

СТРУКТУРНО-ФУНКЦИОНАЛЬНЫЙ АНАЛИЗ БИОПОЛИМЕРОВ И ИХ КОМПЛЕКСОВ

УДК 577.323.24

СТРУКТУРНО-ТЕРМОДИНАМИЧЕСКАЯ ОРГАНИЗАЦИЯ МЕЖГЕННЫХ И ИНТРОННЫХ ПЕРВИЧНЫХ микроРНК ЧЕЛОВЕКА

© 2011 г. Д. В. Посредник¹, Т. В. Романовская¹, А. А. Радишевская², А. А. Хацкевич³,
А. Ю. Гончаревич⁴, В. С. Черепович¹, В. В. Гринеv^{1*}

¹Кафедра генетики биологического факультета Белорусского государственного университета,
Минск, 220030, Республика Беларусь

²School of Biological Sciences, The University of Edinburgh, Edinburgh, EH9 3JR, UK

³Лаборатория молекулярной генетики Института генетики и цитологии Национальной академии наук Беларуси,
Минск, 220072, Республика Беларусь

⁴Department of Veterinary Medicine, University of Cambridge, Cambridge, CB3 0ES, UK

Поступила в редакцию 22.09.2010 г.

Принята к печати 28.10.2010 г.

МикроРНК относятся к классу малых некодирующих РНК, принимающих участие в регуляции экспрессии генов в клетках эукариот. Синтезируются в ядре клетки в виде длинных предшественников, первичные микроРНК проходят многоэтапный биогенез, завершающийся образованием зрелых микроРНК. Один из критических этапов биогенеза микроРНК – распознавание и процессинг первичных микроРНК микропроцессорным комплексом. Считается, что ключевую роль на этом этапе играет вторичная структура первичных микроРНК, хотя специфические структурные элементы до конца не идентифицированы. В результате проведенного нами исследования, направленного на выяснение особенностей структурной и термодинамической организации межгенных и интронных микроРНК, показано, что между этими основными классами микроРНК человека нет статистически значимых различий. Это заключение может служить основанием для объединения последовательностей этих микроРНК в единый набор, который можно использовать как стандарт при поиске новых микроРНК. Кроме того, в структурно-термодинамической организации шпилек первичных микроРНК человека обнаружен ряд характерных особенностей, которые также найдут применение при поиске кандидатов на роль микроРНК в геноме человека и дизайне последовательностей, кодирующих искусственные микроРНК для генной терапии и функциональной геномики.

Ключевые слова: межгенные и интронные микроРНК, первичные микроРНК, вторичная структура, термодинамическая характеристика, сравнительный анализ.

STRUCTURAL AND THERMODYNAMIC FEATURES OF INTERGENIC AND INTRONIC HUMAN PRIMARY microRNAs, by D. V. Posrednik¹, T. V. Ramanouskaya¹, A. A. Radzsheuskaya², A. A. Khatskevich³, A. Y. Goncharevich⁴, V. S. Cherepovich¹, V. V. Grinev^{1*} (¹Department of Genetics, Faculty of Biology, Belarusian State University, Minsk, 220030 Republic of Belarus, *e-mail: grinev_vv@bsu.by; ²School of Biological Sciences, University of Edinburgh, Edinburgh, EH9 3JR, UK; ³Laboratory of Molecular Genetics, Institute for Genetics and Cytology, National Academy of Sciences of Belarus, Minsk, 220072 Republic of Belarus; ⁴Department of Veterinary Medicine, University of Cambridge, Cambridge, CB3 0ES, UK). MicroRNAs are small non-coding RNAs involved in the regulation of gene expression in eukaryotes. Being transcribed in the nucleus in the form of a long primary precursor a primary microRNA undergoes multistep biogenesis that leads to the formation of a mature microRNA. One of the critical biogenesis steps is the recognition and processing of primary microRNAs by microprocessor complex. A primary microRNA secondary structure is considered to play a key role during this biogenesis step, but specific structural motives have not been identified yet. In the result of the study on structural and thermodynamic features of intergenic and intronic microRNAs we have demonstrated that the two main classes of human microRNAs have no statistically significant difference in respect of these parameters. The conclusion may be a base for combining the two class microRNA sequences in a common set for the search for new microRNAs. Specific structural and thermodynamic features in human primary microRNA transcript hairpins have been found; the data may be used for search of new microRNA candidates in human genome and for a design of artificial microRNA sequences for gene therapy and functional genomics purposes.

Keywords: intergenic and intronic microRNAs, primary microRNAs, secondary structure, thermodynamic characterization, comparative analysis.

* Эл. почта: grinev_vv@bsu.by

ВВЕДЕНИЕ

МикроРНК – класс малых РНК, открытых в 1993 году [1]. Гены, кодирующие микроРНК, транскрибируются с помощью РНК-полимеразы II с образованием длинных (до 10 тыс. н.) транскриптов, кепированных по 5'-концу и полиаденилированных по 3'-концу [2, 3]. Такие транскрипты, называемые первичными микроРНК, содержат стабильные шпильки типа “стебель-петля”, которые фланкированы 5'- и 3'-концевыми базальными сегментами и включают собственно микроРНК [4, 5]. На первом этапе биогенеза первичных микроРНК происходит распознавание и вырезание шпильки с помощью микропроцессорного комплекса. Этот процесс протекает в ядре и приводит к высвобождению предшественника микроРНК, или пре-микроРНК [6–8]. В дальнейшем пре-микроРНК системой ядерно-цитоплазматического транспорта (экспорта ядра) направляется в цитоплазму [9], где проходит второй этап биогенеза, в который вовлечен РНКазы III-подобный фермент Dicer, удаляющий из молекулы терминальную неструктурированную петлю [10]. Итоговый РНК-дуплекс инициирует сборку мультисубъединичного RISC-комплекса, который осуществляет расплетание двуцепочечной РНК, селекцию зрелой микроРНК и, используя ее в качестве гида, распознает и разрезает РНК-мишень или подавляет ее трансляцию [11, 12].

Один из критичных этапов описанной выше схеме биогенеза микроРНК – распознавание и процессинг первичных микроРНК микропроцессорным комплексом. На сегодняшний день нет единого представления о том, как это происходит, хотя есть предположение, что ключевую роль в этом процессе играют особенности вторичной структуры первичных микроРНК и способность микропроцессорного комплекса распознавать эти особенности [4, 5]. Между тем, понимание молекулярного механизма процессинга первичных микроРНК необходимо как при исследованиях, связанных с биоинформационным и экспериментальным поиском новых микроРНК, так и при дизайне последовательностей, кодирующих искусственные микроРНК. На сегодняшний день нет ясности и в вопросе о структурных и/или термодинамических различиях в организации двух наиболее многочисленных групп микроРНК человека: межгенных и интронных [13–15], что не позволяет использовать их в составе единого набора как инструмента биоинформационного поиска новых микроРНК. На основании вышеизложенных фактов мы решили провести компьютерный структурно-термодинамический анализ экспериментально подтвержденных межгенных и интронных первичных микроРНК человека, чтобы выявить особенности их организации.

ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

Набор нуклеотидных последовательностей шпильки первичных микроРНК человека. Последовательности шпилек (вместе с фланкирующими базальными сегментами) первичных микроРНК человека получены из базы данных miRBase (выпуск 7.1) [16]. Всего проанализировано 326 последовательностей, из которых на долю экспериментально подтвержденных микроРНК приходится 75.2%, а на долю микроРНК, идентифицированных на основе гомологии, 24.8%. Среди экспериментально подтвержденных микроРНК большая часть представлена межгенными (68.2%) и интронными (22.4%) микроРНК. Все дальнейшие исследования проводили только на этих двух классах микроРНК (всего 222 последовательности), каждый из которых разделяли на два подкласса – по принадлежности к смысловой или антисмысловой цепи геномной ДНК (таблица).

Набор случайных межгенных и интронных нуклеотидных последовательностей генома человека. Случайные межгенные и интронные последовательности сгенерированы на основе эталонной сборки NCBI36/hg18 генома человека с помощью программы RandomGenerator v.01. Последовательности аннотированных интронов человека взяты из таблиц UCSC Known Genes геномного браузера UCSC Genome Browser [17]. При этом перекрывающиеся последовательности интронов, подвергающихся альтернативному сплайсингу, либо интронов, принадлежащих перекрывающимся генам, объединяли в единую последовательность, а последовательности, перекрывающиеся с экзонами, исключали. Последовательности межгенных регионов получены в соответствии с координатами аннотированных генов, скомпилированных геномным браузером UCSC Genome Browser. Последовательность не включали в окончательный список, если она принадлежала какому-либо известному гену микроРНК человека (в соответствии с базой данных miRBase, выпуск 7.1; [16]) или представляла собой аннотированный диспергированный повтор (в соответствии с базой данных Repbase Library повторяющихся последовательностей геномной ДНК человека [18, 19]) либо аннотированный простой или tandemный повтор (в соответствии с данными, полученными с помощью программы Tandem Repeats Finder [20]). В конечном итоге сгенерировано 6660 случайных межгенных и интронных последовательностей генома человека (30-кратный избыток по отношению к экспериментально подтвержденным микроРНК), которые подчинялись тем же закономерностям распределения по хромосомам, участкам генома, принадлежности к смысловой или антисмысловой цепи, а также линейному размеру, что и шпильки экспериментально подтвержденных межгенных и интронных микроРНК человека.

Структурный анализ шпилек первичных микроРНК человека. Вторичную структуру шпилек

Организация базы данных по экспериментально подтвержденным микроРНК человека*

Геномный контекст	Всего	Цепь геномной ДНК	
		+	-
Все микроРНК	245 (100.0%)	148 (60.4%)	97 (39.6%)
Интронные микроРНК	55 (22.4%)	28 (50.9%)	27 (49.1%)
Экзонные микроРНК	3 (1.2%)	1 (33.3%)	2 (66.7%)
Экзон/интронные микроРНК	6 (2.4%)	6 (100.0%)	0 (0.0%)
Межгенные микроРНК	167 (68.2%)	110 (65.9%)	57 (34.1%)
Локализация не установлена	14 (5.7%)	3 (21.4%)	11 (78.6%)

* Приведенные в таблице числовые значения соответствуют количеству экспериментально подтвержденных микроРНК, в скобках указано процентное содержание соответствующего компонента по базе данных miRBase (выпуск 7.1).

первичных микроРНК рассчитывали с помощью программного пакета Mfold 3.2 [21]. Выбрав варианты с термодинамически выгодными вторичными структурами, мы анализировали следующие параметры: 1) нуклеотидный состав шпилечных структур; 2) процент спаренных нуклеотидов; 3) частоту встречаемости неканонической пары нуклеотидов G-U; 4) линейный размер собственно терминальной петли; 5) природу нуклеотидной пары, “закрывающей” собственно терминальную петлю; 6) линейный размер и нуклеотидный профиль спирального участка терминальной неструктурированной петли; 7) тип и линейный размер структуры, прерывающей этот спиральный участок; 8) линейный размер терминальной неструктурированной петли; 9) линейный размер, симметричность и количество внутренних петель и петель-выступов в верхнем и нижнем стебле шпильки.

Термодинамический анализ шпилек первичных микроРНК человека. Минимальную свободную энергию (MFE) шпилек первичных микроРНК рассчитывали с помощью программного пакета Mfold 3.2 по алгоритму минимизации энергии в соответствии с “правилами ближайшего соседа” [21]. Линейную плотность MFE (MFE, нормализованную по длине последовательности) определяли в соответствии с предложениями, изложенными Первушиным (Perouchine D.D.) с соавторами [22]. Сравнительный анализ термодинамических параметров шпилек первичных микроРНК и вторичной структуры случайных межгенных и интронных последовательностей проводили с использованием приложения Zipfold из программного пакета UNAFold [23].

Оценка статистической значимости MFE вторичной структуры нуклеотидных последовательностей. Статистическую значимость MFE последовательности рассчитывали с помощью программного пакета RANFOLD 2.0 в соответствии с правилами рандомизирующего теста [24]. С помощью трех различных алгоритмов (моонуклеотидного и динуклеотидного “тасования”, а также “тасования” по модели Маркова 1-й степени) для каждой ориги-

нальной последовательности генерировали 6000 новых последовательностей. После расчета MFE вторичной структуры каждой из вновь сгенерированных последовательностей проводили статистическое сравнение полученных значений со значением MFE оригинальной последовательности.

