



УДК 577.214(.337+.622+32)+575.852('112+'113)

## МОЛЕКУЛЯРНАЯ ЭВОЛЮЦИЯ И СТРУКТУРА СУБЪЕДИНИЦ ЯДЕРНЫХ РНК-ПОЛИМЕРАЗ ЭУКАРИОТ В СВЕТЕ ЭКЗОН-ИНТРОННОЙ ОРГАНИЗАЦИИ ИХ ГЕНОВ

© 1999 г. Г. В. Шпаковский<sup>#</sup>, Е. Н. Лебеденко

Институт биоорганической химии им. М.М. Шемякина и Ю.А. Овчинникова РАН,  
117871, Москва, ГСП-7, ул. Миклухо-Маклая, 16/10

Поступила в редакцию 30.03.99 г. Принята к печати 05.07. 99 г.

Анализ собственных (для *Schizosaccharomyces pombe*) и литературных (для *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, *Arabidopsis thaliana*, *Homo sapiens* и ряда других Eucarya) данных по экзон-инtronной организации генов, кодирующих субъединицы ядерных РНК-полимераз, показал, что интроны в ортологичных генах из разных организмов расположены не случайным образом: их позиции, спроецированные на карты сравнения аминокислотных последовательностей субъединиц-ортологов, зачастую совпадают у эволюционно очень далеких видов. При этом позиции инtronов, как правило, соответствуют краям зон консервативности или местам возможных изгибов полипептидной цепи, ограничивая элементы вторичной структуры (субъединица Rpb8 с экспериментально установленной пространственной структурой) или структурно-функциональные модули (субъединицы Rpb10 и Rpc10). Это согласуется как с идеей о древнем происхождении инtronов, так и с представлением об эволюции древних белковых последовательностей путем сборки их генов из короткихprotoэкзонов, отобранных Природой еще во времена мира РНК. Сравнительный анализ первичных структур субъединиц РНК-полимераз эукариот позволил обнаружить сигнал ядерной локализации в субъединице Rpb10, а также найти предположительные гомологии субъединицы Rpc10 у некоторых архей.

**Ключевые слова:** организация генов субъединиц ядерных РНК-полимераз эукариот; эволюция экзон-инtronной структуры; модули вторичной структуры; экзонная теория генов.

Установление полной первичной структуры геномов целого ряда организмов (в том числе, двух эукариотических: дрожжей *Saccharomyces cerevisiae* [1, 2] и нематоды *Caenorhabditis elegans* [3]) привело к возникновению геномики как науки о строении, функционировании и молекулярной эволюции геномов и подняло на качественно новый уровень рассмотрение многих фундаментальных общебиологических проблем.

Одна из таких проблем – вопрос о происхождении, функциональной роли и эволюции инtronов, а также о соотношении экзон-инtronной организации генов и пространственной структуры белков, которые они кодируют. Обсуждение этой проблемы началось сразу после открытия в 1977 г. экзон-инtronной структуры генов [4–7]. Почти одновременно возникли две противоположные точки зрения. Одна из них утверждает, что гены исходных организмов были непрерывными и, подобно генам современных бактерий, инtronов не содержали (см. [8–10]). Согласно этой точке зрения, интраны явились относительно недавним приобретением эукариотических генов и возни-

кали в основном в результате процесса, сходного с транспозицией как мы ее знаем сегодня. Противоположная точка зрения [11–13] предполагает, что интраны имеют очень древнее происхождение и появились одновременно с возникновением генов из относительно коротких фрагментов первоначальных наследственных молекул. Впоследствии у некоторых эволюционных ветвей организмов интраны были полностью утрачены. Наиболее полно эта точка зрения была развита Уолтером Гилбертом в рамках Экзонной Теории Генов (ЭТГ) [14, 15]. По этой теории первоначальные гены (ur-genes) прогенота (англ. progenote, первичный организм) возникали из коротких последовательностей, кодирующих полипептиды длиной 15–20 а. о. Эти отобранные эволюцией последовательности явились первичными экзонами, дальнейшее комбинирование и эволюция продуктов которых привели к нынешнему разнообразию белковых структур. Некодирующие концевые последовательности первичных экзонов при их рекомбинации приводили к образованию прерывающихся последовательностей – инtronов, которые можно рассматривать как остатки “эволюционного клея”. В отличие от прокариот, освободившихся от этих “ненужных” остатков, у большинства

<sup>#</sup> Автор для переписки (тел.: (095) 330-65-83; e-mail: gvs@ibch.siobc.ras.ru; факс: (095) 335-71-03).

эукариот интроны претерпевали существенную эволюцию, превращаясь зачастую в гигантские последовательности, выполняющие, возможно, важную функцию в геноме (см., например, [16]).

Говоря об ЭТГ, нельзя не отметить важную роль, которую сыграла в ее разработке и развитии концепция модульной организации пространственной структуры белков, предложенная Митико Го [17]. Согласно этой концепции, разработанной на модели гемоглобина, третичная структура белка может быть разбита на компактные полипептидные модули диаметром до 28 Å, которые, как правило, соответствуют отдельным экзонам. Эта гипотеза позволила исследовательнице правильно предсказать в семействе глобиновых генов позицию нового интрана, который был впоследствии обнаружен в генах леггемоглобинов растений [18, 19]. Следуя этой же идеи, У. Гилберт с соавторами правильно предсказали новый инtron в гене триозофосфатизомеразы [20, 21].

Быстрое накопление данных по первичной структуре генов, число которых возрастает примерно в 10 раз каждые 5 лет, позволило применить методы математической статистики для анализа распределения инtronов в генах и их положения в кодоне (фазы). Так, в группе У. Гилберта была проведена статистическая обработка 44 древних (т. е. имеющих ортологов во всех трех надцарствах – Urchkingdoms – Bacteria, Archaea и Eucarya) белков, являющихся, как правило, ферментами универсальных метаболических путей [22–25]. По-видимому, именно на таких, “дом поддерживавших” (англ. house-keeping) генах только и возможно рассматривать вопросы, связанные с происхождением и ранней эволюцией структуры генов.

Несомненно к таким генам можно отнести и гены компонентов аппаратов транскрипции и трансляции, в частности гены субъединиц ядерных РНК-полимераз I–III, которые до сих пор из рассмотрения выпадали. Частично это было связано с отсутствием структурной информации: гены ядерных РНК-полимераз наиболее хорошо охарактеризованного объекта, *S. cerevisiae*, как правило, инtronов не содержат, а для целого ряда других эукариотических организмов имелась только частичная информация. Так, в случае двудольного растения *Arabidopsis thaliana* и человека (*Homo sapiens*) были идентифицированы многие компоненты РНК-полимераз, но лишь на уровне мРНК [26–28]. Только в последние 2–3 года стала появляться и геномная информация. В декабре 1998 г. было завершено секвенирование генома нематоды *C. elegans* [3], но следует иметь в виду, что экзон-инtronная структура многих генов в этом случае является предположительной, так как предсказана с помощью компьютера.

Более полная информация имеется на сегодняшний день о генах ядерных РНК-полиме-

раз I–III *Schizosaccharomyces pombe*. Недавно усилиями лаборатории Акиры Ишихамы (Национальный институт генетики, Япония) и нашей группы завершена структурная характеристика всех 12 субъединиц РНК-полимеразы II и их генов [26, 29–38]; кроме того, клонированы и секвенированы кДНК и гены 9 из 10 незаменимых субъединиц РНК-полимеразы I *Sz. pombe* (8 из них – в нашей группе) [33, 39–42]. Делящиеся дрожжи *Sz. pombe*, рано разошедшаяся с *S. cerevisiae* ветвь одноклеточных аскомицетов [43, 44], представляют несомненно удобный объект для анализа эволюции экзон-инtronной организации генов, поскольку интроны широко представлены в их геноме [45, 46].

Таким образом, для ядерных РНК-полимераз, особенно РНК-полимеразы II, на сегодняшний день есть данные о структуре генов для всех трех основных царств эукариот: грибов (*S. cerevisiae*, *Sz. pombe*), растений (*A. thaliana*) и животных (*C. elegans*, *Drosophila melanogaster*, *H. sapiens*). Настала пора систематизации и детального анализа этих структур.

