

Медицинский журнал. – 2009. №2. – С. 24-27.

**А.В. Бутвиловский, В.Э. Бутвиловский, Е.А. Черноус**

## **ИЗУЧЕНИЕ СТРАТЕГИИ КОДИРОВАНИЯ БЕЛКОВ**

*Белорусский государственный медицинский университет*

*Рассмотрены методы изучения стратегий кодирования белков (изучение частоты использования претерминальных кодонов, доли ГЦЗ-кодонов, определение показателя относительного использования синонимичных кодонов, дистанции МакАйнерни).*

*Ключевые слова: стратегия кодирования, претерминальные кодоны, ГЦЗ-кодоны, синонимичные кодоны, дистанция МакАйнерни.*

**A.V. Butvilovsky, V.E. Butvilovsky, E.A. Chernous**

## **STUDIES OF PROTEIN'S CODING STRATEGIES**

*We examined methods used to study protein's coding strategies (study of preterminal codons frequencies, GC3-codons portion, determination of relative usage of synonymous codons, McInerney's distance.*

*Key words: coding strategy, GC3-codons, preterminal codons, synonymous codons, McInerney's distance.*

Стратегия кодирования белка – это закономерность использования кодонов в соответствующих ему мРНК и ДНК [11]. Анализ использования кодонов в последовательностях нуклеиновых кислот стал возможным в 70-80-х годах прошлого столетия, когда в международных базах данных появилось достаточное количество секвенированных последовательностей РНК и ДНК [13]. С тех пор был установлен ряд важных закономерностей стратегий кодирования белков [1, 2, 11].

Многообразие возможных стратегий кодирования связано с вырожденностью генетического кода (в среднем на каждую из 20 аминокислот приходится три синонимичных кодона) [14]. В 1980-м году Р. Грэнтсем предположил, что каждый вид организмов имеет

оригинальную стратегию кодирования белков [13]. Позднее установлена вариация использования кодонов и у организмов одного вида, связанная с уровнем экспрессии гена [15, 19, 22], его размером [17], структурой мРНК [18], аминокислотным составом кодируемого белка [20] и другими факторами [14].

Стратегия кодирования оказывает влияние на помехоустойчивость процесса трансляции (путем уменьшения или увеличения частоты претерминальных кодонов), а также его скорость и точность (путем неравномерного использования синонимичных кодонов и неодинакового содержания в клетке изоакцепторных тРНК) [16].

Многие исследователи считают, что ГЦ-насыщенность (суммарное содержание гуанина и цитозина) его мРНК (и соответствующего кодирующего участка ДНК) является важнейшим фактором, определяющим стратегию кодирования белка [1, 8].

ГЦ-насыщенность – это один из факторов, обеспечивающих термодинамическую стабильность молекулы ДНК, поскольку между гуанином и цитозином двух ее цепочек образуются три водородные связи.

В соответствии с теорией Н. Суеоки в качестве фактора, определяющего стратегию кодирования белка, более корректно рассматривать мутационное давление, а ГЦ-насыщенность нуклеиновой кислоты является лишь его отражением [1]. Так, на примере мРНК, кодирующих алкогольдегидрогеназы класса 3 хордовых животных, было установлено, что значения показателей мутационного давления связаны с ГЦ-насыщенностью прямой сильной достоверной корреляционной связью ( $r = 0,98 \pm 0,062$ ,  $p < 0,001$ ) [6].

Мутационное давление – это фактор молекулярной эволюции, дающий материал для естественного отбора и обусловленный повышенной частотой возникновения и фиксации замен А и Т на Г и Ц относительно частоты возникновения и фиксации замен Г и Ц на А и Т (ГЦ-давление), или наоборот (АТ-давление) [11].

При изучении стратегии кодирования наиболее часто анализируются следующие показатели:

1. Частота использования претерминальных кодонов [2, 12].
2. Доля ГЦЗ-кодонов [3, 8].
3. Показатель относительного использования синонимичных кодонов [4, 9].
4. Дистанция Дж. МакАйнерни [7, 21].

