



УДК 577.217.525

ЗАВИСИМОСТЬ УРОВНЯ ЭКСПРЕССИИ ГЕНА В *E. coli* ОТ СТРУКТУРЫ УЧАСТКА ИНИЦИАЦИИ ТРАНСЛЯЦИИ (TIR)

II*. ВТОРИЧНАЯ СТРУКТУРА TIR

© 1995 г. А. И. Гуревич**, Р. С. Есипов, Т. А. Качалина, А. Л. Каюшин

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

Поступила в редакцию 16.03.94 г.

Уровни экспрессии некоторых генов, транскрибирующихся в мРНК с идентичными лидерными последовательностями и даже с идентичными протяженными кодирующими участками генов, могут сильно различаться. С целью установить закономерности такого явления изучены вторичные структуры ряда мРНК, транскрибируемых с серии экспрессионных плазмид. Показано, что влияние вторичной структуры мРНК в районе инициации трансляции (TIR) на эффективность инициации не ограничено возможностью образования шпилек в этом районе; существенную роль играют дальние взаимодействия. При возможности образования комплементарных структур более прочных, чем при взаимодействии участков SD, UB1, UB2 и DB с 16S рРНК, эффективность инициации трансляции и соответственно уровень экспрессии снижается.

Ключевые слова: мРНК, рРНК, участок инициации трансляции, вторичная структура РНК.

Зависимость уровня экспрессии гетерологичных генов в *E. coli* от структуры транскриптов имеет сложный характер. Высокий уровень экспрессии связан с устойчивостью мРНК, которая может быть стабилизирована при наличии шпилек на 5'- и 3'-концах [2 - 4], а также с эффективностью рибосомсвязывающего участка, которая в свою очередь определяется как первичной, так и вторичной структурой этого района мРНК (см. обзор [5]). В предыдущей статье мы рассмотрели значение первичной структуры участка инициации трансляции (TIR) и показали вероятный механизм его взаимодействия с 16S рРНК в инициаторном комплексе [1]. Продолжая изучение структурных особенностей экспрессируемых генов, мы нашли, что уровни экспрессии некоторых генов, транскрибирующихся в мРНК с идентичными лидерными последовательностями, могут сильно различаться. Чтобы установить закономерности такого явления, мы сравнили уровень

экспрессии нескольких генов гетерологичных белков, включив в число исследуемых объектов кроме рассмотренных нами ранее человеческого интерлейкина-3 (hIL3) и человеческого эпидермального фактора роста (hEGF) [1] человеческий интерлейкин-4 (hIL4), гибрид N-концевой последовательности hIL3 и тримера лизин-10-окситоцина (lLOX3) [6] и гибрид человеческого интерлейкина-3 и фактора роста GM-CSF (lL3GMCSF).

На схеме приведено конструирование экспрессионных плазмид с генами этих белков. Ген *hIL4* (7) был вырезан из плазмиды pKil4-9 [7] с помощью рестриктаз *Pst*I и *Hind*III и клонирован в *EcoRV/Hind*III-векторе (8), полученном из плазмиды pTE4 (5) [1] с помощью линкера (A), содержащего N-концевой участок гена *hIL4*.

ATCATATGCATAAATGCGACATCACCCCTGCA
(A) TAGTATACGTATTTACGCTGTAGTGGG

Для получения гена гибрида hIL3-hGMCSF сначала была получена плаزمида hTE2il3g (1f), аналогичная pTE2il3 (1b) [8], но в которой искусственный ген *hIL3* был лишен терминирующего кодона. В результате ген *il3g* кодировал синтез несколько большего белка (12 дополнительных аминокислотных остатков). Этот ген (без дополнительной последовательности) был вырезан вместе с регуляторным участком (2) по сайтам *Bam*HI/*Eco*RI. С другой стороны, искусственный ген GM-CSF (3) был вырезан из плазмиды pUC118gmcsf, предоставленной В.Г. Коробко, с

Сокращения: TIR, mTIR и MTIR – участок инициации трансляции, его нетранслируемая и транслируемая ветви; SD – последовательность Шайна-Дальгарно; ASD – anti-SD, 3'-концевая последовательность 16S рРНК; UB – upstream box, часть mTIR, комплементарная участку AUB – anti-upstream box в 16S рРНК; DB – downstream box, часть MTIR, комплементарная участку ADB – anti-downstream box в 16S рРНК; TE – translational enhancer, усилитель трансляции; RBS – сайт связывания рибосомы; н. – нуклеотидное звено.

