиняет время ее жизни в клетке. Кроме того, полиаденилирование повышает эффективность транспорта мРНК из ядра в цитоплазму. Наконец, поли(А)-хвост играет важную роль в циклизации мРНК, которая способствует многократной трансляции молекул мРНК [3].

Впервые фракцию поли(А)-содержащей клеточной РНК выделили в начале 70-х годов прошлого века при обработке РНК рибонуклеазами, гидролизующими фосфодиэфирные связи за гуаниновыми (РНКза Т1) и пиримидиновыми (РНКза А) нуклеотидами: при такой обработке сегменты поли(А) остаются интактными. В тот же период выяснилось, что эти сегменты не кодируются ДНК, и был обнаружен полиаденилирующий фермент – поли(А)-полимераза (РАР) [2]. Однако оказалось, что для полиаденилирования пре-мРНК необходима не только сама РАР, но и множество вспомогательных белковых факторов, а также определенные нуклеотидные последовательности в пре-мРНК, которые этими факторами распознаются.

Самая важная из этих последовательностей называется сигналом полиаденилирования (PAS). PAS представляет собой гексамер, чаще всего AAUAAA, расположенный за 20–30 н. до 3'-конца мРНК [4]. Консенсусную последовательность гексамера можно представить как NNUANA, ее варианты AAUAAA, A(U/G)UAAA и UAUAAA находят в 79% мРНК с функциональным гексамером. Зачастую одна пре-мРНК содержит несколько редких вариантов PAS [5]; это наблюдается примерно у 54 и 32% генов человека и мыши соответственно. Присутствие более одного, не обязательно неканонического, PAS в пределах одного транскрипта позволяет регулировать длину образующейся мРНК: при выборе проксимального PAS получается более короткий 3'-нетранслируемый район (3'-НТР). Такие изоформы мРНК могут обладать разной стабильностью, транслируемостью и внутриклеточной локализацией вследствие того, что 3'-НТР содержат различные регуляторные последовательности, например сайты связывания микроРНК и регуляторных белков. Таким образом, мРНК с более

коротким 3'-НТР может обеспечивать более высокий уровень синтеза белка. Эти формы мРНК характерны для многих активно пролиферирующих клеток, включая раковые [6–9]. Нельзя не отметить, что PAS не удается обнаружить в 12% полиаденилируемых РНК [5].

Помимо гексамера, на 3'-конце могут находиться еще две области, ответственные за осуществление 3'-процессинга. Та из них, которая лежит по ходу транскрипции ниже сайта полиаденилирования, называется downstream sequence element (DSE) и представляет собой либо GU-богатую (YGUGUUY), либо олиго-U-последовательность (UUUUU) [1, 10]. Выше PAS может располагаться upstream sequence element (USE), причем у млекопитающих USE нередко представлен двумя (или более) пентамерами UGUAN ($N = A > U \geq C/G$). В отсутствие канонического PAS USE может играть главную роль в сборке комплекса 3'-процессинга [2, 11]. Между DSE и PAS (в пределах 11–23 н. от PAS и за 10–30 н. до DSE) находится сайт разрезания, нередко это динуклеотид CA, от которого может зависеть эффективность разрезания [12]. Сайт разрезания часто определен нестрого, тем самым может регулироваться длина транскрипта [6]. Перечисленные последовательности служат сайтами связывания различных белков – факторов 3'-процессинга (рис. 1). У млекопитающих их известно не менее 14, хотя в комплексе 3'-процессинга можно найти куда больше белков – до 85. Важно отметить, что белковый комплекс, осуществляющий 3'-процессинг пре-мРНК, собирается еще в процессе транскрипции на С-концевом домене (CTD) наибольшей субъединицы RNAP II [1].

Один из важнейших факторов 3'-процессинга – CPSF (cleavage and polyadenylation specificity factor) (рис. 1) – узнает сигнал полиаденилирования, а усиливает это взаимодействие второй по важности белок комплекса – CstF (cleavage stimulation factor). Вместе с PAS эти белки привлекают все остальные факторы 3'-процессинга [1]. CPSF состоит из семи компонентов: CPSF100, CPSF73 и симплекина, формирующих ядро комплекса, и периферических Fip1, CPSF160, Wdr33, и CPSF30 [13]. Долго считалось, что самая большая субъединица, CPSF160, отвечает за опознавание PAS несмотря на то, что не содержит, как и все прочие компоненты фактора, канонического RRM (РНК-опознающего мотива) [14]. Однако недавно выяснилось, что CPSF160 не образует УФ-сшивок с гексамером AAUAAA. CPSF160, а также CPSF73 и CPSF30 сшиваются с более вышележащими районами 3'-НТР [8]. Получены свидетельства в пользу того, что с PAS могут взаимодействовать, взаимно дополняя друг друга, Wdr-33 и CPSF30 [9]. Wdr33 – довольно большая субъединица, сравнимая по размеру с CPSF160. Не исключено, впрочем, что CPSF160 отвечает за связывание гексаме-

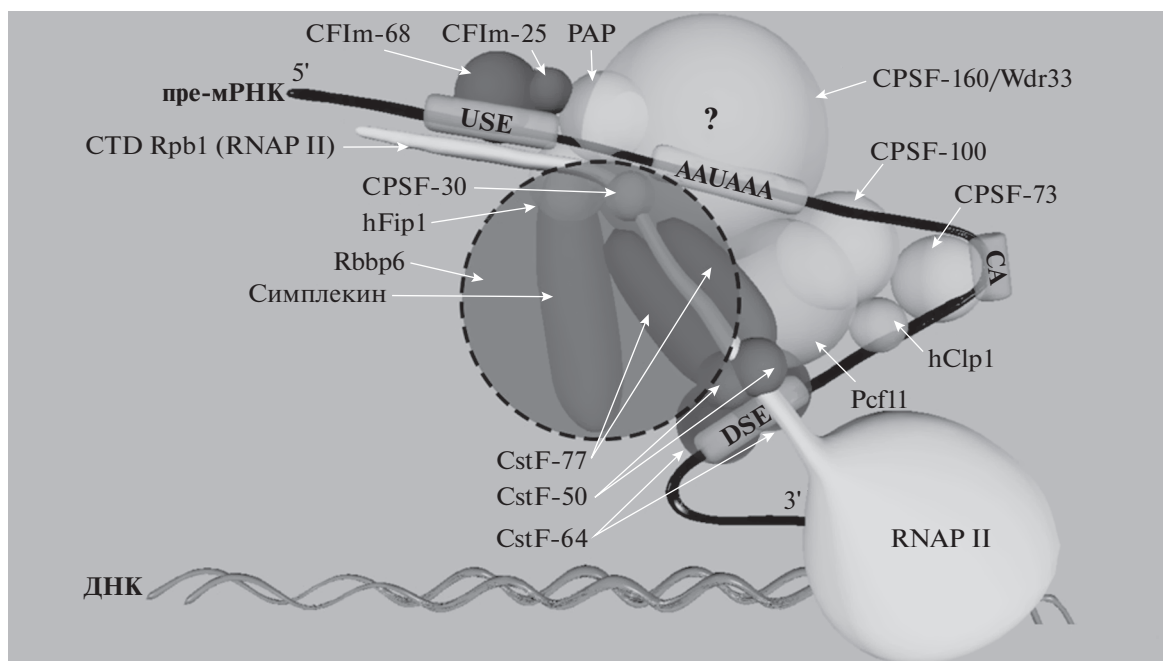


Рис. 1. Схема белкового комплекса, осуществляющего 3'-концевой процессинг пре-мРНК. RNAP II синтезирует транскрипт, который разрезается CPSF-73 в сайте CA. После этого PAP осуществляет нематричный синтез поли(А) на образовавшемся 3'-конце пре-мРНК. Для 3'-концевого процессинга необходим сигнал полиаденилирования (AAUAAA), а также элементы DSE и/или USE. Белки, участвующие в 3'-концевом процессинге, изображены сферическими или овальными структурами и отмечены стрелками. Остальные объяснения см. в тексте. Рисунок отражает данные [1, 8, 9, 13].