Претерминальные кодоны (ПТК) – это кодоны, способные стать терминальными в результате одношаговой мутации и, следовательно, прервать синтез пептидной цепочки [6]. Теоретически, в процессе эволюции частота претерминальных кодонов должна снижаться.. Предположение, сделанное в начале 80-х годов XX века, о том, что естественный отбор постепенно закрепляет в популяции те мутации, которые приводят к снижению ПТК в кодирующих участках мРНК, было критически воспринято сторонниками нейтральной теории эволюции [12, 18]. По данным Ачинович О.В., для мембраносвязанных аденилатциклаз животных характерно уменьшение частоты использования претерминальных кодонов в кодирующих их мРНК в процессе эволюции [2]. При изучении мРНК, кодирующих алкогольдегидрогеназы класса 3 хордовых, установлено, что содержание ПТК контролируется естественным отбором в малой степени, и преимущественно обусловлено мутационным давлением ( $r = -0,81 \pm 0,207$ ,  $p < 0,001$ ) и, следовательно, ГЦ-насыщенностью ( $r = -0,79 \pm 0,216$ ,  $p < 0,001$ ) [1].

Сопоставление частот использования претерминальных кодонов в мРНК не представляет особых трудностей, за исключением случая, когда сравниваемые мРНК транслируются в соответствии с разными таблицами генетического кода. Так, например, кодоны АГА и АГГ являются терминальными в таблице для мРНК, кодирующих митохондриальные белки позвоночных (человека), и соответствуют серину в таблице для мРНК, кодирующих митохондриальные белки беспозвоночных животных. Разное

количество терминальных кодонов в этих двух таблицах генетического кода обуславливает различия количества ПТК в них.

В таблице генетического кода для митохондриальных белков позвоночных терминальными являются 4 кодона (УАА, УАГ, АГА, АГГ), смысловыми – 60 кодонов, претерминальными – 24 кодона, поэтому теоретическая частота ПТК составляет 40,0% ( $24/60 \times 100\%$ ). В таблице генетического кода для митохондриальных белков беспозвоночных животных терминальными являются 2 кодона (УАА, УАГ), смысловыми – 62 кодона, претерминальными – 14 кодонов, что дает теоретическую частоту ПТК, равную 22,6% ( $14/62 \times 100\%$ ).

В 2007 году был разработан способ сравнения частот ПТК в мРНК, транслируемых в соответствии с разными таблицами генетического кода, основанный на вычислении соотношения наблюдаемой и теоретической частоты использования ПТК [10]. Изучение мРНК, кодирующих митохондриальные белки человека (*H. sapiens*), паразитического круглого червя – трихинеллы (*T. spiralis*) и свободноживущего круглого червя – цианорабдитис (*C. elegans*), позволило установить, что среднее соотношение наблюдаемой и теоретически ожидаемой частот претерминальных кодонов в мРНК человека равно  $0,79 \pm 0,034$ , в мРНК трихинеллы –  $1,05 \pm 0,048$ , в мРНК цианорабдитис –  $1,25 \pm 0,036$  [9]. При этом различия между этими показателями человека и *C. elegans* достоверно ( $p < 0,05$ ) отличаются, а человека и *T. spiralis* статистически неразличимы. Это свидетельствует о наличии сходства стратегий кодирования митохондриальных белков человека и трихинеллы в мРНК.

Корреляционная связь между частотой ПТК в мРНК и их ГЦ-насыщенностью является обратной, сильной и достоверной. Этот факт был подтвержден при изучении эволюционных изменений мРНК, кодирующих алкогольдегидрогеназы хордовых, мембраносвязанные аденилатциклазы бактерий, беспозвоночных и позвоночных животных [1, 2, 12].

В случае изучения мРНК, транслируемых в соответствии с разными таблицами генетического кода, более корректно сравнивать соотношение наблюдаемой и теоретической частоты использования ПТК с ГЦ-насыщенностью [9]. Между соотношением наблюдаемых и теоретических частот ПТК и ГЦ-содержанием в мРНК, кодирующих митохондриальные белки человека, трихинеллы и цианорабдитис, обнаружена обратная, сильная и достоверная корреляционная связь ( $r = -0,78 \pm 0,107$ ,  $p < 0,001$ ). Получено уравнение регрессионного анализа для данных показателей:  $y = -0,02x + 1,70$ , где  $y$  – соотношение наблюдаемых и теоретических частот ПТК, а  $x$  – ГЦ-содержание. По данному уравнению вычислено, что при ГЦ-содержании, равном 35%, наблюдаемая частота ПТК равна теоретической (их соотношение равно 1). Следовательно, можно предположить, что при содержании гуанина и цитозина в мРНК более 35% частота претерминальных кодонов будет меньше таковой согласно таблицам генетического кода.