В ядерных РНК-полимеразах I–III можно выделить три типа субъединиц. Во-первых, это две самые большие субъединицы каждого из ферментов, структурно [47–49] и функционально [50–52] гомологичные субъединицам  $\beta$  и  $\beta'$  РНК-полимеразы *E. coli* и представленные во всех трех основных эволюционных доменах (Bacteria, Archaea и Eucarya) [53]. Вторую группу составляют субъединицы Rpb3 и Rpb11 РНК-полимеразы II, а также гомологичные им по структуре субъединицы Rpc40 и Rpc19 из РНК-полимераз I и III, тоже восходящие к бактериальному ферменту (субъединица  $\alpha$ ), однако имеющие с ним лишь ограниченное сходство в рамках так называемого  $\alpha$ -мотива, впервые обнаруженного Мартиндейлом [54]. Функциональная роль гетеродимеров Rpc40–Rpc19 и Rpb3–Rpb11, по-видимому, сходна с ролью гомодимера  $\alpha_2$  в бактериальном ферменте [55–59]. В третью группу входят все остальные, так называемые малые субъединицы, к которым относятся пять субъединиц (Rpb5, Rpb6, Rpb8, Rpb10 и Rpc10), общих для всех трех ферментных комплексов [60–62], и ряд субъединиц, специфичных для каждой из РНК-полимераз [53, 63–67]. В отличие от белков двух первых типов субъединицы третьей группы не имеют гомологов в РНК-полимеразах эубактерий, но для большинства из них есть гомологи в составе РНК-полимераз архей [68–71], причем структурная консервативность, по крайней мере для субъединицы Rpb10, дополняется и функциональным родством [72, 73].

Представленность субъединиц первого и (с определенной оговоркой) второго типов во всех трех надцарствах живых существ указывает на их более древнее происхождение по сравнению с

## Интроны в генах, кодирующих субъединицы ядерной ДНК-зависимой РНК-полимеразы II\*

Ген	Число инtronов							Число позиций инtronов		Комплементация гена-гомолога, в <i>S. cerevisiae</i> ***		Субъединица
	<i>Sz. pom.</i>	<i>S. cer.</i>	<i>A. tha.</i>	<i>C. ele.</i>	<i>D. mel.</i>	<i>H. sap.</i>	другие	всего	с совпадениями**	<i>Sz. pom.</i>	<i>H. sap.</i>	
<i>rpb1</i> <sup>+</sup>	6	0	11	11	3	28	28 <i>M. mus.</i>	44	11	⊖	⊖	Rpb1
<i>rpb2</i> <sup>+</sup>	1	0	24	9	2	28	1 <i>E. oct.</i>	35	2 (♣♣)	⊖	⊖	Rpb2
<i>rpb3</i> <sup>+</sup>	2	0	2	5	0	8	—	11	5 (♥♦♣♣♥)	⊕ <i>ts</i>	⊖	Rpb3
<i>rpb4</i> <sup>+</sup>	3	0	?	2	—	3	—	4	3 (♦♣♥)	⊕	⊕	Rpb4
<i>rpb5</i> <sup>+</sup>	2	0	3	2	0	6	—	11	2 (♣♥)	⊕	⊖	Rpb5
<i>rpb6</i> <sup>+</sup>	1	1	5	0	?	4	—	7	2 (♥♥)	⊕	⊕	Rpb6
<i>rpb7</i> <sup>+</sup>	1	0	5	0	—	7	—	7	5 (♥♣♣♣♣)	⊕	⊕ <i>ts</i>	Rpb7
<i>rpb8</i> <sup>+</sup>	2	0	3	2	—	3	3 <i>P. min.</i>	7	3 (♦♣♥)	⊖	⊕ <i>ts</i>	Rpb8
<i>rpb9</i> <sup>+</sup>	4	0	3	?	0	5	—	8	3 (♥♣♥)	⊕	⊕	Rpb9
<i>rpb11</i> <sup>+</sup>	2	0	4	2	—	3	—	5	3 (♥♥♣)	⊕ <i>ts</i>	⊖	Rpb11
<i>rpb10</i> <sup>+</sup>	2	0	3	?	—	1	—	3	3 (♣♥♦)	⊕	⊕	Rpb10
<i>rpc10</i> <sup>+</sup>	1	0	2	2	2	2	2 <i>U. fab.</i>	2	2 (♦♦)	⊕	⊕	Rpc10

\* В таблице и в рисунках использованы следующие сокращения названий организмов: *A. tha.* – *Arabidopsis thaliana*, *Ar. ful.* – *Archaeoglobus fulgidus*, *C. ele.* – *Caenorhabditis elegans*, *D. mel.* – *Drosophila melanogaster*, *E. oct.* – *Euplotes octocarinatus*, *H. sap.* – *Homosapiens*, *M. mus.* – *Mus musculus*, *P. min.* – *Pedinomonas minutissima*, *Pyr. hor.* – *Pyrococcus horikoshii*, *S. cer.* – *Saccharomyces cerevisiae*, *Sz. pom.* – *Schizosaccharomyces pombe*, *U. fab.* – *Uromyces fabae*.

\*\* Обозначения позиций инtronов как на рис. 1–5.

\*\*\* Плюс в кружочке означает способность соответствующего гена заменять функцию нативного гена-ортолога в клетках *S. cerevisiae*, минус – отсутствие такой способности. Знаками *ts* отмечены случаи, когда в результате межвидовой комплементации образуются термо чувствительные штаммы дрожжей (*ts*-мутанты).

субъединицами третьего типа. Это подтверждает и высокая эволюционная консервативность катализитического центра РНК-полимераз бактерий, архей и эукариот, расположенного на двух самых больших субъединицах каждого из ферментов [50–52, 74, 75]. Вместе с тем, учитывая очень давнее происхождение архей, к древним белкам несомненно можно отнести и субъединицы третьего типа. Все малые субъединицы представляют собой уникальные (то есть практически не имеющие гомологичных доменов в белках других семейств) полипептиды, структурно очень консервативные в эволюции. Опыты по межвидовой комплементации отдельных субъединиц РНК-полимераз I–III в клетках *S. cerevisiae* убедительно доказывают их функциональное родство [26, 27, 29, 32, 65, 72, 73, 76–78].

В данном обзоре мы рассмотрим экзон-инtronную структуру генов, кодирующих субъединицы РНК-полимеразы II, тем более, что пять из них входят в состав также РНК-полимераз I и III. В таблице суммированы данные о числе инtronов в этих генах у разных эукариотических организмов. Сравнение аминокислотных последовательностей всех 12 субъединиц РНК-полимеразы II разных эукариот и сопоставление полученных данных с организацией соответствую-

щих генов позволило выявить в каждом конкретном случае общее число позиций инtronов, известное на сегодняшний день (см. таблицу, а также рис. 1–5).

Оказалось, что интроны в ортологичных генах из разных организмов расположены не случайным образом: их позиции, спроектированные на карты сравнения аминокислотных последовательностей субъединиц-ортологов, зачастую совпадают у эволюционно очень далеких видов, как проиллюстрировано на примере общих субъединиц РНК-полимераз I–III *Rpc10* (рис. 1a), *Rpb6* (рис. 2), *Rpb8* (рис. 4) и *Rpb10* (рис. 5), а также субъединицы *Rpb7*, специфичной для РНК-полимеразы II (рис. 3). Позиции инtronов, как правило, соответствуют краевым точкам зон консервативности или местам возможных изгибов полипептидной цепи (см. рис. 1–5). В ряде случаев позиции инtronов ограничивают реально существующие структурно-функциональные модули, как например кислый (обогащенный остатками аспарагиновой и глутаминовой кислот) *N*-концевой домен субъединицы *Rpb6* [29, 87] (рис. 2), сигнал ядерной локализации белка в случае субъединицы *Rpc10* [32, 88] (рис. 1a) или часть (первую половину) атипичного цинксвязывающего домена RCX<sub>2</sub>CGK♣X<sub>n</sub>RYCCRRM субъединицы *Rpb10*

## Rpc10

(a)