ра, но только в некоторых вариантах PAS, в других же вариантах в комплекс вместо нее включается Wdr33 [9]. CPSF73 содержит домен β -CASP, характерный для эндонуклеаз. Скорее всего, именно CPSF73 разрезает пре-мРНК, хотя в отсутствие других компонентов комплекса эта активность невелика [15]. CStF состоит из трех субъединиц: CstF77, CstF50 и CstF64. Вероятнее всего, он функционирует в виде димера. CstF77 отвечает за димеризацию и взаимодействие с CPSF160, а также служит платформой для двух других субъединиц [16]. CstF64 содержит RRM и требуется для связывания с DSE, но в необходимой мере ее сродство проявляется только в присутствии CPSF. По всей видимости, таким образом контролируется присутствие PAS и DSE на 3'-конце и должное расстояние между ними. CstF50 отвечает за взаимодействие с CTD RNAP II, с ней также могут связываться симплекин и Fip1 [14, 17].

Еще два важных фактора 3'-процессинга — это факторы разрезания I и II (Cleavage factor Im и Pm) (рис. 1). CFIm состоит из двух субъединиц — большой (есть три варианта, наиболее распространена форма 68 кДа) и малой — с массой 25 кДа, причем белок работает в виде гетеротетрамера, состоящего из двух больших и двух малых субъединиц [18, 19]. Несмотря на наличие RRM в большой субъединице, за связывание с РНК отвечает малая субъединица, которая содержит “испор-

ченный” домен Nudix-гидролазы, получивший взамен возможность связываться с РНК. Мишенью CFIm являются USE, содержащие пентамеры UGUAN. Присутствие двух таких элементов на определенном расстоянии друг от друга также позволяет CFIm выпетливать этот участок РНК, и эта его способность лежит, вероятно, в основе механизма выбора альтернативных сайтов полиаденилирования [20]. CFIm обладает большим сродством к дистальному PAS, что показано в опытах с нокдаун-ом CFIm-68. Снижение экспрессии этого белка часто наблюдается в раковых клетках, инвазивность которых при этом повышается [8]. CFIm также играет важную роль в образовании структуры комплекса 3'-процессинга и способен самостоятельно привлекать PAP. Среди других свойств CFIm можно отметить участие в экспорте молекул из ядра (показано для CFIm-68) [8]. CFIm состоит из двух субъединиц, Pcf11 и hClp1 [14]. Pcf11 может связываться с CTD RNAP II, образуя мостик между CTD и транскриптом, что приводит к разборке элонгационного комплекса, т.е. Pcf11 ответственна за терминацию транскрипции — это подтверждено мутационным анализом [21]. Не так давно нашли еще один белок комплекса 3'-процессинга, названный Rbbp6. Он работает в ассоциации с CstF и также участвует в процессах выбора сайта полиаденилирования, особенно в пре-мРНК с AU-богатыми 3'-НТР. Нокдаун Rbbp6 снижает

синтез белков, чьи мРНК имеют такой 3'-НТР, а также способствует выбору более дистального PAS в пре-мРНК [9].

PAP содержит лишь одну субъединицу, и ее каталитический домен расположен в N-концевой части. В С-концевой части находятся регуляторные элементы. К примеру, гиперфосфорилирование С-концевого домена PAP — один из основных способов остановки синтеза белка во время митоза. В средней части белка имеется РНК-связывающий домен, но он взаимодействует с РНК лишь неспецифически, для правильной же работы фермента в комплексе 3'-процессинга особенно важны CPSF и CFIm. Когда поли(А)-хвост, синтезируемый PAP, достигает определенной длины, с ним связывается поли(А)-связывающий белок PABPN1, благодаря чему активность PAP существенно возрастает. PABPN1 отвечает и за терминацию синтеза поли(А)-хвоста, а также принимает участие в экспорте мРНК из ядра [14]. В цитоплазме с поли(А)-хвостом связывается PABPC1, который в процессе трансляции взаимодействует с фактором элонгации eIF4G и некоторыми другими белками, обеспечивая циклизацию мРНК [22].

Несмотря на важную роль факторов 3'-процессинга в поддержании структуры комплекса, центральную позицию в нем занимает CTD наибольшей субъединицы RNAP II (Rpb1). CTD содержит 52 повтора из семи аминокислотных остатков с консенсусной последовательностью YSPTSPS, профиль фосфорилирования аминокислот обуславливает возможность взаимодействия с различными белками, участвующими в процессе транскрипции [23]. Еще в начале транскрипции (на стадии кепирования) с CTD связываются CPSF (CPSF160) и CstF (CstF50 и 64) [24, 25], причем есть данные о том, что CPSF рекрутируется при помощи TFIIID. CFIm также связывается с CTD еще до окончания элонгации [24–26]. О роли CFIm в связывании с CTD сказано выше, за это взаимодействие отвечает первый остаток серина в повторе YSPTSPS [21]. Итак, непосредственно перед началом 3'-процессинга на CTD наибольшей субъединицы RNAP II уже располагаются CPSF, CstF и CFIm. По мере синтеза транскрипта эти три фактора находят на РНК свои целевые последовательности и связываются с ними. Позже они привлекают CFII и PAP, а затем CPSF73 разрезает транскрипт [14, 25]. По-видимому, к моменту начала полиаденилирования все факторы 3'-процессинга, кроме необходимых для этого CPSF и PAP, диссоциируют от комплекса. PAP медленно синтезирует поли(А)-участок из 11–15 н., с которым уже может связываться PABP. По мере роста поли(А)-хвоста до 200–300 н. он сам, связанные с ними молекулы PABP и одна молекула PAP на его конце формируют петлеобразную структуру. PABP и PAP взаимодействуют с CPSF, что останавливает синтез поли(А)-хвоста. В

то же время 3'-часть бывшей пре-мРНК деградируется экзонуклеазой Xrn2, которую привлекает фактор PSF, также, по всей видимости, ассоциированный с CTD RNAP II [3, 27]. Сам комплекс 3'-процессинга в конце концов разбирается факторами экспорта мРНК из ядра, причем есть сведения, что в экспорте может участвовать CFIm68 [28].

Такой процесс, как сплайсинг, также сопряжен с 3'-процессингом. В генах млекопитающих последний экзон намного длиннее прочих, а механизм его сплайсинга сильно отличается от механизма сплайсинга остальных экзонов. Один из факторов сплайсинга, U2AF65, способен рекрутировать CFIm59/25. Известно, что U2AF65 может стимулировать полиаденилирование транскриптов CPSF-независимым путем, вероятнее всего, посредником в этом процессе является CFIm59/25. Такое взаимодействие может быть особенно важным в случае слабых сайтов полиаденилирования. Намного менее описанный механизм эффективен при вовлечении наиболее распространенной формы CFIm68/25 [29]. Известно, что 3'-процессинг стимулируется взаимодействием CPSF с малой ядерной РНК U2, компонентом сплайсосомы [14].