По мнению авторов, данное предположение по-новому раскрывает сущность взаимосвязи между двумя данными показателями и нуждается в тщательной проверке на большем количестве мРНК, транслируемым по разным таблицам генетического кода [9].

ГЦЗ-кодона – это кодона, содержащие в третьем положении гуанин или цитозин, исключая терминальные и невырожденные кодона (в стандартной таблице генетического кода – АУГ (метионин), УГГ (триптофан)) [6].

Определение доли ГЦЗ-кодонов проводится реже, чем вычисление частоты использования претерминальных кодонов. Однако в некоторых случаях (например, при изучении сопряжения эволюции между паразитами и их хозяевами на молекулярно-генетическом уровне) вычисление этого показателя становится крайне необходимым.

Так, при изучении стратегий кодирования митохондриальных белков человека и трихинеллы (как компонентов системы “паразит-хозяин”, формирующейся при трихинеллезе) установлено, что максимальная доля ГЦЗ-кодонов наблюдается в изученных мРНК человека,

меньшая – в мРНК трихинеллы, а минимальная – в мРНК цианорабдитис [8]. Доля ГЦЗ-кодонов в изученных мРНК человека более сходна с таковой в мРНК трихинеллы по сравнению с контролем. Обнаружено, что характер связи между ГЦ-насыщенностью и долей ГЦЗ-кодонов в мРНК, кодирующих митохондриальные белки человека и трихинеллы, сходен ( $r=0,89\pm0,097$ ) и достоверно ( $p<0,05$ ) отличается от такового у цианорабдитис ( $r=0,13\pm0,314$ ) [8].

Синонимичные (эквивалентные, изоакцепторные) кодоны – это кодоны, кодирующие одну и ту же аминокислоту. Группа синонимичных кодонов называется серией кодонов, которые подразделяются на:

- двукратно вырожденные (состоят из двух синонимичных триплетов),
- четырехкратно вырожденные (состоят из четырех триплетов),
- шестикратно вырожденные (состоят из шести триплетов) [6].

Показатель относительного использования синонимичных кодонов (RSCU, relative synonymous codon's usage) предложен для проведения корректных сравнений частот использования синонимичных кодонов в различных сериях. Показатель RSCU вычисляется по формуле:

$$RSCU_K = \frac{RSCU_{CK} \times n_K}{n_{CK}},$$

где  $RSCU_K$  – показатель относительного использования кодона A,  $RSCU_{CK}$  – количество кодонов в серии,  $n_K$  – частота использования кодона A,  $n_{CK}$  – частота использования всех кодонов серии [7].

Рассчитаем в качестве примера RSCU для кодона УГУ в мРНК, кодирующей НАДН-дегидрогеназу б человека (табл. 1 [7]).  $RSCU_{CK}$  равно 2 (серия представлена двумя кодонами – УГУ и УГЦ),  $n_K = 1$  (кодон УГУ используется в данной мРНК 1 раз),  $n_{CK} = 1$  (кодон УГУ используется 1 раз, кодон УГЦ не используется), тогда:

$$RSCU_{УГУ} = \frac{2 \times 1}{1} = 2$$

Таблица 1. Показатели RSCU и количество кодонов УГУ и УГЦ в мРНК, кодирующих субъединицу 6 НАДН-дегидрогеназы человека, трихинеллы и цианорабдитис

Организм/показатель	Кодон УГУ	Кодон УГЦ
Человек	n=1 RSCU=2,0	n= 0 RSCU=0
Трихинелла	n= 0 RSCU=0	n=2 RSCU=2,0
Цианорабдитис	n=3 RSCU=2,0	n= 0 RSCU=0

Кроме того возможно сравнение суммарных показателей RSCU группы кодонов (например, ГЦЗ-кодонов), соответствующих одной и той же аминокислоте (рис. 1 [1]).

