

# **Ekspresja genów heterologicznych w *Escherichia coli* na modelowo ujętym przykładzie interferonu gamma**

*Tomasz Cieślukowski*  
Centrum Mikrobiologii i Wirusologii  
Polska Akademia Nauk  
Łódź

## **1. Wprowadzenie**

**E***scherichia coli* obok *Bacillus subtilis* i *Saccharomyces caerevisiae* pozostaje najczęściej stosowanym mikroorganizmem wykorzystywanym do ekspresji heterologicznych genów. O użyteczności *E. coli* w procesach uzyskiwania obcych białek decyduje nie tylko dobra znajomość genomu, ale i fizjologii tego gatunku bakterii (1). Istotna jest również łatwość hodowli, możliwość uzyskiwania dużych ilości biomasy na prostych podłożach, a także względnie proste sposoby izolacji wytworzonych białek.

Można przyjąć, że niemal każde białko, które nie jest zbyt małe tzn. zawiera więcej niż 80 aminokwasów (aa), ani zbyt duże (500aa), zbyt hydrofobowe, lub ze zbyt dużą liczbą reszt cysteinowych, może podlegać ekspresji w *E. coli* (2). Być może teza ta wymaga dodatkowego zastrzeżenia dotyczącego sekwencji N-końca produktu białkowego, który to koniec warunkuje utrzymanie się białka w cytozolu (3,4).

## **2. Czynniki określające wydajność ekspresji**

Konstruowane i stosowane układy ekspresji heterologicznych genów obejmują wszystkie etapy powstawania produktu białkowego, włącznie z obecnością czynników decydujących o efektywnej transkrypcji, jak i translacji, oraz modyfikacjami postranslacyjnymi. Odnosi się to zwłaszcza do szybkości transkrypcji oraz stabilności transkryptów. Stabilność mRNA zależy od obecności struktur drugorzędowych tak na końcu 5', jak i w sekwencji terminacyjnej, a także od czynników białkowych odpowiedzialnych za degradację informacyjnego RNA. Wydajność procesu transkrypcji natomiast zależy od wybranego promotora, sposobu jego ewentualnej indukcji, sposobu terminacji, a także od stopnia powiązania z replikacją plazmidu niosącego gen. Mogą

na niego wplywać równieć takie czynniki jak pierwszo- i drugorzędowa struktura informacyjnego RNA, które decydują o kinetyce procesu wiązania podjednostki rybosomu i o inicjacji translacji (5).

Kontrola samego procesu translacji (tzn. elongacji translacji) może odbywać się nie tylko przez modyfikację struktury drugorzędowej transkryptu poza obszarem inicjacji, ale równieć przez dobór kodonów — tzw. kodonów optymalnych (6,7,8,9). Stanowi to niemal zasadę konstruowania układu ekspresji heterologicznego genu, w przypadku gdy gen ten uzyskujemy na drodze chemicznej syntezy. Jednak dobór taki nie stanowi najważniejszego z czynników decydujących o efektywnej wartości ekspresji rozumianej jako wypadkowa procesów syntezy i degradacji białka (10) (por. tab. 5). O podatności produktu na działanie komórkowych procesów degradacji przesądza często struktura samej sekwencji kodującej. Dlatego projektowanie i optymalizacja układu ekspresji odnoszą się równieć do procesu degradacji informacyjnego RNA.

Ponadto stosuje się szereg rozwiązań chroniących uzyskany już produkt przed proteolizą. Są to na przykład: kilkakrotne powtórzenie genu powodujące syntezę wielu kopii tego samego białka w postaci jednego łańcucha polipeptydowego (11), połączenie w jeden produkt sekwencji aminokwasowych dwóch różnych białek, czy wreszcie takie zaprojektowanie genu, że bezpośredni produkt ekspresji syntetyzowany jest w formie prekursora podlegającego sekrecji.

Ludzki interferon gamma jest jednym z tych białek, które od dawna wytwarza się w komórkach *E. coli* w bardzo zróżnicowanych układach ekspresyjnych. Długa, prawie piętnastoletnia, historia rekombinowanego interferonu dostarczyła tu szerokiego spektrum przykładów. Właśnie dzięki temu interferon gamma może służyć jako dobra ilustracja stosowanych sposobów heterologicznej ekspresji.

W dalszej części artykułu na przykładzie IFN $\gamma$  przedstawiono uwarunkowania procesu i poziomu ekspresji wynikające ze struktury operonu, kolejno na poziomie transkrypcji, translacji jak też modyfikacji posttranslacyjnych.

### 3. Regulacja ekspresji na etapie transkrypcji

Regulację ekspresji genu heterologicznego w *E. coli* na etapie transkrypcji można rozpatrywać z czterech punktów widzenia: wyboru promotora, sposobu terminacji, wyboru plazmidu (szybkość jego replikacji), oraz stabilności transkryptu. Interferon gamma stanowi tu dobrą egzemplifikację stosowanych rozwiązań, a to z tej przyczyny, że jego transkrypcję uzyskiwano z różnych promotorów, jak np. *lacUV5*, *trp*,  $\lambda$ *PL*,  $\lambda$ *PR*, *tac* i innych.