Характерное для всех клеток полиаденилирование пре-мРНК в ядре — не единственный пример полиаденилирования транскриптов RNAP II. В некоторых случаях сильно отличающийся аппарат полиаденилирования работает в цитоплазме. мРНК, прошедшая все необходимые посттранскрипционные модификации в ядре, после экспорта из него может быть деаденилирована. Фермент, ответственный за это, 3'-эксонуклеаза PARN, оставляет лишь небольшую часть поли(А)-хвоста [30]. Такая РНК как бы запасается клеткой, находится в неактивном, “спящем”, состоянии и может быть мобилизована в определенный момент. Этот процесс, происходящий в некоторых типах клеток, впервые был описан в ооцитах *Xenopus laevis*. Он характерен, в первую очередь, для ооцитов и ранних эмбрионов различных животных — от насекомых до млекопитающих. РНК, которая подвергается цитоплазматическому полиаденилированию, должна иметь PAS, с которым, как и при ядерном полиаденилировании, связывается CPSF. Основная последовательность, специфичная для такой мРНК, — элемент CPE (cytoplasmic polyadenylation element); известна по меньшей мере еще одна вспомогательная последовательность, которая взаимодействует с поли(С)-связывающим белком PCBP2. CPE опознается CPE-связывающими белками CPEB, которые содержат домен цинковых пальцев и RRM-домены. Мишенью самого важного из этих белков, CPEB1, служит последовательность UUUUAU или UUUUAU. Белки CPEB регулируют весь процесс цитоплазматического полиаденилирования, в частности опосредуют работу специфической PAPD4 (GLD2). CPEB способны связываться с целевой мРНК

еще в ядре. Если N-концевой домен белка CPEB не фосфорилирован, то он поддерживает мРНК в деаденилированном состоянии, в ином случае он инициирует цитоплазматическое полиаденилирование, что означает перевод мРНК в транслируемое состояние. Работе фосфорилированного CPEB способствует белок PUM2 (Pumilio), который имеет на мРНК собственный сайт связывания UGUANUAU и особенно важен в случае неканонических CPE [22, 31]. После деаденилирования экзонуклеаза PARN остается связанной с мРНК, на ней же находятся CPEB и CPSF, а также белок, подавляющий трансляцию — маскин. Симплекин CPSF работает в качестве белка скаффолда, поддерживающего все эти факторы. В ответ на фосфорилирование CPEB PAPD4 привлекается из цитоплазмы, а деаденилаза исключается из комплекса, благодаря чему и осуществляется полиаденилирование [32].

Некоторые мРНК, лишённые CPE, могут полиаденилироваться в цитоплазме при помощи менее изученной системы — элементов MBE (MSI-binding elements) в мРНК и взаимодействующих с ними белков Musashi (MSI1 и MSI2). Такой процесс характерен для раннего эмбриогенеза [33]. Около 10% мРНК соматических клеток также могут подвергаться цитоплазматическому полиаденилированию. Показано, что функционирование и даже регуляция CPEB1 и PAPD4 сходны в ооцитах и дендритах, где они, вероятно, нужны при ремоделировании синапсов в процессе научения. Не исключено, что главную роль здесь могут играть другие белки CPEB, во всяком случае, у мышей в этом процессе участвует CPEB3. Известно также, что CPEB1 может быть вовлечен в такие важные процессы, как регуляция митоза, клеточное старение и образование опухолей. Например, старение фибробластов останавливается при выключении CPEB1, при этом сам фактор взаимодействует с белком p53 [22].

AAUAAA-ЗАВИСИМОЕ ПОЛИАДЕНИЛИРОВАНИЕ ТРАНСКРИПТОВ, СИНТЕЗИРУЕМЫХ РНК-ПОЛИМЕРАЗой III

RNAP III — одна из трех основных РНК-полимераз эукариот [34], специализируется на транскрипции генов коротких РНК, не кодирующих белки. (Отметим, что RNAP II синтезирует также не только мРНК, но и различные некодирующие короткие РНК.) RNAP III синтезирует многочисленные тРНК, 5S рРНК, U6, 7SL, 7SK, Y, vt (vault RNA), вирусные РНК (VA1, EBER1 и EBER2) [35]. Эти РНК играют важную роль в синтезе белков, а также в транскрипции, сплайсинге и репликации ДНК. Ни одна из них не подвергается классическому полиаденилированию, что не вызывает удивления по следующим причинам. Во-первых,

очень сильно отличаются структура и регуляция генов, транскрибируемых RNAP II и RNAP III. У последних, как правило, промотор расположен не перед геном, а внутри него [34, 36]. Так, например, в генах тРНК промотор состоит из двух 11-нуклеотидных участков (боксы А и В), причем бокс А расположен в 10–12 н. от точки начала транскрипции, тогда как бокс В отстоит на 33–35 н. от бокса А по ходу транскрипции [36]. Для ассоциации RNAP III с геном тРНК и инициации его транскрипции необходимы лишь два фактора (TFIIIB и TFIIIC) [34, 36]. Терминация транскрипции RNAP III происходит на блоке, состоящем из пяти или более остатков Т, причем фермент диссоциирует от гена по завершении транскрипции первых двух–трех остатков Т [37]. Другими словами, формирование 3'-конца у первичных транскриптов RNAP III (в отличие от транскриптов RNAP II) не требует участия гексамера AATAAA, DSE или USE и никак не связано с разрезанием цепи РНК. Наконец, у RNAP III нет домена, структурно и функционально сходного с CTD большой субъединицы RNAP II, который привлекает CPSF в транскрипционный комплекс и тем самым определяет работу полиаденилирующей “машины”. Таким образом, остается неясным, как в отсутствие такого CTD RNAP III могла бы запускать полиаденилирование своих транскриптов.

В середине 80-х годов многое из описанного еще не было известно. В то время две группы исследователей независимо создали сходные конструкции, которые состояли из промотора RNAP III гена VA1 РНК аденовируса и гена тимидинкиназы герпесвируса без собственного промотора RNAP II [38, 39]. Предполагалось, что такая химерная конструкция должна транскрибироваться RNAP III, включая ген тимидинкиназы. Утверждалось [38], что многие образующиеся транскрипты разрезались там, где должна разрезаться мРНК, и правильно полиаденилировались. Однако в работе [39] пришли к противоположному выводу: мРНК-подобная РНК, транскрибированная RNAP III, не полиаденилировалась, хотя и содержала сигналы, необходимые и достаточные для ее 3'-концевого разрезания и полиаденилирования в том случае, если бы транскрипцию осуществляла RNAP II. По-видимому, справедлив именно этот вывод — есть косвенные свидетельства в пользу того, что в работе [38] наблюдали полиаденилирование РНК, транскрибированных со случайных (скрытых) промоторов RNAP II [39]. В 1995 г. была проведена еще одна сходная работа — ген, кодирующий белок Tat HIV-1 и лишенный собственного промотора RNAP II, был вставлен внутрь гена VA1 ниже его промотора RNAP III [40]. Было ясно показано, что РНК, транскрибированная RNAP III с этого химерного гена, не подвергалась ни кепированию, ни полиаденилированию. Та-

ким образом, большинство результатов свидетельствовало о том, что транскрибированные RNAP III РНК, в норме синтезируемые RNAP II, не способны полиаденилироваться.

В начале 80-х годов при участии одного из нас (Д.К.) были открыты особые повторяющиеся последовательности ДНК [41–44], получившие название SINE (Short Interspersed Elements) [45]. Оказалось, что по геному мыши рассеяно огромное число $((1-5) \times 10^5)$ копий SINE двух семейств – B1 и B2 [41, 46]. В геноме человека число копий SINE Alu достигает 1.1 млн. [47]. Со временем стало ясно, что SINE имеются у подавляющего большинства многоклеточных организмов [48]. Те или иные семейства SINE (в зависимости от момента их возникновения в эволюции) свойственны различным родственным группам организмов (например, B2 найден только у грызунов четырех семейств, а Alu – у всех приматов). Копии SINE, образующие семейство, не идентичны и, как правило, сходство между их нуклеотидными последовательностями лежит в пределах 75–95%. Длина SINE большинства семейств составляет 150–300 п.н. [48]. SINE представляют собой неавтономные мобильные генетические элементы, использующие для своего распространения в геноме обратную транскриптазу, кодируемую мобильными элементами другого типа – LINE (Long Interspersed Elements) [49, 50]. Подавляющее число семейств ведет свое происхождение от тех или иных видов тРНК [48, 51], но есть и такие (B1 и Alu), которые возникли из молекул 7SL РНК [48, 52]. Небольшой третий класс SINE эволюционно связан с 5S рРНК [48, 53]. Во всех случаях последовательности, произошедшие от малых РНК, расположены в “головной” части SINE, в пределах которой находится промотор RNAP III [48, 54]. В SINE, родственных тРНК и 7SL РНК, промотор состоит из боксов А и В, которые упоминались в начале этого раздела. За “головой” SINE следует “тело” и, наконец, “хвост”. Нуклеотидные последовательности тел у большинства семейств SINE не сходны между собой, и их происхождение не известно. У очень многих SINE хвост представляет собой А-богатую последовательность, т.е. в нем сильно преобладают адениловые остатки (это особенно характерно для плацентарных млекопитающих, где большинство SINE используют обратную транскриптазу, кодируемую LINE-1) [48, 50, 55]; у других семейств SINE хвосты образованы tandemными повторами длиной 3–5 п.н. [48, 49].