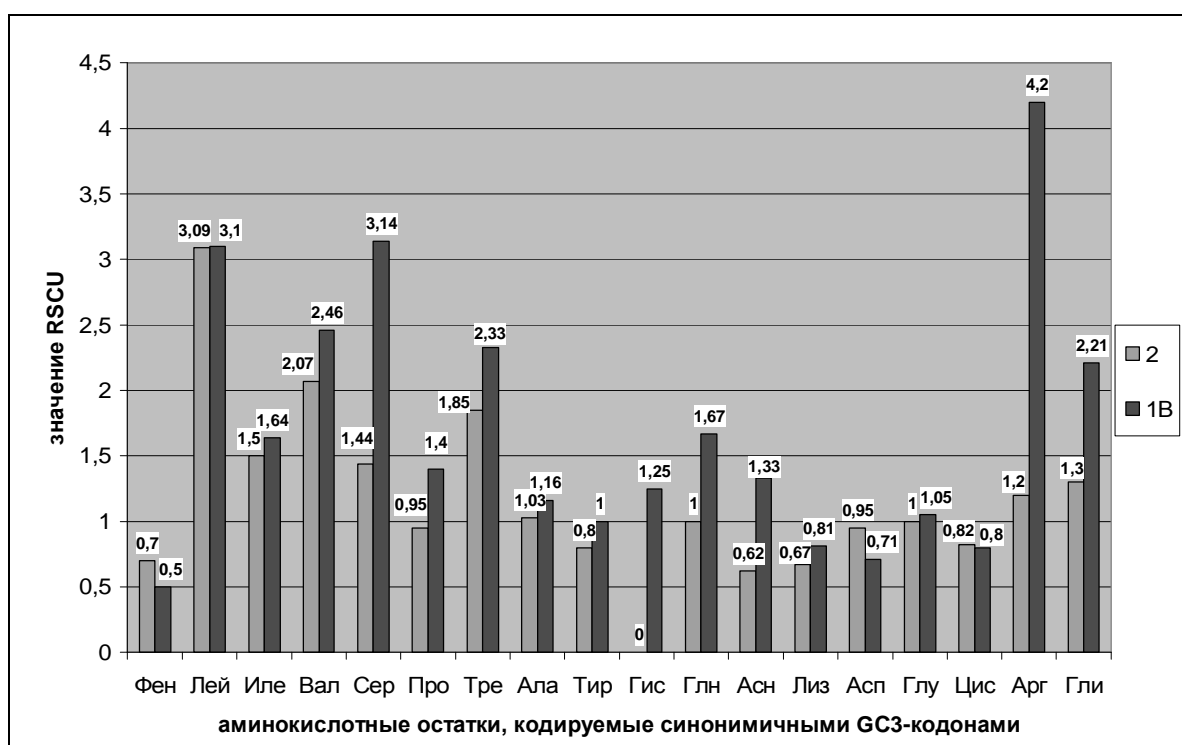


Рисунок 1. Показатели относительного использования синонимичных ГЦЗ-кодонов в мРНК, кодирующих алкогольдегидрогеназы классов 1B и 2 человека

На рис. 1 видно, что суммарные показатели RSCU соответствующих аргинину ГЦЗ-кодонов мРНК, кодирующих алкогольдегидрогеназы классов 1B и С человека, значительно отличаются (4,2 и 1,2, соответственно), а соответствующих цистеину – сходны (0,8 и 0,82).

Дистанция Дж. МакАйнерни ( $D_{jk}$ ) является наиболее новым и перспективным методом изучения стратегий кодирования белков. Вычисление дистанции МакАйнерни проводится по формуле:

$$D_{jk} = \sum_{i=1}^n \frac{abs(RSCU_{ji} - RSCU_{ki})}{n},$$

где  $n$  – число синонимичных вырожденных кодонов для определенной таблицы генетического кода,  $ji$  – кодон  $i$  последовательности  $j$ ,  $ki$  – кодон  $i$  последовательности  $k$  [7, 21].