<i>Sz. pom.</i>	1 MNHPTSTGGTAFNPPRATM <b>IYLCADC</b> GARNTIQA <b>K</b> EVIRCRECGHRVMY <b>KMR</b> TKRMVQFEAR	63 (1)
<i>S. cer.</i>	MSREGFQIPTNLDAAAAGTSQARTATLKY <b>I</b> CAECSSKLSLRTDAVRCKDCGHRIL <b>I</b> KARTKRLVQFEAR	70 (0)
<i>U. fab.</i>	1 MSQSYPNQSVQPNFGRPGQGVPI <b>TLCADC</b> GKDNC <b>I</b> KPREPVRCQEGSRVMYKKR <b>I</b> KRMVQFEAR	67 (2)
<i>A. tha.</i>	1 MDPAPEPVTV <b>C</b> GDCGQ <b>E</b> NTLKSGD <b>V</b> QCRCGYRILYKKR <b>R</b> RVQYEAR	57 (2)
<i>C. ele.</i>	1 MDGSATPGPGQAHLSNSMIY <b>I</b> CGECHAENE <b>I</b> KPKDAIRCRECGYRILYKKR <b>R</b> LMVYDAR	62 (2)
<i>D. mel.</i>	1 MSETSSKDNVKTAMTY <b>I</b> CGECHHENEMRPRDPIRCRECGYRIMYKKR <b>T</b> KRLVVF <b>D</b> AR	57 (2)
<i>H. sar.</i>	1 MDTQKDVQPPPQ <b>M</b> IY <b>I</b> CGE <b>T</b> ENE <b>I</b> KSRDPIRCRECGYRIMYKKR <b>T</b> KRLVVF <b>D</b> AR	58 (2)

Y C C                    C C G R            K R R            AR

(б)

<i>Sz. pom.</i>	1 MNHPTSTGGTAFNPPRATM <b>IYLCADC</b> GARNTIQA--KEVIRCRECGHRVMY <b>KMR</b> TKRMVQFEAR	63
	M <b>Y+C</b> CGA I K + + + C CG R++ <b>K R</b> +	
<i>Ar. ful.</i>	1 MSYVCLICGAEV <b>DID</b> -TEKSLVQ <b>C</b> TNC <b>G</b> GRIL <b>I</b> KPRPLAKKKRVKAI	46
	+ <b>Y C</b> CG EV + D + + + C CG + IL <b>K P R P</b> + RVKAI	
<i>Pyr. hor.</i>	1 MGQYLRVVLISQSLRPVRLLEELWSDVEMPEAV <b>Y</b> RCA <b>K</b> CGREV <b>L</b> LDLSTTRDLRC <b>P</b> Y <b>C</b> GSKILY <b>K</b> P <b>R</b> P <b>K</b> I-PRRVKAI	77
	+ + + <b>Y C</b> + CG + C CG K L + P+ + PR	
<i>Pyr. hor.</i> (L40)	1 MARFPEEARIF <b>K</b> K <b>I</b> LCRG <b>A</b> TNPWG----AE <b>K</b> CR <b>K</b> CG <b>Y</b> KRL-RPKARE-PRGGGR	51

Y C CG                    RC CG R LYK R            R

**Рис. 1.** Сравнение первичных структур субъединицы Rpc10 ядерных РНК-полимераз I-III *Sz. pombe* и ее гомологов из других эукариот (а) и архей (б). Знаки ♦ над аминокислотными последовательностями соответствуют позиции инtronов, расположенных после первого звена кодона данной аминокислоты (фаза 1). В рамку заключен домен, отвечающий за ядерную локализацию белка. В нижних строках приведены инвариантные (а) или консервативные (инвариантные – подчеркнуты) в белках как эукариот, так и архей (б) аминокислотные остатки. Цифры указывают нумерацию аминокислот, цифры в скобках – число инtronов в соответствующем гене (см. [26, 32, 36] и GenBank U81794, AB010072, Z82266, AC004766). Инвариантные аминокислотные остатки выделены темно-серым фоном (а) или соответствующей буквой между последовательностями (б); консервативные – светло-серым фоном (а) или знаком “+” (б). Первичные структуры белков архей взяты из следующих номеров депонирования в GenBank: AE001102 (ген AF0056), AP000007 (ген PHS056) и AP000006 (ген PHS037).

## RpB6

<i>Sz. pom.</i>	1 MSDYEED <b>E</b> DEAF <b>G</b> MDGAV <b>M</b> EEEV <b>D</b> ELEM <b>I</b> DENG <b>Q</b> S <b>Q</b> -----GVSHP <b>G</b> EP <b>S</b> TTV <b>I</b> TED <b>V</b> ASS <b>K</b> T <b>A</b> Q <b>S</b> <b>G</b> KAVAKEDRT <b>TT</b>	72
<i>S. cer.</i>	1 MSDYE <b>E</b> AFND <b>G</b> NEN <b>F</b> ED <b>F</b> DVE <b>H</b> FS <b>D</b> E <b>E</b> TY <b>E</b> KPQFKD <b>G</b> ETTDANG <b>K</b> TIV <b>G</b> GN <b>G</b> P <b>E</b> DF <b>Q</b> Q <b>H</b> E <b>Q</b> IRR <b>K</b> TL <b>K</b> E <b>K</b> A <b>I</b> P <b>K</b> D <b>Q</b> R <b>A</b> T	82
<i>C. ele.</i>	1 MADED <b>D</b> D <b>Y</b> Q <b>D</b> M <b>D</b> ND <b>D</b> F <b>D</b> V <b>D</b> N <b>E</b> M <b>D</b> V <b>I</b> E <b>E</b> ED <b>P</b> Q <b>R</b> P <b>D</b> N <b>E</b> DD <b>D</b> DD <b>N</b> V <b>D</b> N <b>E</b> -----LFD <b>Q</b> CK <b>A</b> V <b>E</b> T <b>S</b> H <b>V</b> T	63
<i>D. mel.</i>	1 MDDAD <b>D</b> Y <b>D</b> ND <b>D</b> D <b>D</b> V <b>G</b> DD <b>D</b> DD <b>V</b> DE <b>D</b> DD <b>E</b> DI <b>Q</b> EE <b>E</b> AD <b>N</b> I <b>E</b> -----I <b>A</b> P <b>G</b> G <b>A</b> GG <b>GG</b> V <b>P</b> K <b>S</b> K <b>R</b> TT	59
<i>H. sap.</i>	1 MSD <b>N</b> E <b>D</b> NF <b>D</b> GG <b>D</b> FF <b>D</b> VE <b>E</b> DE <b>G</b> LL <b>D</b> LEN <b>A</b> <b>E</b> EG <b>Q</b> EN <b>V</b> E-----I <b>L</b> PS <b>G</b> ER <b>P</b> Q <b>A</b> N <b>Q</b> K <b>R</b> TT	54

D D EDD D DDD EEEE DEDDE EEE  
|————— кислый N-концевой домен —————| TT

<i>Sz. pom.</i>	73 PYMT <b>K</b> YERAR <b>I</b> LG <b>T</b> RA <b>L</b> Q <b>I</b> SM <b>N</b> AP <b>V</b> L <b>D</b> LE <b>G</b> ET <b>D</b> P <b>L</b> Q <b>I</b> AM <b>K</b> EL <b>A</b> Q <b>K</b> Q <b>I</b> PL <b>L</b> V <b>R</b> RY <b>L</b> PD <b>G</b> SY-EDWSVAEL <b>I</b>	142 (1)
<i>S. cer.</i>	83 PYMT <b>K</b> YERAR <b>I</b> LG <b>T</b> RA <b>L</b> Q <b>I</b> SM <b>N</b> AP <b>V</b> F <b>D</b> LE <b>G</b> ET <b>D</b> P <b>L</b> RIAM <b>K</b> EL <b>A</b> KK <b>I</b> PL <b>L</b> V <b>R</b> RY <b>L</b> PD <b>G</b> SY-EDWSVEEL <b>I</b> VD <b>L</b>	155 (1)
<i>C. ele.</i>	64 PFMT <b>K</b> YERAR <b>V</b> LG <b>T</b> RA <b>L</b> Q <b>I</b> TAM <b>G</b> P <b>V</b> VE <b>E</b> LE <b>G</b> ET <b>D</b> P <b>L</b> E <b>T</b> AR <b>K</b> EL <b>K</b> Q <b>R</b> R <b>I</b> PI <b>I</b> IV <b>R</b> RY <b>L</b> PD <b>G</b> SY-EDWP <b>T</b> D <b>Q</b> 1 <b>Q</b> LA <b>D</b> W	138 (0)
<i>D. mel.</i>	58 KYMT <b>K</b> YERAR <b>V</b> LG <b>T</b> RA <b>L</b> Q <b>I</b> TAM <b>C</b> P <b>V</b> ME <b>E</b> LE <b>G</b> ET <b>D</b> P <b>L</b> Q <b>I</b> AM <b>K</b> EL <b>K</b> Q <b>K</b> Q <b>I</b> PI <b>I</b> IR <b>R</b> Y <b>L</b> PD <b>H</b> SY-EDWS <b>I</b> DE <b>L</b> IM <b>V</b> D <b>N</b>	131 (?)
<i>H. sap.</i>	55 PYMT <b>K</b> YERAR <b>V</b> LG <b>T</b> RA <b>L</b> Q <b>I</b> TAM <b>C</b> P <b>V</b> ME <b>E</b> LE <b>G</b> ET <b>D</b> P <b>L</b> Q <b>I</b> AM <b>K</b> EL <b>K</b> Q <b>K</b> Q <b>I</b> PI <b>I</b> IR <b>R</b> Y <b>L</b> PD <b>G</b> SY-EDW <b>G</b> VD <b>E</b> LI <b>I</b> TD	127 (4)