В 80-х годах в лаборатории Г.П. Георгиева обнаружили, что в опухолевых клетках мышей присутствует большое количество поли(А)-содержащих РНК длиной от 200 до 500 н., транскрибированных RNAP III с SINE B2 [56–60]. К сходным выводам пришли и другие авторы [61]. Замечено, что А-богатый хвост B2 обычно содержит от одного до четырех гексамеров AATAAA, а на конце

хвоста находится терминатор транскрипции TCTTTT. Показано, что транскрипция RNAP III начинается на первом нуклеотиде B2 и заканчивается на терминаторе, в результате чего образуется РНК длиной около 180 н. Присутствие гексамера(ов) AAUAAA рядом с 3'-концом такой B2 РНК наводило на мысль, что она может полиаденилироваться, хотя синтезируется RNAP III, а не RNAP II. С этой гипотезой хорошо согласовывались следующие факты: 1. устойчивость транскрипции B2 РНК в клетках со средой, содержащей ингибитор RNAP II, α -аманитин (50 мкг/мл); 2. наличие поли(А)-хвостов у B2 РНК, обнаруженных в опытах по клонированию и секвенированию кДНК; 3. гетерогенность длины B2 РНК, обусловленная вариабельностью длины поли(А), предположительно синтезируемой PAP.

Убедительность этих данных несколько снижало то, что опыты проводили на суммарных препаратах клеточной РНК, содержащих транскрипты множества копий B2. Некоторые копии B2 имели неполноценные (укороченные или мутированные) терминаторы, и их транскрипция продолжалась далеко за конец B2. По этой причине образующиеся B2 РНК также могли быть гетерогенными по длине. Кроме того, удлиненные B2 РНК за А-богатым хвостом обычно содержат фрагмент “чистого” поли(А) длиной 15–20 н., который входит в состав самого SINE и образуется в результате транскрипции. Все это усложняло картину и несколько затрудняло однозначную интерпретацию. Так или иначе, способность некоторых транскриптов, синтезируемых RNAP III, подвергаться AAUAAA-зависимому полиаденилированию в то время не нашла своего признания.

На рубеже веков мы (О.Б. и Д.К.) выделили ряд тРНК-родственных SINE из геномов плацентарных млекопитающих [62–65]. Проанализировав их нуклеотидные последовательности вместе с последовательностями ранее описанных SINE этой группы животных, пришли к заключению, что все SINE с А-богатыми хвостами, исходя из особенностей организации этих хвостов, можно разделить на два класса [64]. Один класс обозначили T+; А-богатый хвост таких SINE содержит: (1) один или несколько гексамеров AATAAA; (2) терминатор транскрипции RNAP III (TCTTTT или TTTTTT); и (3) сегмент поли(А)_{10–25}, находящийся на самом конце SINE (рис. 2). Именно к этому классу относится SINE B2, рассмотренный выше. Надо отметить, что в некоторых геномных копиях SINE гексамеры AATAAA и/или терминаторы транскрипции повреждены мутациями, и это первоначально несколько затруднило обособление класса T+. SINE, А-богатым хвостам которых гексамеры AATAAA и терминаторы RNAP III вообще несвойственны, были выделены в класс T-. РНК, транскрибированные RNAP III с таких SINE, не должны полиаденилироваться из-за отсутствия

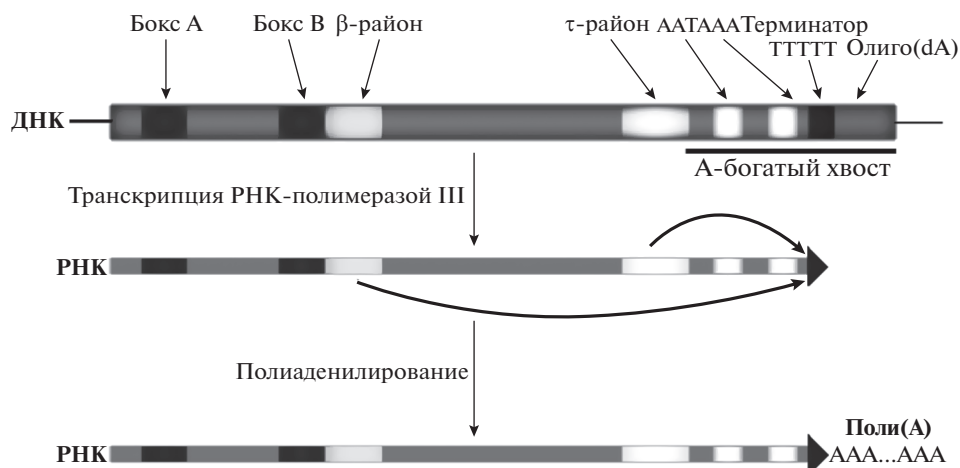


Рис. 2. Схема, иллюстрирующая структуру SINE класса T+, его транскрипцию RNAP III и полиаденилирование образующейся РНК. Дугообразные стрелки символизируют стимулирующее действие бета- и тау-районов на полиаденилирование транскрипта. Остальные объяснения в тексте.

гексамера(ов) AAUAAA. Кроме того, 3'-концы образующихся РНК будут располагаться довольно далеко от хвоста SINE, поскольку в этих случаях транскрипция терминируется на ТТТТТ-блоках, встречающихся в прилежащих к SINE последовательностях.

На новом этапе изучения полиаденилирования РНК, транскрибированных RNAP III, применили подход, позволяющий работать с единичными копиями SINE [66]. Клонированную в плазмидном векторе геномную копию SINE вводят (трансформная трансфекция) в клетки. Через сутки выделяют суммарную клеточную РНК, разделяют ее в денатурирующем полиакриламидном геле (в первых опытах — в агарозном) и подвергают Нозерн-гибридизации, используя в качестве зонда меченую ^{32}P цепь ДНК изучаемого SINE. РНК SINE детектируют с помощью рентгеновской пленки (качественно) и фосфоимеджера (количественно). Полиаденилирование РНК проявляется как РНК SINE (электрофоретическая полоса), удлиненная на 20–300 н. (гетерогенный материал). Первоначально такие опыты провели с копией B2 и с помощью ряда подходов подтвердили, что вся детектируемая РНК, транскрибируется RNAP III, а удлиненная B2 РНК содержит протяженные поли(А)-хвосты [66]. Замена Т на С в обоих гексамерах ААТААА, присутствующих в данной копии B2, приводила к полному исчезновению удлиненной формы РНК, доказывая, что полиаденилирование B2 РНК зависит от AAUAAA [66]. Такие опыты провели с восемью SINE класса T+ и показали, что RNAP III-транскрипты каждого из них способны к AAUAAA-зависимому полиаденилированию [67].