* Предыдущее сообщение см. [1].

** Автор для переписки.

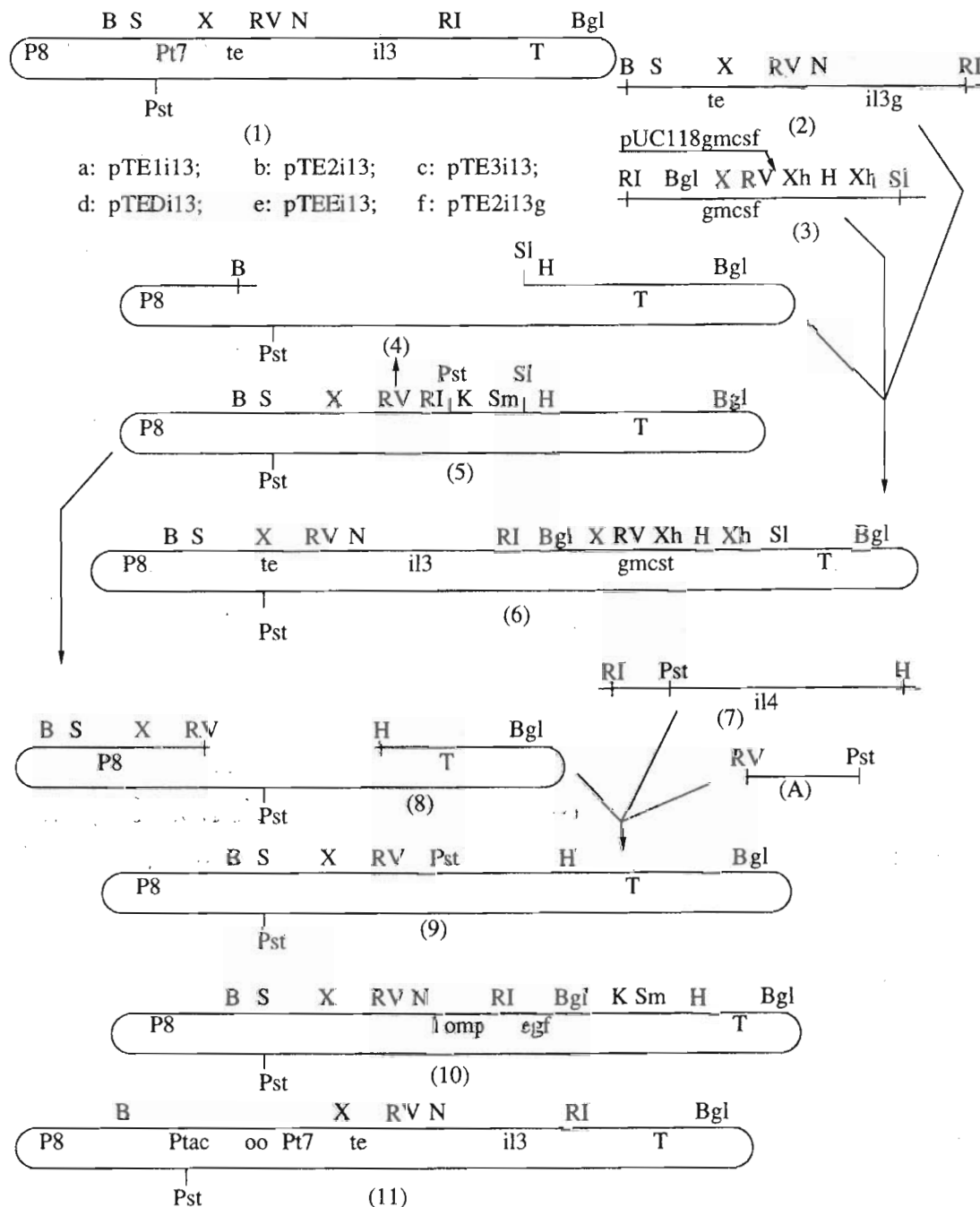


Схема. Получение рекомбинантных плазмид с генами *il3*, *il4*, *il3gmcfsf* и *lompegf*. Указаны сайты рестриктаз (B – *Bam*HI, *Bgl* – *Bgl*II, RI – *Eco*RI, RV – *Eco*RV, H – *Hind*III, K – *Kpn*I, N – *Nde*I, Pst – *Pst*I, S – *Sph*I, Sl – *Sal*I, Sm – *Sma*I, X – *Xba*I, Xh – *Xho*I), участки промоторов (P8, Ptac, Pt7), терминатора, усилителей трансляции (te). А – линкер, содержащий N-концевой участок гена *il4*.