Статистический анализ данных. Статистический анализ данных проводили с помощью пакета статистических программ STATISTICA 7.0. На первом этапе с помощью W-теста Шапиро–Уилка все цифровые данные проверяли на соответствие нормальному распределению. При отклонении от нормального распределения данные сравнивали между собой с помощью U-теста Манна–Уитни. При нормальном распределении использовали независимый t-тест Стьюдента. Статистически значимыми считались различия при $p < 0.05$. Корреляционная взаимосвязь между изучаемыми структурными и/или термодинамическими параметрами оценивали по величине и статистической значимости коэффициента ранговой корреляции Спирмена.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Структурная характеристика межгенных и интронных первичных микроРНК человека

В качестве базовой модели вторичной структуры первичных микроРНК человека в данном исследовании использована модель, предложенная в работе [5] и дополненная в работе [4]. Эта модель описывает первичный транскрипт микроРНК как совокупность термодинамически стабильной шпильки, включающей предшественник микроРНК и собственно зрелую микроРНК, и базальных сегментов (5'- и 3'-концевых фланкирующих последовательностей). В соответствии с предложенной моделью основные элементы шпильки следующие (рис. 1а): 1) терминальная неструктурированная петля (элемент E1), 2) верхняя спираль стебля длиной около 22 п.н. (элемент E2) и 3) нижняя спираль стебля размером около 11 п.н. (элемент E3). При этом терми-

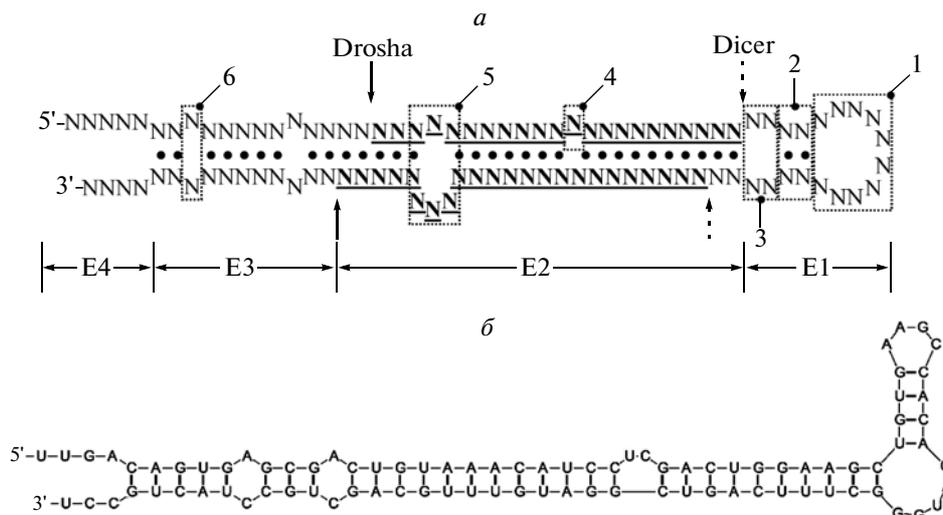


Рис. 1. Вторичная структура шпильки первичных микроРНК человека. *а* – Базовая модель. Фрагмент нуклеотидной последовательности первичной микроРНК записан с помощью условных нуклеотидов N. Нуклеотидная последовательность зрелой микроРНК (в смысловой или антисмысловой цепи верхней спирали стебля) выделена полужирным шрифтом и подчеркнута. На схеме указаны четыре составных элемента структуры: E1 – терминальная неструктурированная петля, E2 – верхняя спираль стебля, E3 – нижняя спираль стебля и E4 – базальные элементы. Сайты разрезания первичной микроРНК ферментами Drosha и Dicer отмечены соответственно сплошными и пунктирными вертикальными стрелками. В терминальной неструктурированной петле пунктирными прямоугольниками выделены собственно терминальная петля (1), спиральный участок (2) и структура, прерывающая этот спиральный участок (3). В стебле шпильки в качестве примеров пунктирными прямоугольниками выделены петля-выступ (4), несимметричная (5) и симметричная (6) внутренние петли. *б* – Пример вторичной структуры шпильки первичной микроРНК человека (has-miR-30a). Вторичная структура рассчитана с помощью программного пакета Mfold 3.2. Представленная конформация соответствует оптимальному сворачиванию нуклеотидной последовательности при MFE = –37.3 ккал/моль.

нальная неструктурированная петля включает собственно терминальную петлю, одну или больше пар оснований, образующих спиральный участок, и структуру (петлю-выступ или внутреннюю симметричную или несимметричную петлю), прерывающую этот спиральный участок (см. рис. 1*а*). Кроме того, как верхняя, так и нижняя спираль стебля содержат одну или несколько петель-выступов либо внутренних симметричных или несимметричных петель, образованных неспаренными нуклеотидами шпильечной структуры (рис. 1*а,б*).

Структурный анализ проведен только для шпилек первичных транскриптов микроРНК, базальные же сегменты (элемент E4 на рис. 1*а*) не анализировали. В образовании таких шпилек задействовано от 55 до 137 н. первичного транскрипта микроРНК – при линейном размере зрелых микроРНК от 17 до 25 н. (данные не показаны). В результате сравнительного структурного анализа шпилек первичных транскриптов межгенных и интронных микроРНК для них не установлено достоверных различий ни по одному из изученных структурных параметров. Также не обнаружено различий между структурными элементами шпилек первичных транскриптов микроРНК, локализованных в смысловой или антисмысловой цепи геномной ДНК. Эти данные послужили для нас основанием к объединению межгенных и интронных микроРНК в

единый набор. Ключевые результаты структурного анализа, проведенного повторно уже на объединенном наборе микроРНК, представлены ниже.

Согласно независимому *t*-тесту Стьюдента и *U*-тесту Манна–Уитни нуклеотидный состав шпилек оказался неоднородным, так как в парах сравнения A/C, A/G, A/U, C/G, C/U и G/U частоты встречаемости нуклеотидов достоверно отличаются при $p < 0.001$ и $p < 0.00001$, соответственно для *t*- и *U*-теста. При этом следует отметить несколько моментов в отношении нуклеотидного состава шпилек первичных микроРНК: 1) частоты встречаемости отдельных нуклеотидов лежат в диапазоне от $21.7 \pm 5.0\%$ (для C) до $29.2 \pm 4.9\%$ (для U) (рис. 2*а*); 2) несмотря на низкое содержание C, частота встречаемости нуклеотидов G + C близка к 50% ($47.4 \pm 7.8\%$), что обусловлено более высокой частотой встречаемости G ($25.8 \pm 4.1\%$), и, кроме того, этот параметр находится в обратной зависимости от частоты встречаемости A ($23.3 \pm 4.6\%$; $r = -0.80$) и U ($29.2 \pm 4.9\%$; $r = -0.79$) (см. рис. 2*а*); 3) *W*-тест Шапиро–Уилка показывает достоверное ($p < 0.00001$) отклонение характера распределения частоты встречаемости C от нормального распределения (рис. 2*б*); 4) статистически значимые ($p < 0.05$) величины коэффициентов ранговой корреляции Спирмена показывают, что частоты встречаемости A и U находятся в обратной зависимости от частоты встречаемости C ($r = -0.69$