MT**K**YERAR LG**T**RA**L**Q**I** M AP V L GETDPL IA KEL IP RRYLPD S EDW L

**Рис. 2.** Сравнение первичных структур ортологов субъединицы RpB6 ядерных РНК-полимераз I-III из различных эукариотических организмов. Знаки ♥ и ♦ над аминокислотными последовательностями отвечают позициям инtronов, расположенных соответственно после второго (фаза 2) или третьего (фаза 0) звена кодона данной аминокислоты. Выделение аминокислотных остатков – как на рис. 1; отмечен кислый N-концевой домен. В скобках указано число инtronов в соответствующих генах (см. [26, 29, 61, 79]).

## Rpb7

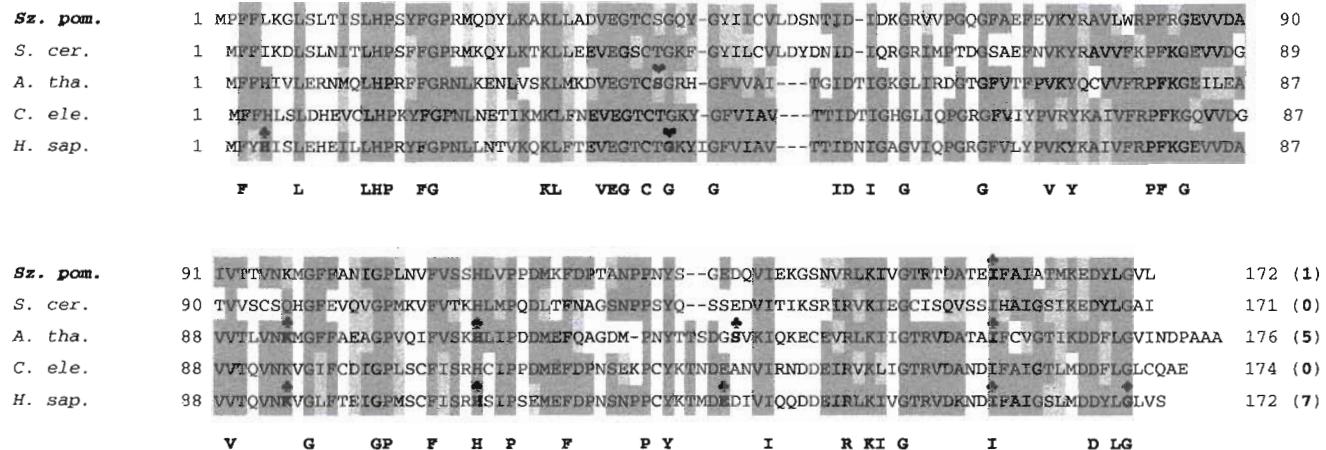


Рис. 3. Сравнение первичных структур ортологов субъединицы Rpb7 ядерной РНК-полимеразы II из различных эукариотических организмов. Все обозначения и выделение аминокислотных остатков как на рис. 1 и 2. Данные по числу и расположению инtronов взяты из публикаций [33, 34, 80, 81] и GenBank AF055916.

## Rpb8

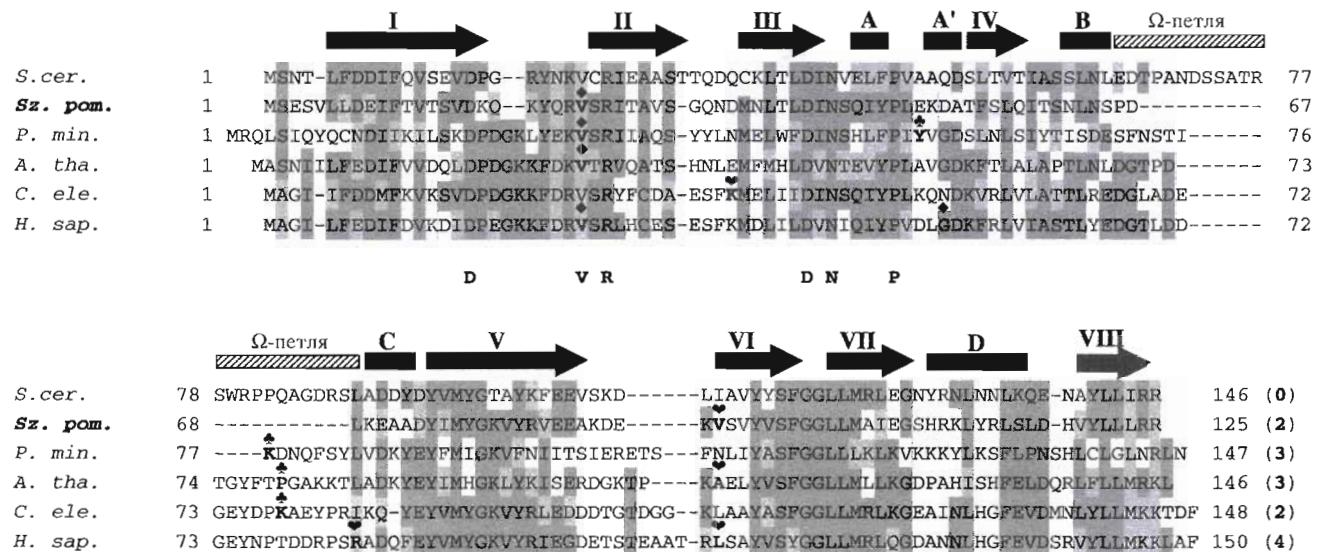
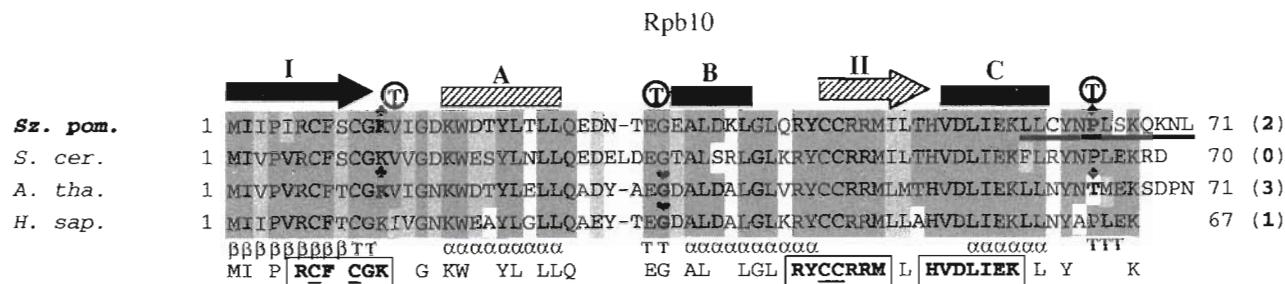


Рис. 4. Сравнение первичных структур ортологов субъединицы Rpb8 ядерных РНК-полимераз I, II и III из различных эукариотических организмов. Выделение аминокислотных остатков как на рис. 1 и 2. Позиции инtronов указаны знаками ♦, ♥ и ♣ (соответственно фаза 1, 2 и 0). Над аминокислотными последовательностями показаны экспериментально установленные элементы вторичной структуры RPB8 *S. cerevisiae* [82]: β-тяжки (стрелки) и α-спирали (прямоугольники), а также неструктурированная Ω-петля. В скобках указано число инtronов в соответствующих генах (см. [26, 35, 83]).