помощью рестриктаз *Eco*RI и *Sal*I. Наконец, оба гена мы лигировали с *Bam*HI/*Sal*I-вектором (4), полученным из плазмиды pTE3 (5).

Таким образом, приведенный выше ряд экспрессионных плазмид обеспечивал конститутивный синтез гетерологичных белков в результате трансляции мРНК (инициированных с промотора P8), имеющих идентичные протяженные

5'-концевые нетранслируемые последовательности и протяженные 3'-концевые участки, оканчивающиеся стабилизирующей мРНК шпилькой р-независимого терминатора (рис. 1, I). В большинстве случаев идентичными также были структуры не только лидерного участка мРНК, но и протяженного участка в кодирующей области (в случае гена *il3* и гибридных генов *ilox3* и

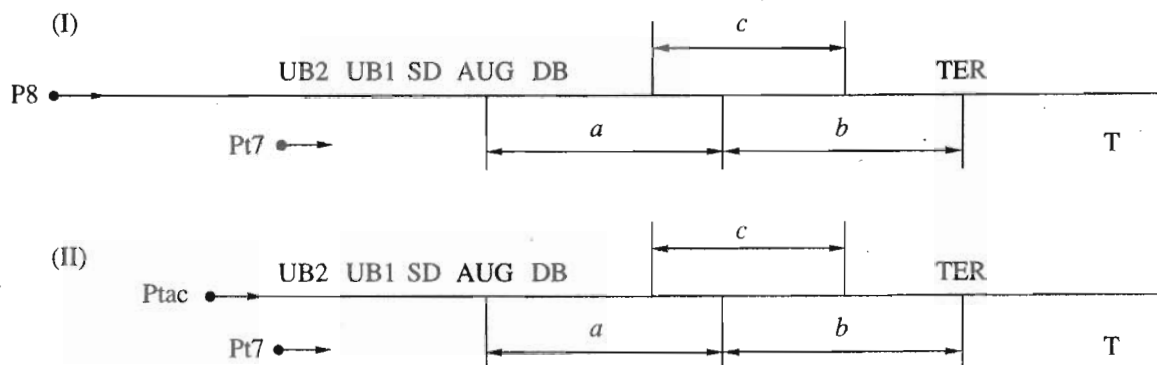


Рис. 1. Строение мРНК генов, экспрессия которых изучена в условиях конститутивного синтеза с промотора P8 (I) и индуцированного синтеза с промоторов Ptac (II) и Pt7 (I и II). Указано положение стартов транскрипции с соответствующих промоторов (•→), расположение в мРНК участков взаимодействия с рибосомой в инициаторном комплексе (UB2, UB1, SD, DB), иницирующего (AUG) и терминирующего (TER) кодонов, терминатора транскрипции (T). *a* – проксимальный участок структурного гена, идентичный в генах *il3*, *ilox3*, *ilgmcsf*; *b* – дистальный участок структурного гена, способный взаимодействовать с лидерным участком мРНК; *c* – центральная часть мРНК, образующая самостоятельный домен. Протяженность лидерного участка мРНК до иницирующего кодона от старта транскрипции с P8 – 296, с Ptac – 158, с Pt7 – 65 н.

ilgmcsf эти участки на рис. 1 обозначены *a*). Несмотря на это, уровни экспрессии изученных генов сильно различались (табл. 1), и если в случае плазмиды *r8il3* низкий уровень биосинтеза белка hIL3 по сравнению с плазмидой *rte2il3* можно было объяснить отсутствием в структуре мРНК участка TE, то для плазмид *rteil4* и в особенности *rteilox3* и *rteilgmcsf*, очевидно, низкий уровень синтеза белка был обусловлен строением удален-

ного от TIR участка *b* в соответствующих мРНК (см. рис. 1).

Другой ряд экспрессионных плазмид с теми же генами, но обеспечивавших индуцированный синтез укороченных с 5'-конца мРНК, инициированных с промотора Ptac либо Pt7 (см. рис. 1, II), был получен путем рекомбинации с плазмидой *rTOTE2il3* (11) [9] по сайтам рестриктаз *XbaI* (либо *EcoRV*) и *PstI* (либо *BglII*) (см. схему).