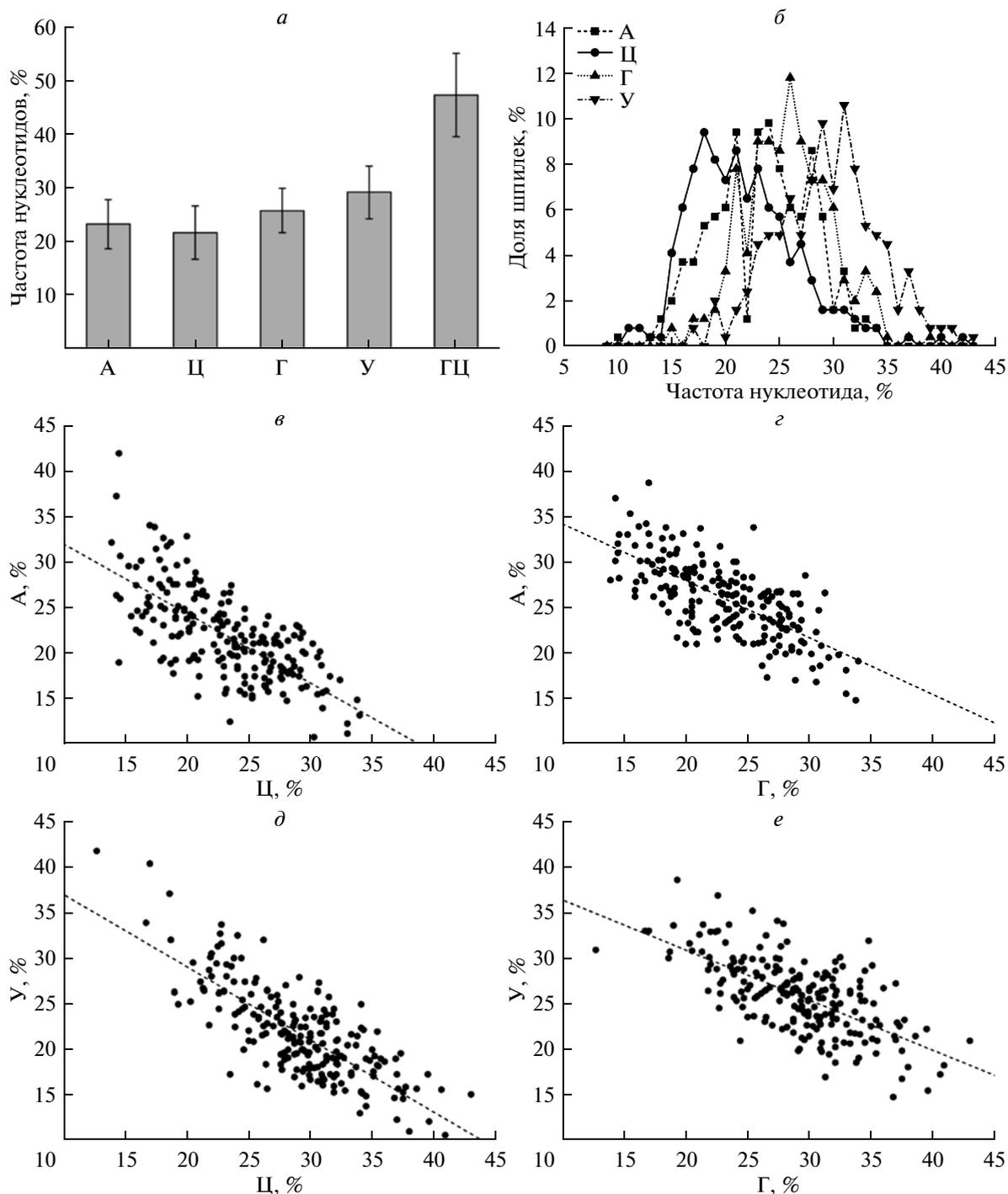


Рис. 2. Нуклеотидный состав шпилек экспериментально подтвержденных межгенных и интронных первичных микроРНК человека. *a* – Обобщенные данные по нуклеотидному составу шпилек представлены как среднее значение \pm стандартное отклонение; *б* – распределение частот встречаемости нуклеотидов в шпильках; *в*, *г* – графики рассеяния для нуклеотидных пар сравнения А/С и А/Г соответственно; *д*, *е* – графики рассеяния для нуклеотидных пар сравнения U/C и U/G соответственно.

и $r = -0.73$ соответственно) и G ($r = -0.70$ и $r = -0.63$ соответственно) и положительно коррелируют друг с другом ($r = 0.31$) (рис. 2*в–е*). Эти данные указывают на неслучайный характер нуклеотидного состава шпилек первичных микроРНК человека и, в част-

ности, на степень вовлеченности нуклеотидов в формирование вторичной структуры РНК-транскрипта. Так, в образование комплементарных пар в шпильках первичных микроРНК человека в среднем вовлечено $73.1 \pm 6.0\%$ нуклеотидов, что обеспе-

чивает стабильную вторичную структуру этих шпилек, причем на долю неканонической пары G-U приходится $12.4 \pm 5.3\%$ всех спаренных нуклеотидов.

На собственно терминальную петлю приходится от 3 до 18 н. первичных микроРНК человека, причем в 75.1% случаев ее размер меньше 10 н., а в 24.9% — не менее 10 н. (данные не показаны). Проведенный нами корреляционный анализ показал, что: 1) существует слабая, но статистически значимая ($p < 0.05$) положительная взаимосвязь между линейным размером собственно терминальной петли и протяженностью спирального участка всей терминальной неструктурированной петли ($r = 0.37$); 2) при увеличении линейного размера собственно терминальной петли протяженность структуры (внутренней петли или петли-выступа), прерывающей спиральный участок, уменьшается ($p < 0.05$; $r = -0.33$); 3) наблюдается хотя и слабая, но статистически значимая ($p < 0.05$) обратная взаимосвязь между линейным размером собственно терминальной петли и суммарной протяженностью внутренних петель шпилек первичных микроРНК ($r = -0.20$).

Анализ частоты встречаемости нуклеотидов, образующих пару, которая “закрывает” собственно терминальную петлю, и последующий корреляционный анализ не выявили какой-либо взаимосвязи между линейным размером самой петли и составом “закрывающей” пары нуклеотидов. Так, обнаружено, что собственно терминальные петли размером 5 и 12 н. преимущественно “закрываются” стабильной парой G-C, а петли размером 8 и 13 н. — термодинамически менее стабильной парой A-U (рис. 3а).