Поскольку РНК, транскрибированные RNAP III с SINE класса T–, *a priori* не способны к полиаде-

нированию, структура семи из них была изменена следующим образом [67, 68]: в А-богатые хвосты этих SINE ввели гексамер(ы) ААТААА, а на самый конец добавили терминатор транскрипции ТТТТТТ. Казалось бы, измененные таким образом SINE класса T– приобретают структуру, свойственную SINE класса T+, и их RNAP III-транскрипты должны быть способны к полиаденилированию. Однако это оказалось не так — опыты по трансфекции клеток HeLa модифицированными SINE класса T– не выявляли полиаденилирования их транскриптов [67]. Это свидетельствовало о том, что SINE класса T+ содержат, помимо ААТААА, какие-то нуклеотидные последовательности, необходимые для полиаденилирования и отсутствующие в SINE класса T–. Создав производные трех SINE класса T+ (B2 мыши, Dip тушканчика и Ves летучей мыши), несущие различные делеции и нуклеотидные замены, мы пришли к выводу, что в этих SINE имеются два района, помимо А-богатого хвоста, важные для полиаденилирования их транскриптов [67] (рис. 2). Первый из них, обозначенный бета, располагается сразу за боксом В промотора (возможно, даже частично перекрывается с ним). По предварительным данным его длина составляет 15–20 п.н. Второй район, названный тау, находится перед А-богатым хвостом, несущим сигналы полиаденилирования (рис. 2). Во многих (7 из 13) SINE класса T+, в том числе Dip и Ves, этот район состоит почти исключительно из остатков С и Т, т.е. представляет собой полипиримидиновую последовательность длиной 30–50 н. [48, 64, 67]. В B2 тау-район не обогащен пиримидиновыми нуклеотидами, а его длина не превышает 18 н. Удаление или замена бета- или тау-районов снижает эффективность полиаденилирования РНК, транскрибированных

RNAP III с SINE, тогда как инактивация обоих этих районов полностью прекращает синтез поли(А)-хвоста [67]. Можно предположить, что с бета- и тау-районами транскриптов SINE связываются какие-то белковые комплексы, активирующие полиаденилирующую “машину”, ассоциированную с гексамером AAUAAA [67]. Вероятно, можно увидеть аналогию между бета- и тау-районами РНК SINE и USE – вспомогательными элементами, расположенными выше сигналов полиаденилирования в пре-мРНК. В качестве рабочей гипотезы можно рассматривать связывание с некоторыми бета- и тау-районами белкового комплекса CFIm, взаимодействующего в случае пре-мРНК с USE и стимулирующего разрезание этой РНК и ее полиаденилирование. По предварительным данным (И. Устьянцев) с тау-районами, образованным полипиримидиновой последовательностью, в клетке связывается особый белок РТВ (polyurimidine tract-binding protein), который, вероятно, вовлечен в полиаденилирование RNAP III-транскриптов SINE. Интересно, что белок РТВ никогда не выявляли как непосредственный участник комплекса полиаденилирования мРНК. Скорее всего, несмотря на AAUAAA-зависимый характер полиаденилирования RNAP III-транскриптов SINE, механизм этого процесса будет значительно отличаться от механизма полиаденилирования мРНК.

Поскольку SINE класса Т+ возникли независимо в разных ветвях филогенетического древа млекопитающих [69], можно сделать вывод о том, что способность RNAP III-транскриптов подвергаться полиаденилированию развилась у SINE благодаря параллельной эволюции. В заключение отметим, что присутствие поли(А) на конце транскриптов SINE удлиняет время жизни этих РНК в клетке [66] и способствует их участию в процессе обратной транскрипции, в результате которого в геноме образуются новые копии SINE [55]. Возможно, в будущем у полиаденилированных РНК, транскрибированных RNAP III с SINE, обнаружатся и какие-то другие клеточные функции.

НЕКАНОНИЧЕСКИЕ ПОЛИ(А)-ПОЛИМЕРАЗЫ И ПОЛИ(А)- ОПОСРЕДОВАННАЯ ДЕГРАДАЦИЯ РНК

Нематричное добавление адениловых нуклеотидов может не только продлевать время жизни эукариотических РНК, но и способствовать ее деградации. Этот феномен описали лишь в 2000 г., когда показали, что при мутациях в ядерной экзосоме – комплексе, ответственном за деградацию различных дефектных или ненужных РНК, – в клетках дрожжей накапливаются полиаденилированные формы отдельных малых РНК, а именно ядерной U4 РНК и малых ядрышковых РНК [70]. Позже выявили полиаденилирование и других

РНК-субстратов, а затем выделили обладающий полиаденилирующей активностью фермент Trf4p, мутации в котором приводили к увеличению времени жизни определенных тРНК [71]. Trf4p является неканонической РАР, он принадлежит к суперсемейству нуклеотидилтрансфераз бета-типа, в которое также входят ферменты, добавляющие ССА на 3'-конец тРНК, и цитоплазматическая РАР Gld2, упомянутая ранее [72]. Оказалось, что подготовкой дефектных РНК к деградации в экзосоме занимается целый комплекс, куда Trf4p входит как один из компонентов. Этот комплекс был назван TRAMP – как сокращение от имен трех его составляющих – РАР Trf4p, РНК-связывающего белка Air2p и хеликазы Mtr4p [73]. Субстратами TRAMP служат тРНК, принявшие неправильную вторичную структуру, малые ядрышковые РНК, малые ядерные РНК, такие как U4, U14 и U6, 5S рРНК, скрытые нестабильные транскрипты (Cryptic Unstable Transcripts, CUT) [74], а также большие рРНК и их предшественники – 25S рРНК и 27S пре-рРНК [75]. Более того, TRAMP может котранскрипционно привлекаться к пре-мРНК, где стимулирует сплайсинг, взаимодействуя непосредственно с его факторами, а затем направляет сплайсированные интроны в экзосому, но в этом пути полиаденилирующая активность роли не играет [76, 77]. В составе комплекса Trf4p может замещаться своим паралогами Trf5p, и есть свидетельства в пользу того, что данная разновидность комплекса опознает несколько другой спектр целевых РНК, хотя большая их часть поступает в TRAMP, содержащий Trf4p [76]. В настоящее время известно, что TRAMP скорее не поли-, а олигоаденилирует целевые РНК, добавляя, по крайней мере в некоторых случаях, лишь три–четыре остатка аденозина [78]. Выяснилось также, что основу комплекса составляют Trf4p и белок с “цинковыми шарнирами” Air2p, а Mtr4p отвечает за транспорт олигоаденилированной РНК к экзосоме. Mtr4p участвует во многих других путях экзосомной деградации, а не только в опосредованных TRAMP. Недавно был предложен механизм работы TRAMP (рис. 3), согласно которому целевая РНК вначале опознается Mtr4p и Air2p, затем 3'-концом проходит в канал внутри Mtr4p, одновременно расплетаясь, что подводит ее к активному центру Trf4p, где осуществляется олигоаденилирование, за которым следует транспортировка в экзосому [74].

В клетках человека найдены два белковых комплекса, схожих с TRAMP. В качестве общего компонента они содержат белок hMtr4p, гомолог дрожжевого белка, но различаются по локализации и составу. Комплекс, названный NEXT (Nuclear Exosome Targeting Complex), обнаружен в ядре вне ядрышек и содержит не только hMtr4p, но также белок с “цинковыми шарнирами” ZCCHC8 и RRM-содержащий белок RBM7. NEXT отвечает

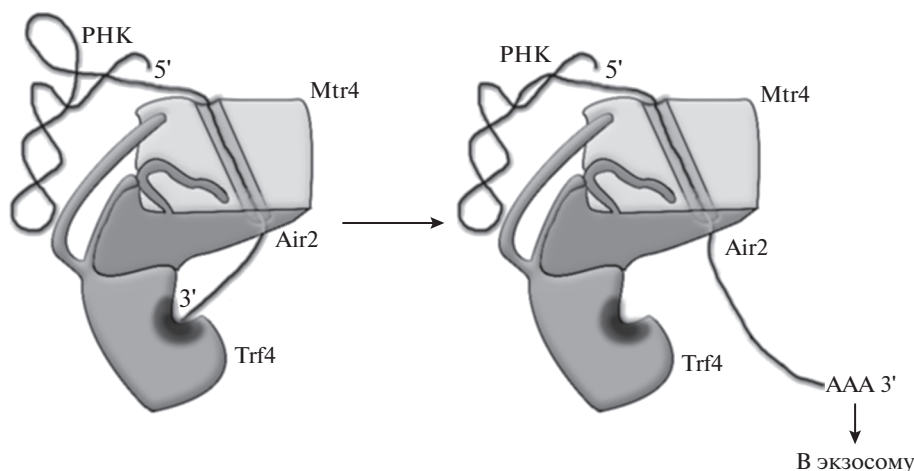


Рис. 3. Схема TRAMP-комплекса дрожжей, узнающего и связывающего РНК с нарушенной структурой. РНК подвергается олигоденилированию субъединицей Trf4, что стимулирует ее деградацию экзосомой.