[26, 30] (рис. 5). Интересно отметить, что для двух самых маленьких субъединиц все известные на сегодняшний день позиции инtronов относятся к категории совпадающих, то есть встречаются хотя бы у двух разных организмов (см. таблицу и рис. 1а и 5).

Анализ мини-субъединиц Rpb10 и Rpc10 привел также к двум находкам. В геноме некоторых архей, *Archaeoglobus fulgidus* [89] и *Pyrococcus horikoshii* [90], мы обнаружили отдаленные гомо-

логи субъединицы Rpb10 (рис. 1б), которая до сих пор считалась чисто эукариотическим приобретением. Кроме того оказалось, что некоторым сходством с Rpb10 обладает еще один белок из *Pyr. horikoshii*, предположительно являющийся рибосомным белком L40. Таким образом, не исключено, что в случае Rpb10 основной структурный модуль был заимствован из более древнего аппарата трансляции. В случае субъединицы Rpb10 (рис. 5) нам удалось идентифицировать возмож-



**Рис. 5.** Сравнение первичных структур ортологов субъединицы Rpb10 ядерных РНК-полимераз I, II и III из различных эукариотических организмов. Все обозначения и выделение аминокислотных остатков как на рис. 1 и 2. Обозначения элементов вторичной структуры как на рис. 4 (элементы, которые не предсказываются однозначно у всех ортологов, даны штриховкой), ① – точка поворота полипептидной цепи. Под аминокислотной последовательностью Rpb10 *H. sapiens* приведена ее вторичная структура ( $\alpha$ - $\alpha$ -спираль,  $\beta$ - $\beta$ -тяж., Т-поворот), предсказанная с помощью программы ALB [84, 85]. В рамки заключены три мотива субъединицы Rpb10, строго консервативные у всех эукариот (см. [30]); подчеркнуты остатки цистеина, формирующие атипичный Zn-связывающий домен [26, 30]. Жирной полосой в первичной структуре белка *Sz. pombe* подчеркнута обсуждаемая в тексте последовательность, соответствующая сигналу ядерной локализации. Данные по числу и расположению инtronов взяты из публикаций [30, 36, 86] и GenBank AC005882.

ный сигнал ядерной локализации этого белка. Во-первых, замечено структурное сходство специфичного только для эукариотических гомологов Rpb10 C-концевого домена LLXYNPLEK (отсутствует у родственных субъединиц РНК-полимераз архей и вирусов животных [26]) с лейцинбогатыми последовательностями сигнала ядерного транспорта [91]. Кроме того, этот же участок (NPLSKQK у *Sz. pombe*) имеет структурное сходство с фрагментом рибосомного белка L36 (NPRHKQK), а, как известно, сборка рибосом происходит в ядрышке после первоначального импорта рибосомных белков из цитоплазмы. Положение одного из совпадающих инtronов (◆) в этом же участке генов Rpb10 у *Sz. pombe* и *A. thaliana* может говорить о том, что фрагмент гена, отвечающий C-концевым аминокислотным остаткам, был, по-видимому, присоединен именно для окончательного формирования этого домена ядерной локализации (рис. 5).

Ярким примером того, что позиции инtronов в генах субъединиц ядерных РНК-полимераз часто совпадают с границами элементов сворачивания (фолдинга) этих полипептидов, является субъединица Rpb8 – на сегодняшний день единственная субъединица ядерных РНК-полимераз эукариот, для которой экспериментально установлена пространственная структура [82]. Действительно, 5 из 7 позиций инtronов, известных на сегодняшний день в генах белка Rpb8 из разных организмов, точно совпадают с началом элементов вторичной структуры ( $\beta$ -тяж. II, III и VI;  $\alpha$ -спирали A' и C) в субъединице RPB8 *S. cerevisiae*. Шестой инtron попадает в область, соответствующую в RPB8 *S. cerevisiae* протяженной неструктурированной  $\Omega$ -петле (рис. 4). Не исключено, что у организмов, имеющих этот инtron (*Pedinomonas minutissima* и *C. elegans*), элемент вторичной структуры (предположительно, более длинная  $\alpha$ -спираль C) начинается именно в этом месте.

Сходная картина наблюдается в случае субъединицы Rpb10. С помощью программы ALB, разработанной А.В. Финкельштейном [84, 85], был проанализирован целый ряд гомологов этой субъединицы, приведенных на рис. 5, а также родственные им субъединицы РНК-полимераз архея *Sulfolobus acidocaldarius* (субъединица N) [70] и вириуса осповакцины (субъединица Rpo7) [92]. Выяснилось, что все три позиции инtronов, которые к тому же относятся к категории совпадающих (см. таблицу), практически соответствуют предсказанным точкам поворота (рис. 5). Используя информацию о позициях инtronов, а также о консервативных структурных доменах [30, 72, 73], мы скорректировали вторичную структуру Rpb10, предсказанную программой ALB (рис. 5). В окончательном варианте  $\beta$ -тяж I и  $\alpha$ -спирали B и C, а также все отмеченные места поворотов, предсказываются с высокой вероятностью.

Рассмотренные выше закономерности экзон-инtronной структуры генов ядерных РНК-полимераз эукариот хорошо согласуются как с идеей о древнем происхождении инtronов, так и с представлением об эволюции древних белковых последовательностей путем сборки их генов из коротких протоэкзонов. Вероятно, первоначальные кодирующие блоки, длина которых не могла превышать 100 нт [93–95], были отобраны Природой еще во времена мира РНК [96–99]. Эти блоки были ограничены некодирующими последовательностями РНК, предшественниками протоинtronов.

Данное утверждение хорошо иллюстрирует пример субъединицы Rpb7 (рис. 3), C-концевая половина которой разбивается совпадающими (NB!) позициями инtronов именно на такие полипептидные модули (14–22 а.о.). Интересно, что все эти инtronы расположены между кодонами (фаза 0 – ♦). В недавней публикации У. Гилберта с соавторами [25] статистическими методами показано, что только нулевая фаза инtronов (♦) кор-