Выявлено, что нуклеотидный профиль спирального участка (рис. 3в), расположенного между собственно терминальной петлей и структурой (внутренней петлей или петлей-выступом), прерывающей этот участок, сходен с нуклеотидным профилем целой шпилеки, за исключением позиции 9, где в спиральном участке на долю C приходится 41.7%, а на долю G 37.5% всех нуклеотидов, встречающихся в этой позиции, в то время как для целой шпилеки это соответствует 21.7 и 25.8% (рис. 3б; логограмма сгенерирована с помощью интернет-сервиса WebLogo [25]). В 59.5% случаев этот спиральный участок прерывается внутренней петлей, а в 40.5% случаев — петлей-выступом, линейный размер которых находится в диапазоне от 2 до 13 н. и от 1 до 11 н. соответственно (данные не показаны).

К отличительным особенностям шпилек первичных микроРНК человека относится наличие внутренних петель и/или петель-выступов, локализованных в верхней и/или нижней спирали стебля. При линейном размере от 2 до 15 н. и общем количестве до 7 штук на одну шпилеку на внутренние петли приходится до 31 н. от всех нуклеотидов шпилеки, причем, как по количеству, так и по линейному размеру, здесь преобладают симметричные внутренние петли (рис. 3з, д). Идентичные результа-

ты получены и для петель-выступов (рис. 3з), однако характер частотного распределения шпилек в зависимости от количества нуклеотидов, входящих на петли-выступы, отличается от внутренних петель (рис. 3е). Интересно, что корреляционный анализ показал хотя и слабую, но статистически значимую ($p < 0.05$) обратную зависимость между общим количеством петель-выступов в шпилеках и размером спирального участка терминальной неструктурированной петли ($r = -0.22$), а также между общим количеством внутренних петель и петель-выступов внутри шпилеки ($r = -0.46$).

Термодинамическая характеристика межгенных и интронных первичных микроРНК человека

Как и в случае структурных параметров, для термодинамического анализа отобрали только шпилеки первичных микроРНК, исключив базальные сегменты. Поскольку между межгенными и интронными микроРНК, а также микроРНК из смысловых и антисмысловых цепи геномной ДНК не обнаружено никаких достоверных различий ни по одному из исследованных термодинамических параметров, то мы представляем результаты обобщенного анализа.

По термодинамической стабильности шпилек первичных микроРНК и случайных межгенных и интронных нуклеотидных последовательностей генома человека сравниваемые последовательности достоверно отличаются друг от друга как по абсолютным значениям MFE, так и по линейной плотности MFE. Разброс значений MFE для шпилек первичных микроРНК человека укладывается в диапазон от -67.6 ккал/моль до -15.9 ккал/моль (с медианой -38.9 ккал/моль), что дает 8.6%-ное перекрытие со случайными межгенными и интронными геномными последовательностями, для которых этот же параметр лежит в диапазоне значений от -27.2 ккал/моль до 0.61 ккал/моль (с медианой -6.2 ккал/моль). Нормализация MFE по длине позволяет уменьшить степень перекрытия между сравниваемыми последовательностями до 1.2% (рис. 4а, б).

Результаты рандомизированного теста, проведенного с помощью программы RANFOLD 2.0, согласуются с вышеприведенными данными. Так, среди шпилек первичных микроРНК человека уровню значимости $p < 0.05$ по всем трем моделям “тасования” нуклеотидов удовлетворяют не менее 95% последовательностей, в то же время среди случайных межгенных и интронных нуклеотидных последовательностей этому требованию удовлетворяет только 2.0% вариантов (рис. 4в, г).

Дестабилизирующий эффект в состоянии шпилек первичных микроРНК человека вносит собственно терминальная петля, а также внутренние петли и петли-выступы верхней и нижней спиралей стебля, которые в совокупности повышают свободную энергию шпилек. Так, анализ собственно терминальных