Гены *il4*, *ilox3* и *ilgmcsf*, транскрибируемые в мРНК с протяженной лидерной последовательностью (с промотора P8), давали низкий уровень экспрессии (см. рис. 1, I, и табл. 1). Однако те же гены, транскрибируемые в мРНК с укороченной с 5'-конца лидерной последовательностью (с промоторов Ptac или Pt7), давали высокий уровень экспрессии (см. рис. 1, I, II, и табл. 2). Очевидно, что в последних случаях исчезновение ингибирующего влияния структуры участка *b* (см. рис. 1) обусловлено исключением дальнего взаимодействия с лидерным участком мРНК.

В связи с этим мы рассмотрели вторичные структуры всех трех типов исследуемых мРНК, инициированных с трех различных промоторов (P8, Ptac или Pt7). Модели вторичных структур мРНК были рассчитаны по программе "Вторичная структура РНК" пакета программ "Genebee" (Л.И. Бродский, МНПП "Гендальф" и Институт ФХБ им. Белозерского, МГУ).

Из рассмотрения моделей следует, что 3'-концевой домен всех изученных мРНК, охватывающий около 90 н., характеризуется наличием шпильки терминатора и еще одного структурированного участка (см. рис. 2А).

5'-Концевые домены всех мРНК, инициированных с промоторов Ptac (охватывающий около 70 н.) и Pt7 (около 20 н.), характеризуются наличием

Таблица 1. Конститутивный биосинтез белков на мРНК, инициированных с промотора P8

Экспрессионная плаزمид	Ген	Содержание (%) белкового продукта в тотальном лизате штамма-производителя <i>E. coli</i>		
		HB101	SG20050	TG1
<i>r8il3</i>	<i>il3</i>	<1	<1	<1
<i>rte1il3</i>	»	2.8	3.4	3.4
<i>rte2il3</i>	»	24.7	27.9	12.0
<i>rte3il3</i>	»	7.0	8.0	4.6
<i>rte4il3</i>	»	5.7	7.9	3.8
<i>rte5il3</i>	»	15.9	16.2	3.8
<i>rte2il3g</i>	<i>il3g</i>	12.0	17.3	–
<i>rteil4</i>	<i>il4</i>	<1	<1	–
<i>rte1ompef</i>	<i>lompef</i>	3.7	2.4	2.8
<i>rte2lompef</i>	»	4.2	3.1	2.4
<i>rte3lompef</i>	»	4.4	2.9	2.4
<i>rte1ilgmcsf</i>	<i>ilgmcsf</i>	<1	<1	<1
<i>rte1ilox3</i>	<i>ilox3</i>	<1	<1	<1

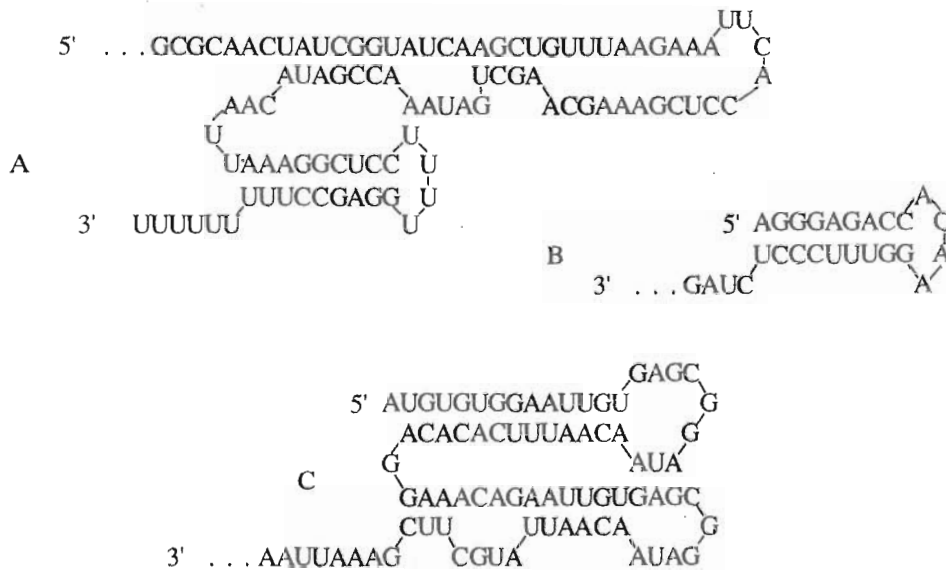


Рис. 2. Вторичная структура концевых доменов мРНК. А – 3'-концевой домен, В – 5'-концевой домен мРНК, инициированных с промотора Pt7, С – 5'-концевой домен мРНК, инициированных с промотора Ptac.