за направление в экзосому главным образом вышедших промоторных транскриптов (Promoter Upstream Transcripts, PROMPT) – аналогов CUT дрожжей, и, что важно, не аденилирует свои субстраты [79, 80]. В ядрышках работает другой комплекс, более похожий на TRAMP и названный hTRAMP, в его состав, кроме hMtr4p, входят белки PAPD5 (hTRF4-2) и ZCCHC7 (hAIR2). Этот комплекс проявляет аденилирующую активность и опосредует деградацию в экзосоме неправильно свернутых пре-рРНК. Показано также, что при нокдауне неканонической PAPD5 возрастает концентрация онкогенной микроРНК miR-21 [79, 81].

Если у эукариот роль полиаденилирования как механизма деградации, а не стабилизации РНК, описана сравнительно недавно, то давно показано, что у прокариот полиаденилирование маркирует именно деградирующие РНК. В клетках *Escherichia coli* обнаружены PAPI и полинуклеотидфосфорилаза (PNPase), которая добавляет случайную последовательность нуклеотидов, лишь наполовину состоящую из аденинов. Поли(А)-хвосты, синтезируемые PAPI, довольно небольшие, обычно от 15 до 30 н., причем их длина может зависеть от типа РНК. Полиаденилированию в *E. coli* подвергаются и кодирующие (чаще с Rho-независимыми терминаторами), и не кодирующие РНК (не менее 2% от всей клеточной РНК), особенно дефектные, и в 90% случаев эту реакцию осуществляет PAPI. Такие РНК подвергаются более быстрой деградации в противоположность каноническому полиаденилированию в клетках эукариот. Поли(А)-хвост у прокариотической РНК может опознаваться РНКазой E в составе деградосомы – комплекса, функционально схожего с экзосомой, а также, возможно, РНКазой R. Значительное повышение концентрации PAPI токсич-

но для бактериальной клетки из-за уничтожения жизненно важных РНК, либо из-за перерасхода АТФ. Интересно, что в митохондриях также имеются свои PAP, и длина поли(А)-хвоста в митохондриальных РНК достигает 40–60 н. Вероятнее всего, он также отмечает РНК, нуждающуюся в деградации [82].

Данный обзор написан при финансовой поддержке программы Президиума РАН “Молекулярная и клеточная биология” и Российского фонда фундаментальных исследований (14-04-00616-а).

СПИСОК ЛИТЕРАТУРЫ

1. Mandel C.R., Bai Y., Tong L. (2008) Protein factors in pre-mRNA 3'-end processing. *Cell. Mol. Life Sci.: CMLS.* **65**, 1099–1122.
2. Proudfoot N.J. (2011) Ending the message: poly(A) signals then and now. *Genes Dev.* **25**, 1770–1782.
3. Wigington C.P., Williams K.R., Meers M.P., Bassell G.J., Corbett A.H. (2014) Poly(A) RNA-binding proteins and polyadenosine RNA: new members and novel functions. *Wiley interdisciplinary Rev. RNA.* **5**, 601–622.
4. Manley J.L. (1988) Polyadenylation of mRNA precursors. *Biochim. Biophys. Acta.* **950**, 1–12.
5. Beaudoin E., Freier S., Wyatt J.R., Claverie J.M., Gautheret D. 2000. Patterns of variant polyadenylation signal usage in human genes. *Genome Res.* **10**, 1001–1010.
6. Tian B., Hu J., Zhang H., Lutz C.S. (2005) A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucl. Acids Res.* **33**, 201–212.
7. Edwalds-Gilbert G., Veraldi K.L., Milcarek C. (1997) Alternative poly(A) site selection in complex transcription units: means to an end? *Nucl. Acids Res.* **25**, 2547–2561.
8. Martin G., Gruber A.R., Keller W., Zavolan M. (2012) Genome-wide analysis of pre-mRNA 3' end processing