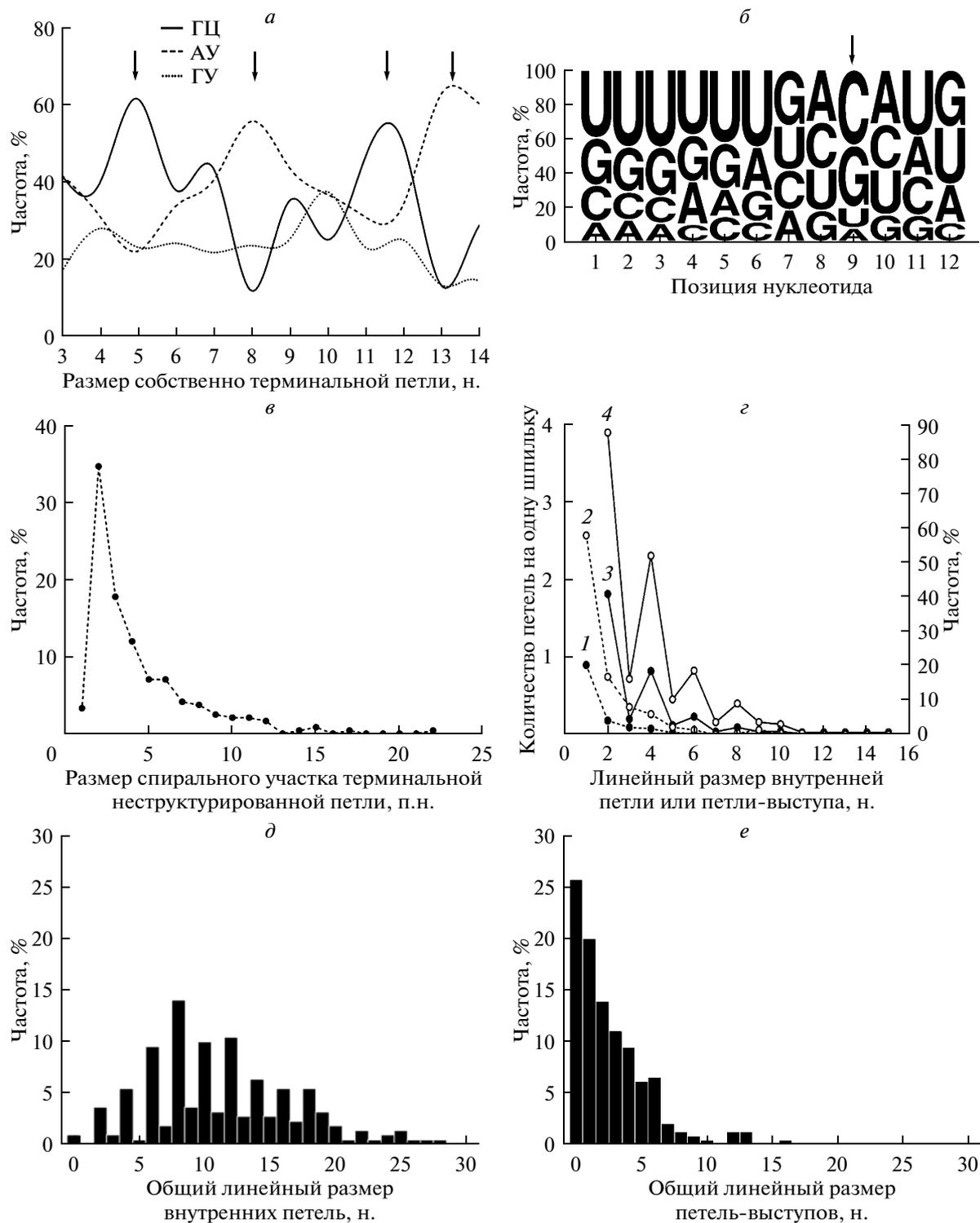


Рис. 3. Особенности организации шпилек экспериментально подтвержденных межгенных и интронных первичных микроРНК человека. *а* – Частота встречаемости G-C, A-U и G-U пар нуклеотидов, закрывающих собственно терминальную петлю, в зависимости от ее линейного размера. Вертикальными стрелками отмечены критические пики. *б* – Нуклеотидный профиль спирального участка терминальной неструктурированной петли. Показаны первые 12 позиций нуклеотидов. Первая позиция соответствует паре нуклеотидов, закрывающей концевую петлю. Вертикальной стрелкой отмечена позиция со значимым частотным отклонением нуклеотидов. *в* – Размер спирального участка терминальной неструктурированной петли шпилек первичных микроРНК человека. *г* – Распределение линейного размера внутренних петель и петель-выступов по их общей частоте и количеству на одну шпильку первичных микроРНК человека. *1* – Петли-выступы, количество; *2* – петли-выступы, частота; *3* – внутренние петли, количество; *4* – внутренние петли, частота. *д* – Частотное распределение шпилек первичных микроРНК человека по общему линейному размеру их внутренних петель. *е* – Частотное распределение шпилек первичных микроРНК человека по общему линейному размеру их петель-выступов.

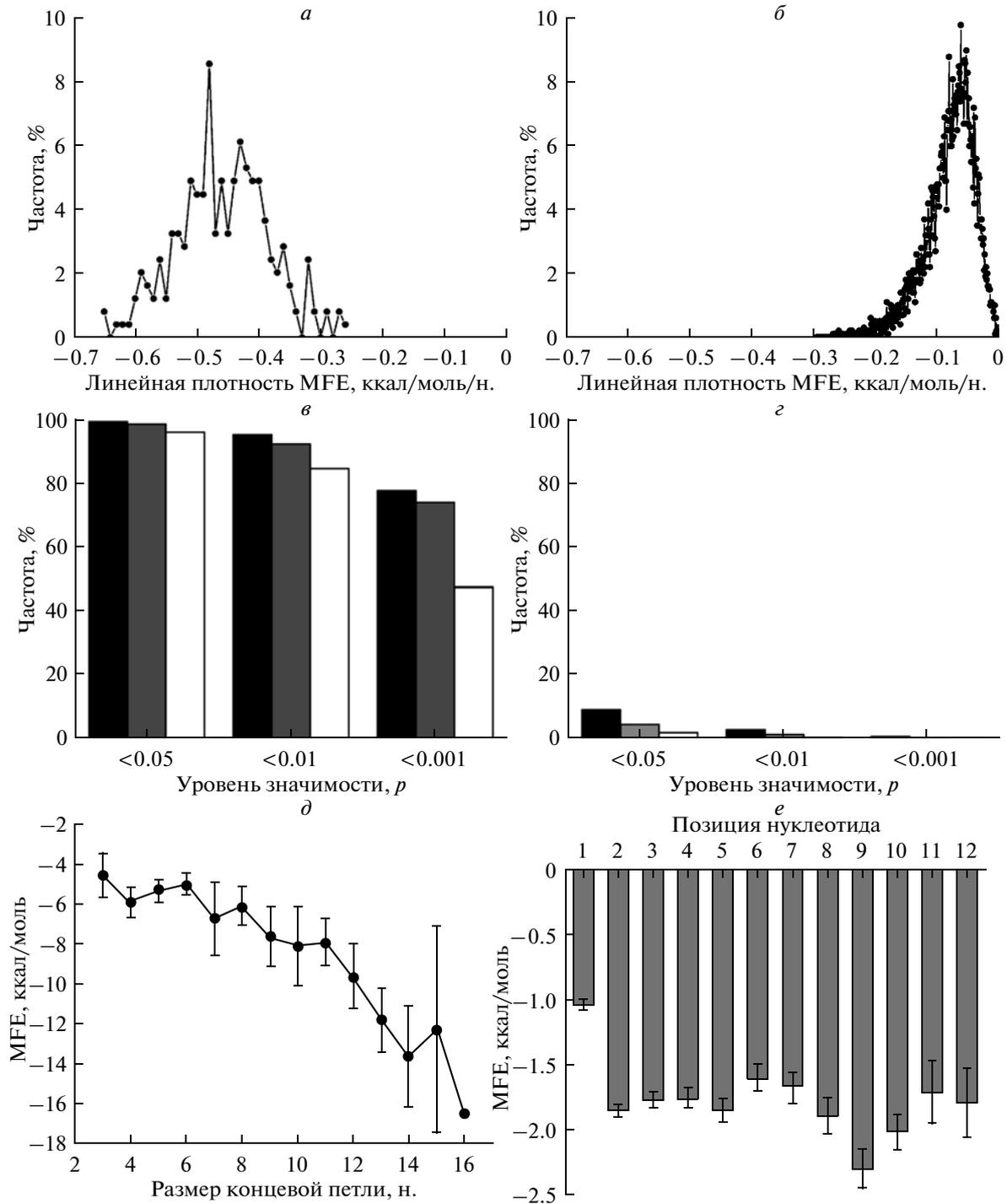


Рис. 4. Термодинамические характеристики шпилек межгенных и интронных первичных микроРНК человека. *a* – Линейная плотность MFE шпилек первичных микроРНК человека; *б* – линейная плотность MFE случайных межгенных и интронных нуклеотидных последовательностей генома человека; *в* – статистическая значимость MFE шпилек первичных микроРНК человека; *з* – статистическая значимость MFE вторичной структуры случайных межгенных и интронных нуклеотидных последовательностей генома человека. Представлены результаты расчетов уровня значимости с использованием мононуклеотидной (■) и динуклеотидной модели (▒), а также модели Маркова 1-й степени (□). В частях *в* и *з* рисунка показан процент последовательностей, удовлетворяющих соответствующему уровню значимости. *д* – Взаимосвязь между размером собственно терминальной петли и термодинамической стабильностью спирального участка терминальной неструктурированной петли шпилек первичных микроРНК человека. Показаны первые 12 позиций нуклеотидов. Первая позиция соответствует паре нуклеотидов, “закрывающей” собственно терминальную петлю.