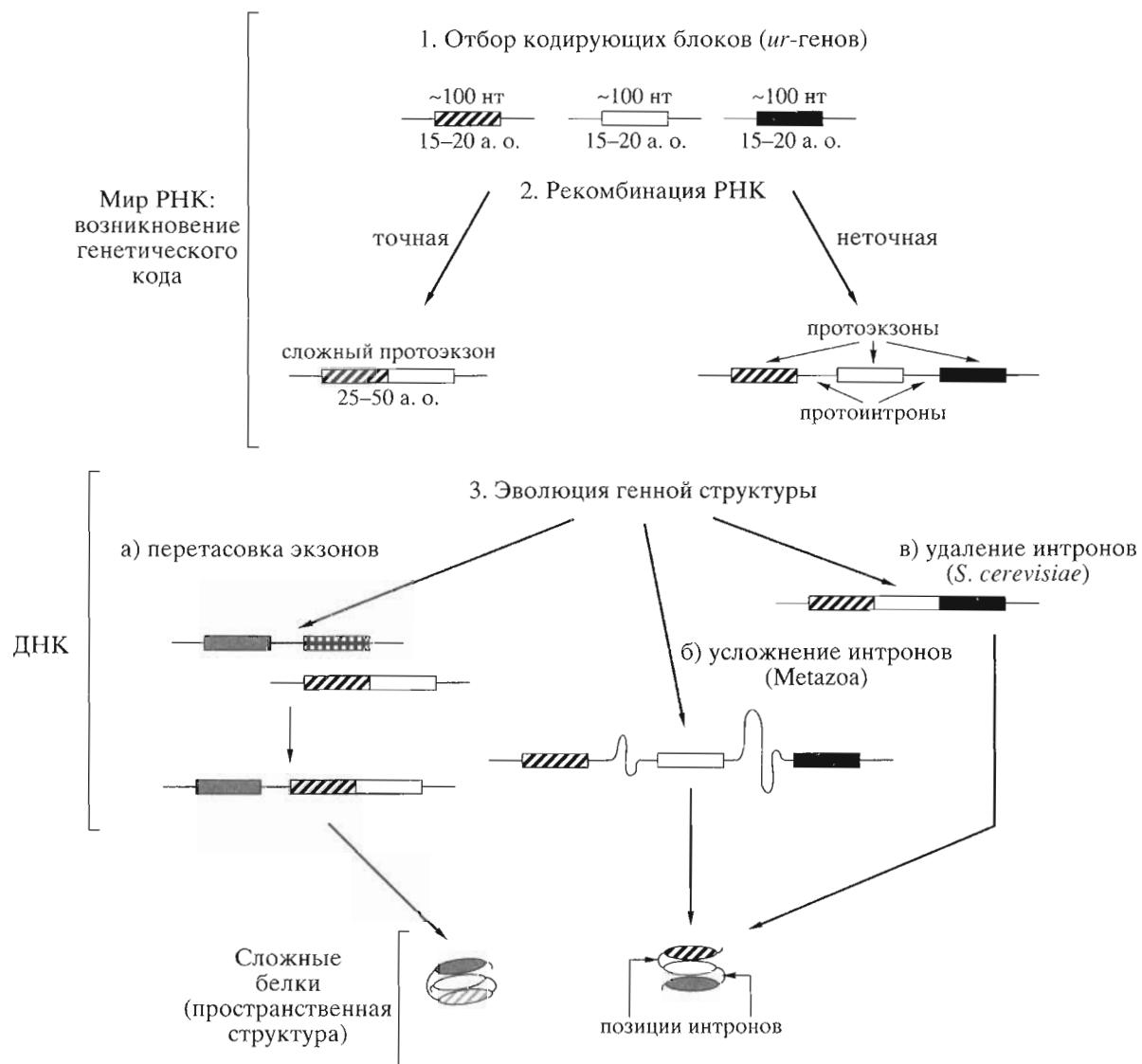


Рис. 6. Схема возникновения и эволюции экзон-инtronной структуры генов и формирования модульной структуры субъединиц РНК-полимераз у эукариот.

релирует со структурой древних белков. В этом смысле *C*-концевая половина Rpb7 представляет собой как бы законсервированную в эволюции структуру, своего рода "окаменелость". *N*-Конец этой же субъединицы демонстрирует последующие этапы эволюции организаций гена, включающие, быть может, рекомбинацию РНК по участкамprotoинtronов (рис. 6). Такая рекомбинация может быть точной и приводить к исчезновению соответствующих protoинtronов и образованию более длинных полипептидных модулей (25–50 а.о.), и/или неточной и, в этом случае, сохранять и видоизменять соответствующие protoинтраны, обеспечивая их эволюцию (включая смену фазы, небольшое смещение положения, изменение длины и даже возникновение новых функций). Возможность рекомбинаций такого типа была продемонстрирована в недавних работах А. Четверина с со-

трудниками по негомологичной рекомбинации РНК в бесклеточной системе (см. [100]).

Дальнейшая и наиболее существенная эволюция инtronов происходила уже на уровне ДНК-содержащих генов. При этом происходили такие процессы как перетасовка экзонов, удаление инtronов и/или их существенное усложнение. Перетасовка экзонов как один из основных путей возникновения новых белков и их эволюции детально обсуждалась в работах Гилберта и др. [14, 15, 22], и мы не будем здесь на этом останавливаться. Интересными примерами элиминирования инtronов являются дрожжи *S. cerevisiae* и нуклеоморф хлорархнотифтной водоросли *P. minutissima*. В случае *S. cerevisiae* почти все интраны в генах РНК-полимераз (за исключением совпадающего интрана в гене *RPB6* [61], см. рис. 2) были удалены благодаря чрезвычайно развитой у этого орга-

низма системе гомологичной рекомбинации на уровне ДНК. У *P. minutissima* параллельно с беспримерным редуцированием генома нуклеоморфа интроны укорачивались до самого минимума. В результате самые короткие из известных на сегодняшний день сплайсирующихся инtronов длиной 18–19 п. о. найдены в геноме этого нуклеоморфа, в частности в гене *RPB8* [83, 101] (см. рис. 4, а также [102]). Обсуждение путей усложнения инtronов (у высших организмов) с приобретением ими определенных функций выходит за рамки настоящего обзора. К эволюции инtronов на уровне ДНК, возможно, имеет отношение и тот факт, что интроны обильно представлены в геноме двудольного растения *A. thaliana* (см. таблицу), что наверняка связано с ролью полиплоидии в эволюции растений.

Модули, из которых построены субъединицы РНК-полимераз эукариот, практически не встречаются в других белках, что свидетельствует об уникальности и эволюционной консервативности базового аппарата транскрипции Eucarya. Косвенным подтверждением его древнего происхождения служит тот факт, что субъединица Rpb8, не имеющая явных гомологов по первичной структуре, оказалась сходной по пространственной структуре с некоторыми рибосомными белками бактерий (S1, S17), которые относятся к олигонуклеотид/олигосахаридсвязывающим белкам (OB-семейство) [82, 103, 104].

Таким образом, эволюция экзон-инtronной организации генов субъединиц ядерных РНК-полимераз эукариот и процесс формирования пространственной структуры этих субъединиц могут быть представлены в виде следующей схемы (рис. 6), восходящей во многом к концепции модульной организации пространственной структуры белков, предложенной М. Го [17] и Экзонной Теории Генов, развитой У. Гилбертом [14, 15]. В рамках этой схемы, руководствуясь сравнением экзон-инtronных структур всех известных генов-ортологов, в первичной структуре каждой из субъединиц можно выделить пептидные модули, ограниченные позициями инtronов. Такая разбивка субъединиц ядерных РНК-полимераз на небольшие модули повышает надежность предсказания вторичной структуры этих уникальных белков. Кроме того, пространственные структуры самых коротких, консервативных и важных в функциональном отношении (по данным мутагенеза) пептидных модулей могут быть рассчитаны методами конформационного анализа и проверены экспериментально физико-химическими методами с использованием соответствующих синтетических пептидов. Комбинированием предсказанных полипептидных модулей в синтезе белков *de novo*, исходя при этом из имеющихся в литературе данных о взаимодействующих доменах различных субъединиц, могут быть получены интересные белковые инструменты анализа

структур и функции базового аппарата транскрипции эукариот (например, блокаторы той или иной стадии сборки транскрипционного комплекса). Что касается самых маленьких общих субъединиц ядерных РНК-полимераз I–III (R<sub>p</sub>c10 и R<sub>p</sub>b10), то они являются удобными модельными объектами для создания алгоритма предсказания третичной структуры белков на основе вторичной структуры отмеченных в данной работе полипептидных модулей. Работа по созданию такого алгоритма может проводиться параллельно с экспериментальным анализом пространственной структуры этих белков (нативных или рекомбинантных) с помощью спектроскопии ЯМР.

Когда этот обзор уже был принят к печати, появилась статья (Darcy T.J., Hausner W., Awery D.E., Edwards A.M., Thomm M., Reeve J.N. “*Methanobacterium thermoautotrophicum* RNA Polymerase and Transcription *in vitro*.” J. of Bacteriology. 1999. V. 181. № 14. P. 4424–4429.), сообщающая об обнаружении белка, гомологичного субъединице R<sub>p</sub>c10, в очищенных препаратах РНК-полимеразы архея *Methanobacterium thermoautotrophicum* (субъединица P, ген *rpoP*=ORF MTH0680.5), так что присутствие гомолога эукариотической субъединицы R<sub>p</sub>c10 в РНК-полимеразе архей можно считать доказанным. Кроме того, недавно установленные первичные структуры генов *hRPB8* и *hRPB10α* (личное сообщение М. Виньерона, Страсбург, Франция) хорошо подтверждают основные идеи, предложенные в этом обзоре.