устойчивой вторичной структуры со свободной энергией шпильки $\Delta G^0 = -16 \dots -18$ ккал/моль (см. рис. 2В, С), которые, очевидно, должны придавать устойчивость мРНК (ср. работы [2, 3, 5]). С другой стороны, 5'-концевые последовательности всех мРНК, инициированных с промотора P8 на протяжении 100 н., не были вовлечены в образование прочных вторичных структур. В то же время последующий участок во многих из рассматриваемых мРНК оказался включенным в образование вторичной структуры с участием транслируемой области; в случае мРНК P8-*te2il3*, а также P8-*te1il3*, P8-*te3il3*, P8-*il3*, P8-*tedil3*, P8-*teeil3*, P8-*te2il3g* вторичная структура охватывала нуклеотиды 105 - 185 (нетранслируемая часть) и 580 - 680 (С-концевой транслируемый участок гена *hll3*) (рис. 3А).

При этом остающуюся центральную часть мРНК можно разделить на два домена: 1) включающий TIR, 2) остальная часть структурного гена (участок *c* на рис. 1). Подобного типа структурную организацию имеет и мРНК P8-*teil4*, в которой вторичной структурой связаны участки 134 - 202 (нетранслируемая часть) и 749 - 809 (участок после терминирующего кодона) (см. рис. 3В), а в центральной части также имеется домен, охватывающий всю остальную транслируемую часть гена.

В то же время, если сравнивать инициированные с разных промоторов (P8, Ptac, Pt7) мРНК описанных выше генов, можно прийти к следующим представлениям о включении во вторичные структуры областей TIR и прилегающих участков (см. рис. 4).

При рассмотрении моделей мы учитывали, что в шпильке 442 - 492 16S рРНК в основании распо-

ложен прочный стебель (5G · C-пар), а в остальной части, включающей последовательности AUB1 и AUB2 (456 - 487), свободная энергия самокомплементарных участков мала ($\Delta G^0 > -5$ ккал/моль). В шпильке 1410 - 1540 участок, включающий последовательность ADB, вовлечен во вторичную структуру с сравнительно низкой свободной энергией ($\Delta G^0 = -6 \dots -7$ ккал/моль).

Между структурами TIR мРНК *te2il3* и *toteil3* не наблюдается существенной разницы; в то же время в TIR в *teilo3* участок DB включен в значительно более прочную вторичную структуру ($\Delta G^0 -13.9$ ккал/моль), чем в случае TIR *toteilo3* ($\Delta G^0 -10.2$ ккал/моль). Аналогично различаются между собой TIR *teilgmcsf* и *toteilgmcsf*.

Таблица 2. Индуцированный биосинтез белков на мРНК, инициированных с промоторов Ptac и Pt7

Экспрессионная плаزمид	Ген	Содержание белкового продукта в тотальном лизате штамма-производителя <i>E. coli</i>		
		TG1	AG10	BL21 (DE3)
pkkil4-12	<i>il4</i>	15.0	1.5	2.0
ptote2il3	<i>il3</i>	20.0	3.8	3.2
ptoteilgmcsf	<i>ilgmcsf</i>	4.7	7.8	18.6
ptoteiliox3	<i>ilox3</i>	26.6	13.0	18.0
pteiliox3	<i>ilox3</i>	-	-	9.2
pteilgmcsf	<i>ilgmcsf</i>	-	-	12.4