петель показал, что их дестабилизирующая энергия лежит в диапазоне от 2.7 до 6.7 ккал/моль, напрямую коррелируя с линейным размером самой петли (данные не показаны). Кроме того, нами обнаружена статистически значимая ($p < 0.05$) прямая взаимосвязь ($r = 0.36$) между размером собственно терминальной петли шпильки и термодинамической стабильностью спирального участка терминальной неструктурированной петли (рис. 4д).

Термодинамический профиль спирального участка терминальной неструктурированной петли, полученный для первых двенадцати нуклеотидных позиций, в целом согласуется с данными по его нуклеотидному профилю (см. рис. 3б). Так, особенность нуклеотидного состава по позиции 9, отмеченная нами ранее, отражается и на термодинамической стабильности этого участка: значение MFE для этой позиции составляет величину -2.3 ± 0.7 ккал/моль, что достоверно ниже значения этого же показателя для остальных позиций (рис. 4е).

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

В представленном исследовании проведен структурно-термодинамический анализ шпилек межгенных и интронных микроРНК человека, который основан на результатах расчета вторичной структуры изучаемых последовательностей с помощью программного пакета Mfold 3.2. Выбор этого программного пакета обусловлен двумя причинами: во-первых, на данный момент это давно используемая и проверенная многими исследователями программа, которая позволяет провести надежные расчеты вторичной структуры нуклеиновых кислот [21]; во-вторых, это единственный в своем роде программный пакет, расчеты которого в отношении первичных микроРНК получили экспериментальное подтверждение (точность $\geq 83\%$) [26].

Путем сравнительного анализа шпилек первичных межгенных и интронных микроРНК мы установили отсутствие между ними достоверных различий по всем исследованным структурным или термодинамическим параметрам. Полученные результаты позволяют предположить, что, несмотря на постулируемые различия в биогенезе межгенных и интронных первичных микроРНК [8, 13–15], они не связаны со структурными или термодинамическими особенностями организации их первичных транскриптов.

Особенности организации вторичной структуры первичных транскриптов микроРНК человека могут быть описаны одной из двух конкурирующих в настоящее время моделей: 1) предложенной группой под руководством Брайна Р. Куллена (Bryan R. Cullen) [5] и 2) разработанной Ханом (Han J.) с соавторами [4]. Обе модели экспериментально обоснованы и нацелены на объяснение феномена распознавания первичных транскриптов микроРНК

микропроцессорным комплексом ядра. Согласно первой модели, успешный биогенез шпилечных структур, входящих в состав первичных микроРНК, связан с распознаванием микропроцессорным комплексом их терминальных петель, причем линейный размер последних должен быть не менее 10 н. Поскольку эта модель основана на данных мутационного анализа лишь четырех различных микроРНК человека, мы проверили ее ключевое положение — наличие у шпильки терминальной петли размером не менее 10 н., — путем биоинформационного анализа всех экспериментально подтвержденных для клеток человека межгенных и интронных микроРНК. Полученные нами результаты указывают на то, что лишь небольшая часть (24.9%) всех шпилек удовлетворяет этому требованию — обычно размер терминальной петли меньше 10 н. (75.1%).

В данной работе собственно терминальная петля рассматривалась как часть терминальной неструктурированной петли шпилек первичных микроРНК. Такой подход позволил нам установить, что увеличение линейного размера и снижение термодинамической стабильности собственно терминальной петли компенсируется удлинением спирального участка терминальной неструктурированной петли и повышением его термодинамической стабильности, а также уменьшением линейного размера внутренней петли или петли-выступа, прерывающей этот спиральный участок. Справедливо и обратное утверждение: чем меньше собственно терминальная петля, тем короче и менее стабилен спиральный участок терминальной неструктурированной петли, тем крупнее и менее стабильна та структура, которая его прерывает.

Ранее показано, что область ортологичных генов микроРНК, кодирующая терминальную неструктурированную петлю, характеризуется высокой вариабельностью [27]. Это открытие послужило основанием для предположения о том, что нуклеотидный состав терминальной неструктурированной петли шпилек первичных микроРНК не играет важной роли в биогенезе и функционировании микроРНК [3, 27]. Тем не менее, наши данные указывают на то, что составные части терминальной неструктурированной петли взаимосвязаны между собой и с остальными частями шпильки. Поэтому можно предположить, что этот структурный элемент шпилек все же важен для биогенеза микроРНК, на что указывают и данные других авторов [5]. Однако остается не совсем ясной причина обнаруженной нами взаимосвязи между линейным размером собственно терминальной петли и тем, какая пара нуклеотидов ее “закрывает”. Кроме того, остается открытым и вопрос о биологической значимости высокой частоты встречаемости (до 80%) нуклеотидов G и C в позиции 9 спирального участка терминальной неструктурированной петли.

Собственно терминальная петля, а также петля, прерывающая спиральный участок терминальной неструктурированной петли, представляют собой компоненты единой системы петель в шпильках первичных микроРНК. Проведенный нами детальный анализ частоты встречаемости, а также общий анализ типа, количества и размера внутренних петель и петель-выступов стебля шпилек первичных микроРНК показал, что: 1) внутренние петли и петли-выступы являются неотъемлемым компонентом вторичной структуры первичных микроРНК, отражающим, возможно, эволюционно сложившийся компромисс между необходимостью иметь определенную нуклеотидную последовательность зрелой микроРНК и необходимостью сохранить стабильную вторичную структуру, распознаваемую микропроцессорным комплексом в ядре и Dicer и RISC-комплексом в цитоплазме; 2) наличие таких структур позволяет удерживать шпильки в определенном диапазоне термодинамической стабильности, что, возможно, является важным фактором осуществления успешной “обработки” шпилек Dicer и RISC-комплексом [26].