- reveals a decisive role of human cleavage factor I in the regulation of 3' UTR length. *Cell Repts.* **1**, 753–763.
9. Shi Y., Manley J.L. (2015) The end of the message: multiple protein–RNA interactions define the mRNA polyadenylation site. *Genes Dev.* **29**, 889–897.
 10. Chou Z.F., Chen F., Wilusz J. 1994. Sequence and position requirements for uridylyte-rich downstream elements of polyadenylation signals. *Nucl. Acids Res.* **22**, 2525–2531.
 11. Venkataraman K., Brown K.M., Gilmartin G.M. (2005) Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition. *Genes Dev.* **19**, 1315–1327.
 12. Chen F., MacDonald C.C., Wilusz J. (1995) Cleavage site determinants in the mammalian polyadenylation signal. *Nucl. Acids Res.* **23**, 2614–2620.
 13. Shi Y., Di Giammartino D.C., Taylor D., Sarkeshik A., Rice W.J., Yates J.R., 3rd, Frank J., Manley J.L. (2009) Molecular architecture of the human pre-mRNA 3' processing complex. *Mol. Cell.* **33**, 365–376.
 14. Chan S., Choi E.A., Shi Y. (2011) Pre-mRNA 3'-end processing complex assembly and function. *Wiley interdisciplinary Rev. RNA.* **2**, 321–335.
 15. Mandel C.R., Kaneko S., Zhang H., Gebauer D., Vethantham V., Manley J.L., Tong L. (2006) Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature.* **444**, 953–956.
 16. Bai Y., Auperin T.C., Chou C.Y., Chang G.G., Manley J.L., Tong L. (2007) Crystal structure of murine CstF-77: dimeric association and implications for polyadenylation of mRNA precursors. *Mol. Cell.* **25**, 863–875.
 17. Murthy K.G., Manley J.L. (1992) Characterization of the multisubunit cleavage-polyadenylation specificity factor from calf thymus. *J. Biol. Chem.* **267**, 14804–14811.
 18. Kim S., Yamamoto J., Chen Y., Aida M., Wada T., Handa H., Yamaguchi Y. (2010) Evidence that cleavage factor Im is a heterotetrameric protein complex controlling alternative polyadenylation. *Genes Cells.* **15**, 1003–1013.
 19. Yang Q., Gilmartin G.M., Double S. (2010) Structural basis of UGUA recognition by the Nudix protein CFI(m)25 and implications for a regulatory role in mRNA 3' processing. *Proc. Natl. Acad. Sci. USA.* **107**, 10062–10067.
 20. Yang Q., Gilmartin G.M., Double S. (2011) The structure of human cleavage factor I(m) hints at functions beyond UGUA-specific RNA binding: a role in alternative polyadenylation and a potential link to 5' capping and splicing. *RNA Biol.* **8**, 748–753.
 21. Zhang Z., Gilmour D.S. (2006) Pcf11 is a termination factor in *Drosophila* that dismantles the elongation complex by bridging the CTD of RNA polymerase II to the nascent transcript. *Mol. Cell.* **21**, 65–74.
 22. Charlesworth A., Meijer H.A., de Moor C.H. (2013) Specificity factors in cytoplasmic polyadenylation. *Wiley Interdisciplinary Rev. RNA.* **4**, 437–461.
 23. Собенникова М.В., Шематорова Е.К., Шпаковский Г.В. 2007. С-концевой домен (CTD) субъединицы Rpb1 ядерной РНК-полимеразы II и его роль в транскрипционном цикле. *Молекуляр. биология.* **41**, 433–449.
 24. McCracken S., Fong N., Yankulov K., Ballantyne S., Pan G., Greenblatt J., Patterso S.D., Wickens M., Bentley D.L. (1997) The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature.* **385**, 357–361.
 25. Hirose Y., Manley J.L. (2000) RNA polymerase II and the integration of nuclear events. *Genes Dev.* **14**, 1415–1429.
 26. Dantoni J.C., Murthy K.G., Manley J.L., Tora L. (1997) Transcription factor TFIID recruits factor CPSF for formation of 3' end of mRNA. *Nature.* **389**, 399–402.
 27. Kaneko S., Rozenblatt-Rosen O., Meyerson M., Manley J.L. (2007) The multifunctional protein p54nrb/PSF recruits the exonuclease XRN2 to facilitate pre-mRNA 3' processing and transcription termination. *Genes Dev.* **21**, 1779–1789.
 28. Ruepp M.D., Aringhieri C., Vivarelli S., Cardinale S., Paro S., Schumperli D., Barabino S.M. (2009) Mammalian pre-mRNA 3' end processing factor CF I m 68 functions in mRNA export. *Mol. Biol. Cell.* **20**, 5211–5223.
 29. Millevoi S., Loulergue C., Dettwiler S., Karaa S.Z., Keller W., Antoniou M., Vagner S. (2006) An interaction between U2AF 65 and CF I(m) links the splicing and 3' end processing machineries. *EMBO J.* **25**, 4854–4864.
 30. Kim J.H., Richter J.D. 2006. Opposing polymerase-deadenylase activities regulate cytoplasmic polyadenylation. *Mol. Cell.* **24**, 173–183.
 31. Villalba A., Coll O., Gebauer F. (2011) Cytoplasmic polyadenylation and translational control. *Curr. Opin. Genet. Dev.* **21**, 452–457.
 32. Norbury C.J. (2013) Cytoplasmic RNA: a case of the tail wagging the dog. *Nature Rev. Mol. Cell Biol.* **14**, 643–653.
 33. Charlesworth A., Wilczynska A., Thampi P., Cox L.L., MacNicol A.M. (2006) Musashi regulates the temporal order of mRNA translation during *Xenopus* oocyte maturation. *EMBO J.* **25**, 2792–2801.
 34. Schramm L., Hernandez N. (2002) Recruitment of RNA polymerase III to its target promoters. *Genes Dev.* **16**, 2593–2620.
 35. Dieci G., Fiorino G., Castelnovo M., Teichmann M., Pagano A. (2007) The expanding RNA polymerase III transcriptome. *Trends Genet.: TIG.* **23**, 614–622.
 36. Orioli A., Pascali C., Pagano A., Teichmann M., Dieci G. (2012) RNA polymerase III transcription control elements: themes and variations. *Gene.* **493**, 185–194.
 37. Arimbasseri A.G., Rijal K., Maraia R.J. (2013) Transcription termination by the eukaryotic RNA polymerase III. *Biochim. Biophys. Acta.* **1829**, 318–330.
 38. Lewis E.D., Manley J.L. (1986) Polyadenylation of an mRNA precursor occurs independently of transcription by RNA polymerase II *in vivo*. *Proc. Natl. Acad. Sci. USA.* **83**, 8555–8559.
 39. Sisodia S.S., Sollner-Webb B., Cleveland D.W. (1987). Specificity of RNA maturation pathways: RNAs transcribed by RNA polymerase III are not substrates for

- splicing or polyadenylation. *Mol. Cell. Biol.* **7**, 3602–3612.
40. Gunnery S., Mathews M.B. (1995) Functional mRNA can be generated by RNA polymerase III. *Mol. Cell. Biol.* **15**, 3597–3607.
 41. Kramerov D.A., Grigoryan A.A., Ryskov A.P., Georgiev G.P. (1979) Long double-stranded sequences (ds-RNA-B) of nuclear pre-mRNA consist of a few highly abundant classes of sequences: evidence from DNA cloning experiments. *Nucl. Acids Res.* **6**, 697–713.
 42. Krayev A.S., Kramerov D.A., Skryabin K.G., Ryskov A.P., Bayev A.A., Georgiev G.P. (1980) The nucleotide sequence of the ubiquitous repetitive DNA sequence B1 complementary to the most abundant class of mouse fold-back RNA. *Nucl. Acids Res.* **8**, 1201–1215.
 43. Krayev A.S., Markusheva T.V., Kramerov D.A., Ryskov A.P., Skryabin K.G., Bayev A.A., Georgiev G.P. (1982) Ubiquitous transposon-like repeats B1 and B2 of the mouse genome: B2 sequencing. *Nucl. Acids Res.* **10**, 7461–7475.
 44. Deininger P.L., Jolly D.J., Rubin C.M., Friedmann T., Schmid C.W. (1981) Base sequence studies of 300 nucleotide renatured repeated human DNA clones. *J. Mol. Biol.* **151**, 17–33.
 45. Singer M.F. (1982) SINES and LINES: highly repeated short and long interspersed sequences in mammalian genomes. *Cell.* **28**, 433–434.
 46. Mouse Genome Sequencing Consortium. (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature.* **420**, 520–562.
 47. International Human Genome Sequencing Consortium. (2001) Initial sequencing and analysis of the human genome. *Nature.* **409**, 860–921.
 48. Vassetzky N.S., Kramerov D.A. (2013) SINEBase: a database and tool for SINE analysis. *Nucl. Acids Res.* **41**, D83–D89.
 49. Kajikawa M., Okada N. (2002) LINES mobilize SINES in the eel through a shared 3' sequence. *Cell.* **111**, 433–444.
 50. Dewannieux M., Esnault C., Heidmann T. (2003) LINE-mediated retrotransposition of marked Alu sequences. *Nat. Genet.* **35**, 41–48.
 51. Daniels G.R., Deininger P.L. (1985) Repeat sequence families derived from mammalian tRNA genes. *Nature.* **317**, 819–822.
 52. Ullu E., Tschudi C. (1984) Alu sequences are processed 7SL RNA genes. *Nature.* **312**, 171–172.
 53. Gogolevsky K.P., Vassetzky N.S., Kramerov D.A. (2009) 5S rRNA-derived and tRNA-derived SINES in fruit bats. *Genomics.* **93**, 494–500.
 54. Kramerov D.A., Vassetzky N.S. (2011) SINES. *Wiley Interdisciplinary Rev. RNA.* **2**, 772–786.
 55. Dewannieux M., Heidmann T. (2005) L1-mediated retrotransposition of murine B1 and B2 SINES recapitulated in cultured cells. *J. Mol. Biol.* **349**, 241–247.
 56. Kramerov D.A., Lekakh I.V., Samarina O.P., Ryskov A.P. (1982) The sequences homologous to major interspersed repeats B1 and B2 of mouse genome are present in mRNA and small cytoplasmic poly(A) + RNA. *Nucl. Acids Res.* **10**, 7477–7491.
 57. Grigoryan M.S., Kramerov D.A., Tulchinsky E.M., Revasova E.S., Lukanidin E.M. (1985) Activation of putative transposition intermediate formation in tumor cells. *EMBO J.* **4**, 2209–2215.
 58. Kramerov D.A., Tillib S.V., Lekakh I.V., Ryskov A.P., Georgiev G.P. (1985) Biosynthesis and cytoplasmic distribution of small poly(A)-containing B2 RNA. *Biochim. Biophys. Acta.* **824**, 85–98.
 59. Kramerov D.A., Tillib S.V., Ryskov A.P., Georgiev G.P. (1985) Nucleotide sequence of small polyadenylated B2 RNA. *Nucl. Acids Res.* **13**, 6423–6437.
 60. Kramerov D.A., Tillib S.V., Shumyatsky G.P., Georgiev G.P. (1990) The most abundant nascent poly(A) + RNAs are transcribed by RNA polymerase III in murine tumor cells. *Nucl. Acids Res.* **18**, 4499–44506.
 61. Bladon T.S., Fregeau C.J., McBurney M.W. (1990) Synthesis and processing of small B2 transcripts in mouse embryonal carcinoma cells. *Mol. Cell. Biol.* **10**, 4058–4067.
 62. Serdobova I.M., Kramerov D.A. (1998) Short retrotransposons of the B2 superfamily: evolution and application for the study of rodent phylogeny. *J. Mol. Evol.* **46**, 202–214.
 63. Borodulina O.R., Kramerov D.A. (1999) Wide distribution of short interspersed elements among eukaryotic genomes. *FEBS Lett.* **457**, 409–413.
 64. Borodulina O.R., Kramerov D.A. (2001) Short interspersed elements (SINES) from insectivores. Two classes of mammalian SINES distinguished by A-rich tail structure. *Mamm. Genome.* **12**, 779–786.
 65. Borodulina O.R., Kramerov D.A. (2005) PCR-based approach to SINE isolation: simple and complex SINES. *Gene.* **349**, 197–205.
 66. Borodulina O.R., Kramerov D.A. (2008) Transcripts synthesized by RNA polymerase III can be polyadenylated in an AAUAAA-dependent manner. *RNA.* **14**, 1865–1873.
 67. Borodulina O.R., Golubchikova J.S., Ustyantsev I.G., Kramerov D.A. (2016) Polyadenylation of RNA transcribed from mammalian SINES by RNA polymerase III: Complex requirements for nucleotide sequences. *Biochim. Biophys. Acta.* **1859**, 355–365.
 68. Roy-Engel A.M., El-Sawy M., Farooq L., Odom G.L., Perepelitsa-Belancio V., Bruch H., Oyeniran O.O., Deininger P.L. (2005) Human retroelements may introduce intragenic polyadenylation signals. *Cytogenet. Genome Res.* **110**, 365–371.
 69. Kramerov D.A., Vassetzky N.S. (2011) Origin and evolution of SINES in eukaryotic genomes. *Heredity.* **107**, 487–495.
 70. van Hoof A., Lennertz P., Parker R. (2000) Yeast exosome mutants accumulate 3'-extended polyadenylated forms of U4 small nuclear RNA and small nucleolar RNAs. *Mol. Cell. Biol.* **20**, 441–452.
 71. Kadaba S., Krueger A., Trice T., Krecic A.M., Hinnebusch A.G., Anderson J. (2004) Nuclear surveillance and degradation of hypomodified initiator tRNA^{Met} in *S. cerevisiae*. *Genes Dev.* **18**, 1227–1240.
 72. Reinisch K.M., Wolin S.L. (2007) Emerging themes in non-coding RNA quality control. *Curr. Opin. Struct. Biol.* **17**, 209–214.