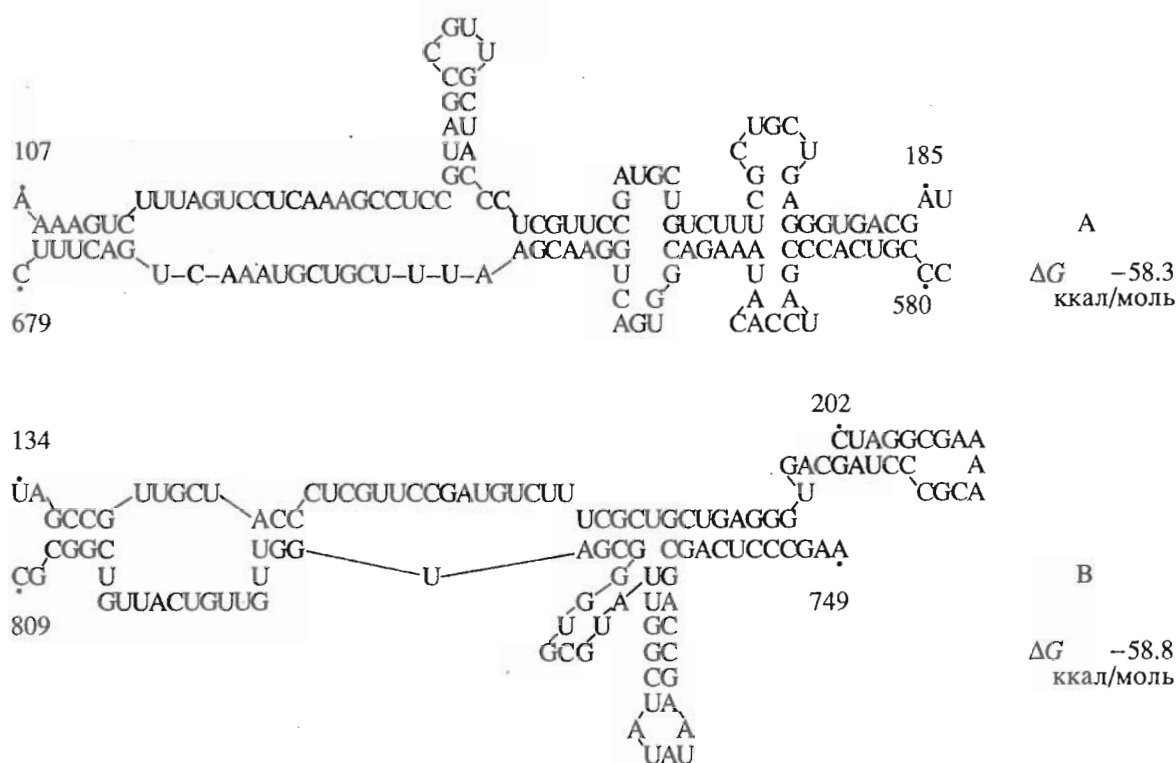


Рис. 3. Вторичные структуры участков мРНК генов *il3* (А) и *il4* (В), инициированных с промотора P8. Указаны номера нуклеотидов в мРНК.

В предыдущем сообщении [1] уже отмечались различия в первичной структуре TIR генов *te1il3*, *te2il3*, *te3il3*, *tedil3* и *teeil3*, а также *telompf*, *tedlompf* и *teelompf*. Однако эти различия принципиально не изменяют характер взаимодействий в структуре TIR, хотя можно отметить, что в мРНК *te1il3* наблюдается лишь слабое взаимодействие в участке UB2 ($\Delta G^0 < -4$ ккал/моль), а в мРНК *te3il3* внутримолекулярное взаимодействие с участием UB1 слабее, чем в случае *te2il3*.

Следует отметить отсутствие в мРНК генов *il4* и *il3* в плаزمидях *rckil4-9* и *r8il3* участков UB1 и UB2. Что же касается мРНК *teil4*, то TIR в этом случае в отличие от TIR *te2il3* имеет участок DB, включенный во вторичную структуру.

Чтобы выяснить, каким же образом сказывается вторичная структура TIR на уровне экспрессии генов, мы сравнили уровень биосинтеза упомянутых выше белков в различных штаммах-продукентах в стандартных условиях конститутивного (с промотора P8) и индуцированного (с промоторов Ptac и Pt7) синтеза. Данные анализа тотальных лизатов клеток продукентов после гель-электрофореза и сканирования гелей, прокрашенных ку-масси R-250, приведены в табл. 1 и 2.