Результаты проведенного нами термодинамического анализа согласуются с данными других авторов, показавших, что шпильки первичных транскриптов микроРНК обладают высокой термодинамической стабильностью, отличающей их от других, менее структурированных, классов РНК [24, 28]. Так что наши данные лишь уточняют и дополняют уже имеющиеся представления о термодинамической стабильности первичных микроРНК. В частности, используя в качестве контроля обширный набор случайных межгенных и интронных геномных последовательностей, мы показали, что по нормализованному значению MFE шпильки первичных транскриптов микроРНК только на 1.2% перекрываются со случайными последовательностями. Следовательно, нормализованное значение MFE может быть одним из критериев дискриминации между ложными микроРНК-подобными последовательностями и реальными кандидатами на новые микроРНК, а также альтернативой такому параметру, как статистическая значимость MFE, используемому при поиске новых микроРНК [24].

Таким образом, результаты проведенного нами исследования показывают, что между межгенными и интронными микроРНК нет статистически значимых структурных или термодинамических различий, что является основанием для включения их в единый набор, который можно использовать как инструмент для биоинформационного поиска новых микроРНК. Кроме того, мы обнаружили ряд особенностей в структурно-термодинамической организации шпилек первичных транскриптов микроРНК человека, что может быть использовано при биоинформационном поиске новых микроРНК в геноме человека, а также при дизайне последовательностей, ко-

дирующих искусственные микроРНК для решения задач геномной терапии и функциональной геномики.

Работа получила финансовую поддержку со стороны Международной ассоциации по содействию кооперации с учеными из новых независимых стран бывшего Советского Союза (ИНТАС, грант № 05-109-4921), а так же Белорусского республиканского фонда фундаментальных исследований (грант № Б09МС-028).

СПИСОК ЛИТЕРАТУРЫ

1. Lee R.C., Ambros V. 2001. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science*. **294**, 862–864.
2. Cai X., Hagedorn C.H., Cullen B.R. 2004. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA*. **10**, 1957–1966.
3. Olena A.F., Patton J.G. 2010. Genomic organization of microRNAs. *J. Cell. Physiol.* **222**, 540–545.
4. Han J., Lee Y., Yeom K.-H., Nam J.-W., Heo I., Rhee J.-K., Sohn S.Y., Cho Y., Zhang B.-T., Kim V.N. 2006. Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell*. **125**, 887–901.
5. Zeng Y., Yi R., Cullen B.R. 2005. Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha. *EMBO J.* **24**, 138–148.
6. Bartel D.P. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*. **116**, 281–297.
7. Han J., Lee Y., Yeom K.-H., Kim Y.-K., Jin H., Kim V.N. 2004. The Drosha-DGCR8 complex in primary microRNA processing. *Gen. Dev.* **18**, 3016–3027.
8. Naqvi A.R., Islam M.N., Choudhury N.R., Haq Q.M.R. 2009. The fascinating world of RNA interference. *Int. J. Biol. Sci.* **5**, 97–117.
9. Bohnsack M.T., Czaplinski K., Gorlich D. 2004. Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA*. **10**, 185–191.
10. Kim V.N. 2005. MicroRNA biogenesis: coordinated cropping and dicing. *Nat. Rev. Mol. Cell. Bio.* **6**, 1–11.
11. Gregory R.I., Chendrimada T.P., Cooch N., Shiekhattar R. 2005. Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. *Cell*. **123**, 631–640.
12. Погаев Е.И., Боринская С.А., Исламгулов Д.В., Григоренко А.П. 2008. МикроРНК человека в норме и патологии. *Молекуляр. биология*. **42**, 751–764.
13. Kim Y.K., Kim V.N. 2007. Processing of intronic microRNAs. *EMBO J.* **26**, 775–783.
14. Ruby J.G., Jan C.H., Bartel D.P. 2007. Intronic microRNA precursors that bypass Drosha processing. *Nature*. **448**, 83–86.
15. Ying S.-Y., Lin S.-L. 2006. Current perspectives in intronic microRNAs (miRNAs). *J. Biomed. Sci.* **13**, 5–15.
16. Griffiths-Jones S., Grocock R.J., van Dongen S., Bateman A., Enright A.J. 2006. miRBase: microRNA sequences, targets and gene nomenclature. *Nucl. Acids Res.* **34**, D140–D144.
17. Karolchik D., Baertsch R., Diekhans M., Furey T.S., Hinrichs A., Lu Y.T., Roskin K.M., Schwartz M., Sugnet C.W., Thomas D.J., Weber R.J., Haussler D.,

- Kent W.J. 2003. The UCSC genome browser database. *Nucl. Acids Res.* **31**, 51–54.
18. Kohany O., Gentles A.J., Hankus L., Jurka J. 2006. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinfo.* **7**, 474/doi:10.1186/1471-2105-7-474.
19. Smit A.F. 1999. Interspersed repeats and other elements of transposable elements in mammalian genomes. *Curr. Opin. Genet. Dev.* **9**, 657–663.
20. Benson G. 1999. Tandem repeats finder: A program to analyze DNA sequences. *Nucl. Acids Res.* **27**, 12–17.
21. Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucl. Acids Res.* **31**, 3406–3415.
22. Pervouchine D.D., Graber J.H., Kasif S. 2003. On the normalization of RNA equilibrium free energy to the length of the sequence. *Nucleic Acids Res.* **31**, e49.
23. Markham N.R., Zuker M. 2005. DINAMelt web server for nucleic acid melting prediction. *Nucl. Acids Res.* **33**, W577–W581.
24. Bonnet E., Wuyts J., Rouzé P., Van de Peer Y. 2004. Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences. *Bioinformatics.* **20**, 2911–2917.
25. Crooks G.E., Hon G., Chandonia J.-M., Brenner S.E. 2004. WebLogo: A sequence logo generator. *Genome Res.* **14**, 1188–1190.
26. Krol J., Sobczak K., Wilczynska U., Drath M., Jasinska A., Kaczynska D., Krzyzosiak W.J. 2004. Structural features of microRNA (miRNA) precursors and their relevance to miRNA biogenesis and their relevance to miRNA biogenesis and small interfering RNA/short hairpin RNA design. *J. Biol. Chem.* **279**, 42230–42239.
27. Altuvia Y., Landgraf P., Lithwick G., Elefant N., Pfeffer S., Aravin A., Brownstein M.J., Tuschl T., Margalit H. 2005. Clustering and conservation patterns of human microRNAs. *Nucl. Acids Res.* **33**, 2697–2706.
28. Clote P., Ferré F., Kranakis E., Krizanc D. 2005. Structural RNA has lower folding energy than random RNA of the same dinucleotide frequency. *RNA.* **11**, 578–591.