## СПИСОК ЛИТЕРАТУРЫ

1. Goffeau A., Barrell B.G., Bussey H., Davis R.W., Dujon B., Feldmann H., Galibert F., Hoheisel J.D., Jacq C., Johnston M. et al. // Science. 1996. V. 274. P. 546–567.
2. Goffeau A., Aert R., Agostini-Carbone M.L. et al. // Nature. 1997. V. 387. Suppl. To issue 6632.
3. Science. 1998. V. 282. P. 2012–2018 (<http://www.sanger.ac.uk/>; <http://genome.wustl.edu/gsc/gschmpg.html>).
4. Chow L.T., Gelinas R.E., Broker T.R., Roberts R.J. // Cell. 1977. V. 12. P. 1–8.
5. Berget S.M., Moore C., Sharp P.A. // Proc. Natl. Acad. Sci. USA. 1977. V. 74. P. 3171–3175.
6. Breathnach R., Mandel J.L., Chambon P. // Nature. 1977. V. 170. P. 314–319.
7. Brack C., Tonegawa S. // Proc. Natl. Acad. Sci. USA. 1977. V. 74. P. 5652–5656.
8. Rogers J. // Trends in Genet. 1989. V. 5. P. 213–216.
9. Cavalier-Smith T. // Trends in Genet. 1989. V. 7. P. 145–148.
10. Palmer J.D., Logsdon Jr. J.M. // Curr. Opin. Genet. Dev. 1991. V. 1. P. 470–477.
11. Gilbert W. // Nature. 1978. V. 271. P. 501.
12. Doolittle W.F. // Nature. 1978. V. 272. P. 581–582.
13. Blake C.C.F. // Nature. 1978. V. 273. P. 267–268.
14. Gilbert W. // Cold Spring Harbor Symp. Quant. Biol. 1987. V. 52. P. 901–905.

15. Gilbert W., Glynnas M. // Gene. 1993. V. 135. P. 137–144.
16. Patrushev L.I. // Biochem. Mol. Biol. Internat. 1997. V. 41. P. 851–860.
17. Go M. // Nature. 1981. V. 291. P. 90–93.
18. Jensen E.O., Paludan K., Hyldig-Nielsen J.J., Jorgensen P., Marcker K.A. // Nature. 1981. V. 291. P. 677–679.
19. Moens L., Vanfleteren J., De Baere I., Jellie A.M., Tate W., Trotman C.N.A. // FEBS Lett. 1992. V. 312. P. 105–109.
20. Gilbert W., Marchionni M., McKnight G. // Cell. 1986. V. 46. P. 151–154.
21. Tittiger C., Whyard S., Walker V.K. // Nature. 1993. V. 361. P. 470–472.
22. Gilbert W., de Souza S.J., Long M. // Proc. Natl. Acad. Sci. USA. 1997. V. 94. P. 7698–7703.
23. De Souza S.J., Long M., Schoenbach L., Roy S.W., Gilbert W. // Gene. 1997. V. 205. P. 141–144.
24. Long M., de Souza S.J., Rosenberg C., Gilbert W. // Proc. Natl. Acad. Sci. USA. 1998. V. 95. P. 219–223.
25. De Souza S.J., Long M., Klein R.J., Roy S., Lin S., Gilbert W. // Proc. Natl. Acad. Sci. USA. 1998. V. 95. P. 5094–5099.
26. Shpakovski G.V., Acker J., Wintzerith M., Lacroix J.-F., Thuriaux P., Vigneron M. // Mol. Cell. Biol. 1995. V. 15. P. 4702–4710.
27. Khazak V., Estoja J., Cho H., Majors J., Sonoda G., Testa J.R., Golemis E.A. // Mol. Cell. Biol. 1998. V. 18. P. 1935–1945.
28. Larkin R.M., Guilfoyle T.J. // J. Biol. Chem. 1998. V. 273. P. 5631–5637.
29. Shpakovski G.V. // Gene. 1994. V. 147. P. 63–69.
30. Шпаковский Г.В., Лебеденко Е.Н. // Биоорган. химия. 1996. Т. 22. С. 938–940.
31. Sakurai H., Miyao T., Ishihama A. // Gene. 1996. V. 180. P. 63–67.
32. Шпаковский Г.В., Лебеденко Е.Н. // Биоорган. химия. 1997. Т. 23. С. 441–448.
33. Шпаковский Г.В., Лебеденко Е.Н. // Биоорган. химия. 1997. Т. 23. С. 988–991.
34. Sakurai H., Ishihama A. // Gene. 1997. V. 196. P. 165–174.
35. Шпаковский Г.В., Прошкин С.А., Каюшин А.Л., Коростелева М.Д., Лебеденко Е.Н. // Биоорган. химия. 1998. Т. 24. С. 119–125.
36. Шпаковский Г.В., Прошкин С.А., Лебеденко Е.Н. // Молекуляр. биология. 1998. Т. 32. С. 285–290.
37. Sakurai H., Kimura M., Ishihama A. // Gene. 1998. V. 221. P. 11–16.
38. Шпаковский Г.В., Баранова Г.М. // Биоорган. химия. 1999. Т. 25. (в печати).
39. Yamagishi M., Nomura M. // Gene. 1988. V. 74. P. 503–515.
40. Hirano T., Konoha G., Toda T., Yanagida M. // J. Cell Biol. 1989. V. 108. P. 243–253.
41. Шпаковский Г.В., Шематорова Е.К. // Биоорган. химия. 1998. Т. 24. С. 933–937.
42. Шпаковский Г.В., Шематорова Е.К. // Биоорган. химия. 1998. Т. 25. С. 791–796.
43. Nishida H., Sugiyama J. // Mycoscience. 1994. V. 35. P. 361–366.
44. Sipiczki M. // Ant. Leeuw. 1995. V. 68. P. 119–149.
45. Prabhala G., Rosenberg G. H., Kaufer N. F. // Yeast. 1992. V. 8. P. 171–182.
46. Zhang M.Q., Marr T.G. // Nucl. Acids Res. 1994. V. 22. P. 1750–1759.
47. Ovchinnikov Yu.A., Monastyrskaya G.S., Gubanov V.V., Guryev S.O., Salomatina I.S., Shuyaeva T.M., Lipkin V.M., Sverdlov E.D. // Nucleic Acids Res. 1982. V. 10. P. 4035–4044.
48. Allison L.A., Moyle M., Shales M., Ingles C.J. // Cell. 1985. V. 42. P. 599–610.
49. Sweetser D., Nonet M., Young R.A. // Proc. Natl. Acad. Sci. USA. 1987. V. 84. P. 1192–1196.
50. Grachev M.A., Lukhtanov E.A., Mustaev A.A., Zaychikov E.F., Abdukayumov M.N., Rabinov I.V., Richter V.I., Skoblov Y.S., Chistyakov P.G. // Eur. J. Biochem. 1989. V. 180. P. 577–581.
51. Riva M., Carles C., Sentenac A., Grachev M.A., Mustaev A.A., Zaychikov E.F. // J. Biol. Chem. 1990. V. 265. P. 16498–16503.
52. Zaychikov E., Martin E., Denissova L., Kozlov M., Heuman H., Nikiforov V., Goldfarb A., Mustaev A. // Science. 1996. V. 273. P. 107–109.
53. Thuriaux P., Sentenac A. // The Molecular and Cellular Biology of the Yeast Saccharomyces: Gene Expression / Eds J.R. Broach, J.R. Pringle, E.W. Jones. Cold Spring Harbor; N.Y.: Cold Spring Harbor Laboratory Press, 1992. V. II. P. 1–48.
54. Martindale D.W. // Nucl. Acids Res. 1990. V. 18. P. 2953–2959.
55. Lalo D., Carles C., Sentenac A., Thuriaux P. // Proc. Natl. Acad. Sci. USA. 1993. V. 90. P. 5524–5528.
56. Ulmasov T., Larkin R. M., Guilfoyle T.J. // J. Biol. Chem. 1996. V. 271. P. 5085–5094.
57. Larkin R.M., Guilfoyle T.J. // J. Biol. Chem. 1997. V. 272. P. 12824–12830.
58. Kimura M., Ishiguro A., Ishihama A. // J. Biol. Chem. 1997. V. 272. P. 25851–25855.
59. Svetlov V., Nolan K., Burgess R.R. // J. Biol. Chem. 1998. V. 273. P. 10827–10830.
60. Valenzuela P., Bell G.I., Weinberg F., Rutter W.J. // Biochem. Biophys. Res. Commun. 1976. V. 71. P. 1319–1325.
61. Woychik N.A., Liao S.-M., Kolodziej P., Young R.A. // Genes Dev. 1990. V. 4. P. 313–323.
62. Carles C., Treich I., Bouet F., Riva M., Sentenac A. // J. Biol. Chem. 1991. V. 266. P. 24092–24096.
63. Gadot O., Mariotte-Labarre S., Chedin S., Quemeineur E., Carles C., Sentenac A., Thuriaux P. // Mol. Cell. Biol. 1997. V. 17. P. 1787–1795.
64. Woychik N.A., Young R.A. // Transcription Mechanisms and Regulation / Eds R.C. Conaway, J.W. Conaway. N.Y.: Raven Press, 1994. P. 227–242.
65. Chédin S., Riva M., Schultz P., Sentenac A., Carles C. // Genes Dev. 1998. V. 12. P. 3857–3871.
66. Шпаковский Г.В., Лебеденко Е.Н. // Биоорган. химия. 1997. Т. 24. С. 877–880.
67. Chédin S., Ferri M.L., Peyroche G., Andrau J.-C., Jourdain S., Lefevre O., Werner M., Carles C., Sentenac A. // Cold Spring Harbor Symp. Quant. Biol. 1998. V. 63. P. 381–389.
68. Klenk H.P., Palm P., Lottspeich F., Zillig W. // Proc. Natl. Acad. Sci. USA. 1992. V. 89. P. 407–410.
69. Lanzendorfer M., Langer D., Hain J., Klenk H.-P., Holz I., Arnoldhammer I., Zillig W. // System. Appl. Microbiol. 1994. V. 16. P. 656–664.