Из полученных результатов следует, что в тех случаях, когда сайты связывания TIR с 16S рРНК (UB1, UB2, SD, DB) не участвуют в образовании вторичной структуры мРНК (см. рис. 4, 2, 7, 8),

достигается максимальный уровень экспрессии генов. При одинаковой первичной структуре сайтов уровень экспрессии снижается при включении их во вторичную структуру; при этом, если свободная энергия возникающих шпилек не ниже порогового значения $\Delta G^0 = -9 \dots -10$ ккал/моль, наблюдаемое снижение невелико. Однако уровень экспрессии резко падает, если свободная энергия образования шпилек уменьшается. Это может быть следствием выведения соответствующих участков из взаимодействия с 16S рРНК (ср. рис. 4, 1 и 2, 4 и 5, 2 и 6).

Таким образом, кроме первичной структуры TIR, определяющей структуру сайтов взаимодействия мРНК с 16S рРНК в инициаторном комплексе, на эффективность инициации трансляции и соответственно на уровень экспрессии генов существенное влияние может оказывать вторичная структура мРНК, включающая комплементарные взаимодействия между элементами TIR и удаленными от TIR районами мРНК.

ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

Получение плазмид *rFPCP10il3*, *rTE1il3*, *rTE2il3*, *rTE3il3* описано в работе [7], *rTEDil3*, *TEEil3*, *rTelompf*, *rTEDlompf*, *rTEElompf* – в предыдущем сообщении [1], плазмид *rTE2ilox3* и

1) pte1il3

GGGUGACGAUCCCGCAAAGCGGAUCCGCAUGCUCUAGAAAAUAAUUUUGUUUAACUUUA
 UB2 _____ UB1 _____
 UCGCC -11.1

SD _____ DB _____
 AGAAGGAGAUUCAUAUAGGCUCCGAUGACCCAGACCACCUCUCUGAAAACCCUCUUGGGUU
 → ← -8.8

2) pte2il3

GAAAUAAUACGACUCACUAUAGGGAGACCACAACGGUUUCCCUAGAAAAUAAUUUUGU
 UB2 _____ UB1 _____
 → ← -10.3 ← -5.2 ←

UB1 _____ SD _____ DB _____
 UUAACUUUAAGAAGGAGAUUCAUAUAGGCUCCGAUGACCCAGACCACCUCUCUGAAAACCU
 → ← -8.8 ← → ← -7.1

3) p8il3

AAAAGCGGAUCCAAGGAGAUUCAUAUAGGCUCCGAUGACCCAGACCACCUCUCUG
 SD _____ DB _____
 -11.1 UCGCC → ← -8.8 ← → GUGU -7.1

4) ptoteil3, ptoteilo3 и ptoteilmcsf

GAAAUAAUACGACUCACUAUAGGGAGACCACAACGGUUUCCCUAGAAAAUAAUUUUGU
 UB2 _____ UB1 _____
 -10.8 UCCCU CAAAG -8.1

UB1 _____ SD _____ DB _____
 UUAACUUUAAGAAGGAGAUUCAUAUAGGCUCCGAUGACCCAGACCACCUCUCUGAAAACCU
 → ← -8.8 ← → UGGU -10.2 ← -7.1

5) pteilo3 и pteilmcsf

GAAAUAAUACGACUCACUAUAGGGAGACCACAACGGUUUCCCUAGAAAAUAAUUUUGU
 UB2 _____ UB1 _____
 -10.8 → ←

UB1 _____ SD _____ DB _____
 UUAACUUUAAGAAGGAGAUUCAUAUAGGCUCCGAUGACCCAGACCACCUCUCUGAAAACCU
 → ← -8.8 ← → ACUGGG → ← -7.1
 -13.9

6) pteil4

GAAAUAAUACGACUCACUAUAGGGAGACCACAACGGUUUCCCUAGAAAAUAAUUUUGU
 UB2 _____ UB1 _____
 -10.8 → ←

UB1 _____ SD _____ DB _____
 UUAACUUUAAGAAGGAGAUUCAUAUAGCAUAAAUGCGACAUCACCCUGCAGGAAAUCAUAAA
 → ← -9.4 UACGC AGUGG -9.1

7) pkil4-9

AAUUGUGAGCGGAUACAUAUUUCACACAGGAAACAGAAUUAUGCAUAAAUGCGACAUCACCCU
 UB1 _____ SD _____ DB _____

8) ptelompf

GAAAUAAUACGACUCACUAUAGGGAGACCACAACGGUUUCCCUAGAAAAUAAUUUUGU
 UB2 _____ UB1 _____
 -10.8 → ←

UB1 _____ SD _____ DB _____
 UUAACUUUAAGAAGGAGAUUCAUAUGAUGAAGCGCAAUAUUCUGGCAGUGAUCGUCCUGC
 → ← -8.2

Рис. 4. Вторичные структуры TIR мРНК указанных плазмид. Сверху отчеркнуты участки связывания с рибосомой в инициаторном комплексе. Стрелками отмечены шпильки в пределах TIR, снизу приведены комплементарные последовательности шпилек за пределами TIR. Приведены также свободные энергии шпилек.

pTOTE2il3 – в работах [4] и [8] соответственно. Структура плазмид изображена на схеме.