Долгое время основным недостатком, ограничивающим широкое применение метода МакАйнерни, являлось отсутствие формулы для расчета ошибки одноименной дистанции, что не позволяло статически сравнивать сходство стратегий кодирования большого количества белков между собой. В 2007 году предложен способ вычисления ошибки дистанции МакАйнерни ( $SE_{D_{jk}}$ ), основанный на оригинальном вычислении варианты [5]:

$$SE_{D_{jk}} = SD_{D_{jk}} / \sqrt{n},$$

где  $n$  – число пар сравниваемых признаков,  $SD_{D_{jk}} = \sqrt{(\sum d^2)/n}$ ,  $d = V_B - M_B$ ,

$$V_B = abs(RSCU_{ji} - RSCU_{ki})/n.$$

Применение данного способа оказалось эффективным для определения сходства картин использования синонимичных кодонов в мРНК, кодирующих субъединицы 1–6, 4L НАДН-дегидрогеназы, цитохром  $b$ , субъединицы 1–3 цитохром- $c$ -оксидазы, субъединицу 6 АТФ-синтазы человека, трихинеллы и цианорабдитис (табл. 2 [7]).

Установлено, что достоверные отличия дистанций МакАйнерни характерны для девяти изученных мРНК, кодирующих (75,0±13,06%) белки человека и трихинеллы (по сравнению с таковыми для мРНК человека и цианорабдитис). Однако полученные дистанции для шести из них (для мРНК, соответствующих субъединице 1 НАДН-дегидрогеназы, цитохрому  $b$ , субъединицам 1-3 цитохром- $c$ -оксидазы, субъединице 6 АТФ-синтазы) меньше таковых контроля, а для трех из них (соответствующих субъединицам 2, 4, 6 НАДН-дегидрогеназы) выше таковых контроля. При коэволюции в системе «паразит–хозяин»



человека и трихинеллы на молекулярно-генетическом уровне следовало бы ожидать достоверно меньшие дистанции МакАйнерни опыта по сравнению с контролем для всех изученных мРНК.

Таблица 2. Дистанции МакАйнерни для мРНК, кодирующих митохондриальные белки человека и трихинеллы (опыт), а также человека и цианорабдитис (контроль)

Фермент/организм	Опыт	Контроль	p
НАДН-дегидрогеназа-1	0,010±0,0015	0,018±0,0020	<0,01
НАДН-дегидрогеназа-2	0,022±0,0016	0,016±0,0016	<0,05
НАДН-дегидрогеназа-3	0,013±0,0022	0,017±0,0021	>0,05
НАДН-дегидрогеназа-4	0,024±0,0017	0,017±0,0018	<0,05
НАДН-дегидрогеназа-4L	0,023±0,0023	0,020±0,0024	>0,05
НАДН-дегидрогеназа-5	0,023±0,0013	0,020±0,0021	>0,05
НАДН-дегидрогеназа-6	0,021±0,0021	0,015±0,0019	<0,05
Цитохром <i>b</i>	0,009±0,0013	0,018±0,0017	<0,001
Цитохром- <i>c</i> -оксидаза-1	0,008±0,0009	0,016±0,0016	<0,001
Цитохром- <i>c</i> -оксидаза-2	0,010±0,0014	0,017±0,0016	<0,01
Цитохром- <i>c</i> -оксидаза-3	0,010±0,0014	0,017±0,0016	<0,01
АТФ-синтаза 6	0,008±0,0013	0,018±0,0019	<0,001

Это наводит на мысль о том, что существуют некоторые особенности данной методики изучения картин использования кодонов. Для их определения был проведен сравнительный анализ показателей RSCU для кодонов некоторых изучаемых мРНК человека, трихинеллы и цианорабдитис [7]. Установлено, что особенностью метода МакАйнерни является то, что биологическая значимость результатов напрямую связана с числом кодонов нуклеиновой кислоты, по которым вычисляется дистанция. Эта особенность была рассмотрена авторами на примере кодонов УГУ и УГЦ (соответствуют цистеину) мРНК, кодирующих субъединицу 6 НАДН-дегидрогеназы человека, трихинеллы и цианорабдитис (табл. 1).

Анализируя значения показателя относительного использования синонимичных кодонов в мРНК, кодирующих субъединицу 6 НАДН-дегидрогеназы, можно прийти к выводу, что картины использования кодонов УГУ и УГЦ в данной мРНК человека сходны с таковыми в мРНК цианорабдитис и диаметрально противоположны таковым мРНК

трихинеллы. На этом примере видно, что данные значения RSCU получены по малому числу синонимичных кодонов (1–3) и, следовательно, обладают низкой репрезентативностью. Такие значения RSCU будут вносить существенный вклад в вычисляемую на их основании дистанцию МакАйнерни, анализ которой может привести к ошибочным выводам.