70. Langer D.J., Hain J., Thuriaux P., Zillig W. // Proc. Natl. Acad. Sci. USA. 1995. V. 92. P. 5768–5772.
71. Eloranta J.J., Kato A., Teng M.S., Weinzierl R.O.J. // Nucl. Acids Res. 1998. V. 26. P. 5562–5567.
72. Шпаковский Г.В., Лебеденко Е.Н., Тюрьо П. // Биоорган. химия. 1997. Т. 23. С. 110–117.
73. Gadal O., Shpakovski G.V., Thuriaux P. // J. Biol. Chem. 1999. V. 274. P. 8421–8427.
74. Grachev M.A., Mustaev A.A., Zaychikov E.F., Lindner A.J., Hartmann G.R. // FEBS Lett. 1989. V. 250. P. 317–320.
75. Wlassoff W.A., Kimura M., Ishihama A. // J. Biol. Chem. 1999. V. 274. P. 5104–5113.
76. McKune K., Moore P.A., Hull M.W., Woychik N.A. // Mol. Cell. Biol. 1995. V. 15. P. 6895–6900.
77. Voutsina A., Riva M., Carles C., Alexandraki D. // Nucl. Acids Res. 1999. V. 27. P. 1047–1055.
78. Shpakovski G.V., Shematorova E.K. // Curr. Genet. 1999. V. 36. P. 208–214.
79. Pusch C., Wang Z., Roe B., Blin N. // Genomics. 1996. V. 34. P. 440–442.
80. Ulmasov T.N., Guilfoyle T.J. // J. Biol. Chem. 1992. V. 267. P. 23165–23169.
81. Schoen T.J., Chandrasekharappa S.C., Guru S.C., Mazuruk K., Chader G.J., Rodriguez I.R. // Biochim. Biophys. Acta. 1997. V. 1353. P. 39–49.
82. Krapp S., Kelly G., Reischl J., Weinzierl R.O.J., Matthews S. // Nature Struct. Biol. 1998. V. 5. P. 110–114.
83. Gilson P., McFadden G. // Proc. Natl. Acad. Sci. USA. 1996. V. 93. P. 7737–7742.
84. Финкельштейн А.В. // Докл. АН СССР. 1975. Т. 223. С. 744–747.
85. Ptitsyn O.B., Finkelstein A.V. // Biopolymers. 1983. V. 22. P. 15–25.
86. Acker J., Murroni O., Mattei M.-G., Kedinger C., Vignerion M. // Genomics. 1996. V. 32. P. 86–90.
87. Nouraini S., Archambault J., Friesen J.D. // Mol. Cell. Biol. 1996. V. 16. P. 5985–5996.
88. Treich I., Carles C., Riva M., Sentenac A. // Gene Expr. 1992. V. 2. P. 31–37.
89. Klenk H.P., Clayton R.A., Tomb J.F., White O., Nelson K.E. et al. // Nature. 1997. V. 390. P. 364–370.
90. Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y. et al. // DNA Res. 1998. V. 5. P. 55–76.
91. Lee M., Silver P.A. // Current Opinion in Genetics & Development. 1997. V. 7. P. 212–219.
92. Amegadzie B.Y., Ahn B.Y., Moss B. // J. Virol. 1992. V. 66. P. 3003–3010.
93. Crick F.H.C. // J. Mol. Biol. 1968. V. 38. P. 367–379.
94. Eigen M., Gardiner W., Schuster P., Winkler-Oswatitsch R. // Sci. Am. 1981. V. 244. P. 88–118.
95. De Duve C. // Gene. 1993. V. 135. P. 29–31.
96. Orgel L.E. // Trends Biochem. Sci. 1998. V. 23. P. 491–495.
97. Gilbert W. // Nature. 1986. V. 319. P. 618.
98. De Duve C. “Blueprint for a Cell: The Nature and Origin of Life”. Neil Patterson Publishers, Carolina Biological Supply Company, Burlington, NC, 1991.
99. Zhang B., Cech T.R. // Nature. 1997. V. 390. P. 96–100.
100. Chetverin A.B., Chetverina H.V., Demidenko A.A., Ugarov V.I. // Cell. 1997. V. 88. P. 503–513.
101. Gilson P.R., Maier U.-G., McFadden G.I. // Current Opin. Genet. & Dev. 1997. V. 7. P. 800–806.
102. Шпаковский Г.В. // Биоорган. химия. 1998. Т. 24. С. 154–155.
103. Murzin A.G. // EMBO J. 1993. V. 12. P. 861–867.
104. Ramakrishnan V., White S.W. // Trends Biochem. Sci. 1998. V. 23. P. 208–212.

## Molecular Evolution and Structure of Subunits of Nuclear RNA Polymerases of Eukaryotes in Context of the Exon–Intron Organization of Their Genes

G. V. Shpakovski<sup>#</sup> and E. N. Lebedenko

Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences,  
ul. Miklukho-Maklaya 16/10, GSP-7 Moscow, 117871 Russia

Analysis of literary data (for *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, *Arabidopsis thaliana*, *Homo sapiens*, and some other Eucarya) and our data (for *Schizosaccharomyces pombe*) on the exon–intron organization of the genes encoding subunits of nuclear RNA polymerases showed that introns in the orthologous genes from different organisms are arranged nonrandomly, namely, their positions, if projected on the map of the comparison of the amino acid sequences of the orthologous subunits, not infrequently coincide in evolutionarily distant species. As a rule, intron positions correspond to the boundaries of the structurally conserved regions (domains) or to the sites of possible turns of the polypeptide chain. For example, introns flank the secondary structure elements in the Rpb8 subunit with the known three-dimensional structure or the structure–function modules in subunits Rpb10 and Rpc10. These facts are in agreement with the idea of the ancient origin of introns, and with the notion of evolution of ancient protein sequences through the assembly of their genes from short protoexons selected by the nature as far back as the RNA world times. Comparative analysis of the primary structures of the subunits of eukaryotic RNA polymerases allowed us to reveal a nuclear localization signal in subunit Rpb10 and some hypothetical archaeal homologues of subunit Rpc10.

**Key words:** subunits of nuclear RNA polymerases, gene organization; evolution of the exon–intron structure; secondary structure modules; exon theory of genes.

<sup>#</sup> To whom correspondence should be addressed; phone: +7 (095) 330-6583; fax: +7 (095) 335-7103; e-mail: gvs@ibch.siobc.ras.ru.