Использованные реагенты и условия экспериментов описаны в сообщении [1]. Денситометрическое сканирование покрашенных кумасси R-250 гелей проводили на хроматосканере CS-930 (Shimadzu).

Использованы штаммы *E. coli*: HB101 [F^- , hsdS (r_B^- , m_B^-), *recA*, *ara-14*, *proA*, *leu*, *thi*, *lacY*, *galK2*, *str*, *xyl-5*, *mtl-1*, *supE44*]; SG20050 [*lon*, *lac*, Tc^R]; TG1 [$\Delta(lac-pro)$, *thi*, *supE*, *hsd-5/F'*, *traD36*, *proA*⁺*B*⁺, *lacI*^q, *lacZ* Δ M15]; AG10 [*lon*, *lac*, Tc^R/F' , *traD36*, *proA*⁺*B*⁺, *lacI*^q, *lacZ* Δ M15]; BL21(DE3) [F^- , *ompT*, r_B^- , m_B^-]. Штамм AG10 получен конъюгацией штаммов TG1 и SG20050.

СПИСОК ЛИТЕРАТУРЫ

1. Гуревич А.И., Есипов Р.С., Качалина Т.А., Каюшин А.Л., Коростелева М.Д. // Биоорганич. химия. 1995. Т. 21. № 2. С. 117 - 123.
2. Emory S.A., Bouvet P., Belasco J.G. // Genes and Development. 1992. V. 6. № 1. P. 135 - 148.
3. Newbury S., Smith N.H., Robinson E.C., Hiles I.D., Higgins C.F. // Cell. 1987. V. 48. № 2. P. 297 - 310.
4. Chen C.-Y.A., Beatty J.T., Cohen S.N., Belasco J.G. // Cell. 1988. V. 52. № 4. P. 609 - 619.
5. McCarthy E.G., Gualerzi C. // Trends Genet. 1990. V. 6. № 3. P. 78 - 85.
6. Гуревич А.И., Качалина Т.А., Каюшин А.Л., Коростелева М.Д., Мирошников А.И. // Биоорганич. химия. 1993. Т. 19. № 6. С. 629 - 632.
7. Батчикова Н.В., Кулагина М.А., Луценко С.В., Смирнов В.А., Каневский В.Ю., Рязанова Л.А., Назимов И.В., Сонина Н.В., Синягина Е.А., Ажаев А.Л. // Биоорганич. химия. 1992. Т. 18. № 5. С. 660 - 669.
8. Гуревич А.И., Скапцова Р.В., Луценко С.В., Смирнов В.А., Куркин А.Н., Ажаев А.В. // Биоорганич. химия. 1991. Т. 17. № 5. С. 647 - 652.
9. Луценко С.В., Гуревич А.И., Птицын Л.Р., Рязанова Л.А., Смирнов В.А. // Биоорганич. химия. 1992. Т. 18. № 3. С. 391 - 397.

Dependence of the Gene Expression Level in *E. coli* on the Structure of the Translation Initiation Region. II. Secondary Structure of the Translation Initiation Region¹

A. I. Gurevich², R. S. Esipov, T. A. Kachalina, and A. L. Kayushin

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

Abstract – The expression levels of genes that are transcribed to give mRNAs with identical leader sequences and even with identical extended coding regions may differ considerably. In order to determine the mechanism of this phenomenon, secondary structures of some mRNAs synthesized from a series of expression plasmids were studied. It was shown that the effect of the mRNA secondary structure in the translation initiation region on the initiation efficiency is due not only to the hairpin formation in this region but also to long-range interactions. When complementary structures tighter than those resulted from the interaction of regions SD, UB1, UB2, and DB with 16S rRNA are formed, the efficiency of the translation initiation and, consequently, the expression level decrease.

Key words: mRNA, rRNA, translation initiation region, RNA secondary structure.

¹ For previous communication see [1].

² To whom correspondence should be addressed.