Очевидным путем устранения данного недостатка методики МакАйнерни является совместный анализ и вычисление дистанций максимально возможного (в пределах поставленной цели исследования) количества последовательностей нуклеиновых кислот. При учете установленной особенности дистанция МакАйнерни может быть рекомендована в качестве эффективного способа определения сходства картин использования синонимичных кодонов в последовательностях РНК и ДНК [7].

В таблице 2 видно, что значения RSCU по 12-ти изученным мРНК человека ближе к таковым трихинеллы, по сравнению с контролем. При этом общая дистанция МакАйнерни для 12-ти изученных мРНК человека и трихинеллы ( $0,010 \pm 0,0009$ ) в 1,5 раза меньше таковой для мРНК человека и цианорабдитис ( $0,015 \pm 0,0017$ ;  $p < 0,05$ ). Этот факт свидетельствует о статистически значимом сходстве картин использования синонимичных кодонов в мРНК, кодирующих исследуемые митохондриальные белки человека и трихинеллы [9].

Несмотря на достигнутые успехи в установлении основных закономерностей стратегий кодирования белков, разработка и апробация новых методов их изучения остается перспективным направлением развития биоинформатики, молекулярной эволюции и биологии. Это связано и с возможностями практического применения результатов исследований в области стратегий кодирования белков для проведения обратной трансляции последовательностей белков в соответствующие им последовательности нуклеиновых кислот (в случае первичного секвенирования аминокислотных последовательностей) и определения белок-кодирующих и белок-некодирующих участков геномов.

Работа выполнена в рамках гранта БРФФИ № Б08М-084 от 1.04.2008 г.

Литература:

1. Алкогольдегидрогеназы хордовых животных: монография / А.В. Бутвиловский, Е.В. Барковский, В.Э. Бутвиловский; под общ. ред. Е.В. Барковского. – Минск: БГМУ, 2007. – 144 с.
2. Барковский, Е.В. Мембраносвязанные аденилатциклазы: монография / Е.В. Барковский, О.В. Ачинович; под общ. ред. Е.В. Барковского. – Мн.: БГМУ, 2005. – 134 с.
3. Бутвиловский, А.В. Об использовании претерминальных кодонов и кодонов, содержащих гуанин и цитозин в нуклеотидных последовательностях мРНК алкогольдегидрогеназ человека 1-7 типов / А.В. Бутвиловский, Е.В. Барковский // Материалы международного симпозиума “Молекулярные механизмы регуляции функции клетки”. – Тюмень: Издательство “ВекторБук”, 2005. – С. 275-277.
4. Бутвиловский, А.В. Стратегия кодирования алкогольдегидрогеназ человека / А.В. Бутвиловский, Е.В. Барковский, В.Э. Бутвиловский // Вестник ВГМУ. - Витебск, 2006, №3. – С. 24–29.
5. Бутвиловский, А.В. Способ вычисления ошибки дистанции МакАйнерни и его практическое применение для решения прикладных задач / А.В. Бутвиловский, А.О. Одинцов // Материалы докладов XV Международной конференции студентов, аспирантов и молодых ученых «Ломоносов» / Отв. ред. И.А. Алешковский, П.Н. Костылев, А.И. Андреев. [Электронный ресурс] — М.: Издательство МГУ; СП МЫСЛЬ, 2008. — 1 электрон. опт. диск (CD-ROM); 12 см. - Систем. требования: ПК с процессором 486 +; Windows 95; дисковод CD-ROM; Adobe Acrobat Reader. [Адрес ресурса в сети интернет: <http://www.lomonosov-msu.ru/2008/>].
6. Изменения в процессе эволюции мутационного давления в последовательностях мРНК, кодирующих алкогольдегидрогеназы класса 3 хордовых животных / Бутвиловский А.В. [и др.] // Медицинский журнал. – Минск, 2007. – №1. – С. 22–25.
7. Особенности применения методики МакАйнерни для решения прикладных задач / В.Э. Бутвиловский, А.В. Бутвиловский, Е.В. Барковский, Ю.И. Линник, А.В. Давыдов // Техника и технологии: инновации и качество: материалы Междунар. науч.-практ. конф., 23-24 ноября 2007 г., Барановичи, Респ. Беларусь / редкол.: В.В. Таруц (гл. ред.) [и др.]. – Барановичи: РИО БарГУ, 2007. – С. 306-309.
8. О сходстве стратегий кодирования митохондриальных белков человека и трихинеллы. Часть 1. ГЦ-насыщенность, доля ГЦЗ-кодонов и частота использования претерминальных кодонов / В.Э. Бутвиловский, Ю.И. Линник, А.В. Бутвиловский, Е.В. Барковский // Медицинский журнал. – Минск, 2007. – №3. – С. 39–41.
9. О сходстве стратегий кодирования митохондриальных белков человека и трихинеллы. Часть 2. Картина использования синонимичных кодонов. Содержание аминокислотных групп GARP и FYMINK / В.Э. Бутвиловский, Е.В. Барковский, А.В. Бутвиловский, Ю.И. Линник // Медицинский журнал. – Минск, 2007. – №4. – С. 39–42.
10. Способ сравнения частот использования претерминальных кодонов в мРНК, транскрибуемых в соответствии с разными таблицами генетического кода // В.Э. Бутвиловский, А.В. Бутвиловский, Е.А. Черноус, Н.С. Климович // От классических методов генетики и селекции к ДНК-технологиям (к 95-летию со дня рождения академика Н.В. Турбина). Международная научная конференция, Гомель, 2-5 октября 2007 г.: Материалы конференции / ред. колл.: А.В. Кильчевский и др.: Институт генетики и цитологии НАН Беларуси. – Минск, Право и экономика, 2007. – С. 148.
11. Справочник терминов молекулярной эволюции и филогенетики. Учебно-методическое пособие / В.Э. Бутвиловский [и др.] // Мн.: БГМУ, 2006. – 40 с.
12. Хрусталева, В.В. Частота использования претерминальных кодонов в матричных РНК бактериальных аденилатциклаз / В.В. Хрусталева, Е.В. Барковский // Здоровоохранение. – 2006. – №2. – С. 17–20.