73. LaCava J., Houseley J., Saveanu C., Petfalski E., Thompson E., Jacquier A., Tollervey D. (2005) RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. *Cell*. **121**, 713–724.
74. Falk S., Weir J.R., Hentschel J., Reichelt P., Bonneau F., Conti E. (2014) The molecular architecture of the TRAMP complex reveals the organization and interplay of its two catalytic activities. *Mol. Cell*. **55**, 856–867.
75. Dez C., Houseley J., Tollervey D. (2006) Surveillance of nuclear-restricted pre-ribosomes within a subnuclear region of *Saccharomyces cerevisiae*. *EMBO J.* **25**, 1534–1546.
76. San Paolo S., Vanacova S., Schenk L., Scherrer T., Blank D., Keller W., Gerber A.P. (2009) Distinct roles of non-canonical poly(A) polymerases in RNA metabolism. *PLoS Genet.* **5**, e1000555.
77. Kong K.Y., Tang H.M., Pan K., Huang Z., Lee T.H., Hinnebusch A.G., Jin D.Y., Wong C.M. (2014) Cotranscriptional recruitment of yeast TRAMP complex to intronic sequences promotes optimal pre-mRNA splicing. *Nucl. Acids Res.* **42**, 643–660.
78. Jia H., Wang X., Liu F., Guenther U.P., Srinivasan S., Anderson J.T., Jankowsky E. 2011. The RNA helicase Mtr4p modulates polyadenylation in the TRAMP complex. *Cell*. **145**, 890–901.
79. Lubas M., Christensen M.S., Kristiansen M.S., Domanski M., Falkenby L.G., Lykke-Andersen S., Andersen J.S., Dziembowski A., Jensen T.H. (2011) Interaction profiling identifies the human nuclear exosome targeting complex. *Mol. Cell*. **43**, 624–637.
80. Lubas M., Andersen P.R., Schein A., Dziembowski A., Kudla G., Jensen T.H. (2015) The human nuclear exosome targeting complex is loaded onto newly synthesized RNA to direct early ribonucleolysis. *Cell Repts*. **10**, 178–192.
81. Boele J., Persson H., Shin J.W., Ishizu Y., Newie I.S., Sokilde R., Hawkins S.M., Coarfa C., Ikeda K., Takayama K., Horie-Inoue K., Ando Y., Burroughs A.M., Sasaki C., Suzuki C., Sakai M., Aoki S., Ogawa A., Hasegawa A., Lizio M., Kaida K., Teusink B., Carninci P., Suzuki H., Inoue S., Gunaratne P.H., Rovira C., Hayashizaki Y., de Hoon M.J. (2014) PAPD5-mediated 3' adenylation and subsequent degradation of miR-21 is disrupted in proliferative disease. *Proc. Natl. Acad. Sci. USA*. **111**, 11467–11472.
82. Mohanty B.K., Kushner S.R. (2011) Bacterial/archaeal/organellar polyadenylation. *Wiley Interdisciplinary Rev. RNA*. **2**, 256–276.

CANONICAL AND NON-CANONICAL POLYADENYLATION OF RNA

I. G. Ustyantsev, J. S. Golubchikova, O. R. Borodulina, D. A. Kramerov*

Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow 119991

*e-mail: kramerov@eimb.ru

Polyadenylation is non-template addition of adenosine nucleotides at 3' end of RNA that occurs following transcription. Polyadenylation results in the formation of poly(A) tail up to 250–300 nucleotides long. The review deals with the polyadenylation and consists of three parts. The first section considers the classical process of mRNA terminus formation by the cleavage of transcript synthesized by RNA polymerase II and the coupled poly(A) tail synthesis by a canonical polyadenylate polymerase. Nucleotide sequences which are necessary for mRNA cleavage and synthesis of poly(A) tail are considered in detail, in particular, a polyadenylation signal AAUAAA. Numerous proteins and their complexes involved in the cleavage and polyadenylation of mRNA are described. The significance of poly(A) tail for prolongation of lifetime of mRNAs and stimulation of their translation are discussed. The second section presents the data demonstrating the ability for the AAUAAA-dependent polyadenylation of RNAs transcribed by RNA polymerase III from certain SINEs (Short Interspersed Elements). Peculiar structural and functional features of RNA polymerase III make the process of polyadenylation of RNA synthesized by this enzyme unusual. The second part of the review describes in greater detail the history of recent developments in this study area, in particular, the discovery of AAUAAA-dependent polyadenylation of RNAs synthesized by RNA polymerase III that has not been previously discussed in other reviews. Data concerning the AAUAAA-independent polyadenylation catalyzed by non-canonical TRAMP poly(A)-polymerases (Trf4 and Trf5) are presented in the third section. These enzymes promote rapid degradation of RNAs by adding to them a short poly(A) tail. This mechanism allows for recognition, poly(A)-marking and elimination of incorrectly folded non-coding transcripts (e.g. ribosomal and transfer RNAs).

Keywords: canonical and non-canonical poly(A) polymerases, signals and elements of polyadenylation, protein factors of polyadenylation, RNA polymerases II and III, SINE