13. Codon catalog usage and the genome hypothesis / R. Grantham [et al]. // Nucl. Acids Res. – 1980. – Vol. 8. – P. 49–62.
14. Codon usage between genomes is constrained by genome-wide mutational process / S.L. Chen [et. al] // Proc. Natl. Acad. Sci. – 2004. – V. 101. – P. 3480–3485.
15. Codon usage patterns in *Escherichia coli*, *Bacillus subtilis*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Drosophila melanogaster* and *Homo sapiens*; a review of the considerable within-species diversity / P.M. Sharp [et. al]. // Nucl. Acids Res. – 1988. – V.16. – P. 8207–8211.
16. Cutter, A.D. The evolution of biased codon and amino acid usage in nematode genomes / A.D. Cutter, J.D. Wasmuth, M.L. Blaxter // Mol. Biol. Evol. – 2006. – V. 23. – P. 2302–2315.
17. Eyre-Walker, A. Synonymous codon bias is related to gene length in *Escherichia coli*: selection for translational accuracy? / A. Eyre-Walker // Mol. Biol. Evol. – 1996. – Vol. 13. – P. 864–872.
18. Hasegawa, M. Secondary structure of MS2 phage RNA and bias in code word usage / M. Hasegawa, T. Yasunaga, T. Miyata // Nucl. Acids Res. – 1979. – V. 7. – P. 2073–2079.
19. Ikemura, T. Codon usage and tRNA content in unicellular and multicellular organisms / T. Ikemura // Mol. Biol. Evol. – 1985. – V. 2. – P. 13–34.
20. Lobry, J.R. Hydrophobicity, expressivity and aromaticity are the major trends of amino-acid usage in 999 *Escherichia coli* chromosome-encoded genes / J.R. Lobry, C. Gautier // Nucl. Acids Res. – 1994. – V. 22. – P. 3174–3180.
21. McInerney, J.O. GCUA: General codon usage analysis / J.O. McInerney // Bioinformatics. – 1998. – Vol. 14(4). – P.372-373.
22. Rocha, E.P.C. Codon usage bias from tRNA's point of view: redundancy, specialization and efficient decoding for translation optimization / E.P.C. Rocha // Genome Res. – 2004. – V. 14. – P. 2279–2286.