Po raz pierwszy rekombinowany interferon gamma uzyskano w zespole Goeddela i wsp. (12). W komórkach *E. coli* szczepu W3110 plazmid zawierający promotor *trp* pozwalał na ekspresję do 250 u IFN $\gamma$ /ml ekstraktu komórkowego (mierzoną zdolnością wiązania przeciwciał). Niemal w tym samym czasie w komórkach WA801, z pochodnej plazmidu pBR322 zawierającej pro-



motor *lacUV5* uzyskano 486-850 u/ml (13). Zastosowanie promotora  $\beta$ -lakta-mazy w pochodnej pBR322 w komórkach LE392 doprowadziło do wydajności wytwarzania interferonu o jeden rząd większej (około 2560 u/ml) (14). Ostatni plazmid zawierał zmodyfikowaną sekwencję Shine-Dalgarno (SD), w jej postaci optymalnej w sensie zdolności wiązania rybosomu tzn. AGGAGG. Dopiero wprowadzenie silnych promotorów *T5P25* i *tac* pozwoliło na uzyskanie preparatycznych ilości — określonych wielkością efektu cytopatycznego — rzędu odpowiednio  $1,2 \times 10^5$  u/ml (15) i  $80 \times 10^6$  u/ml (16) (tab.1).

TABELA 1

ZESTAWIENIE UKŁADÓW EKSPRESYJNYCH, W KTÓRYCH WYTWARZANO LUDZKIE BIAŁKO IFN $\gamma$  W KOMÓRKACH *E. coli*, UZYSKANA WYDAJNOŚĆ, METODY JEJ POMIARU

Plazmid	Promotor	Szczep <i>E. coli</i>	Wydajność ekspresji	Sposób pomiaru aktywności IFN $\gamma$	Piśmien-nictwo
pIFN $\gamma$ trp48	trp	W3110	250 u/ml	zdolność wiązania przeciwciał	(2)
pGIF5	lacUV5	WA802	486 u/ml	efekt cytopatyczny: wirus Sindbis/amniocyty	(13)
pJP <sub>1</sub> R <sub>3</sub> -IFN $\gamma$	T5P25	LE392	$80 \times 10^6$ u/ml	efekt cytopatyczny: wirus encefalomyocarditis/komórki linii raka płuc A549	(14)
pJR <sub>1</sub> -IFN $\gamma$	Ap <sup>R</sup>	LE392	2560 u/ml	efekt cytopatyczny: wirus encefalomyocarditis/komórki linii raka płuc A549	(14)
pCGtac502	tac	MC1061	$1,2 \times 10^5$ u/ml	efekt cytopatyczny: wirus encefalomyocarditis/komórki WISH	(15)
pLPPY	llp + lac	W3110	21% białka komórkowego	nie określano	(17)
pL <sub>29</sub> IF <sub>6</sub>	T5P25	LE392	48% białka komórkowego	nie określano	(32)

W poszukiwaniu dobrych promotorów wykorzystano układy tandemowe, jak np. kombinację promotora *llp* i regionu promotora laktozy (17), które również pozwoliły na uzyskiwanie dużych ilości białka.

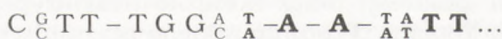
Systemy regulacji ekspresji przez kontrolę transkrypcji polegają również na ścisłej kontroli jej inicjacji. Prototyp takiego plazmidu ekspresyjnego opierał się na odwracalnym ukierunkowaniu wklonowanego promotora. Białko Int faga lambda indukowane impulsem termicznym powodowało „właściwy” zwrot odcinka promotora (18). Dwuplazmidowy układ ekspresyjny, w którym jeden z plazmidów nosił gen białka heterologicznego i odwracalny promotor, drugi zaś — indukowalny gen integrazy, pozwalał na uzyskiwanie białek toksycznych dla *E. coli* (19).



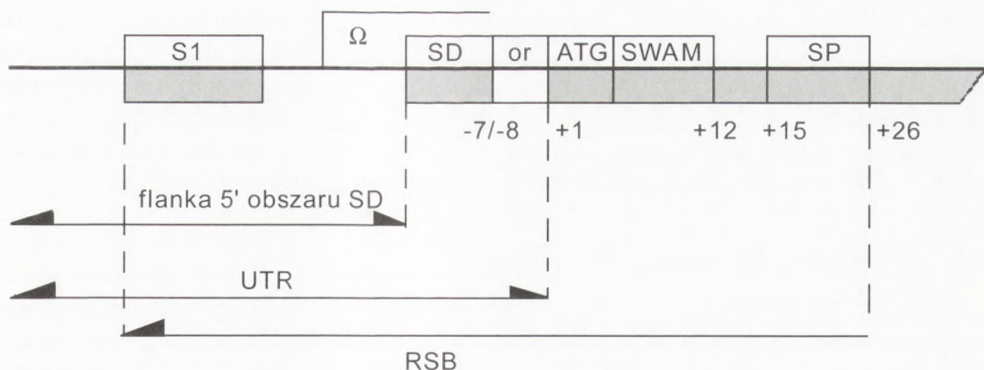
Dobór terminatora transkrypcji zasadniczo zmienia obraz ekspresji. Terminator genu 3 faga T4 dwukrotnie podnosi poziom syntezy IFN $\gamma$  w układzie zawierającym promotor PL i sekwencję Shine-Dalgarno pochodzącą z regionu atenuatora tryptofanu *E. coli*. Ten sam terminator — w przypadku gdy obszar SD pochodzi z faga MS2 — podnosi poziom syntezy IFN $\gamma$  trzykrotnie w porównaniu z układem nie zawierającym żadnego sygnału terminacji (20). Bardzo wysoki poziom ekspresji komórkowej uzyskano klonując pod kontrolą promotora PN25, gen interferonu zakończony terminatorem  $t_0$  faga lambda. Innym wartym odnotowania przykładem jest plazmid pDS6. Zawiera on te same elementy kontroli transkrypcji, ale posiada podwójne miejsce wiązania rybosomu, czym dodatkowo wzmacnia wydajność translacji (21).

Stabilność transkryptu zależy nie tylko od struktury 3' jak i 5' końców mRNA, lecz także od obecności w komórce białek biorących udział w procesach degradacji transkryptu (np. RNaza E, RNaza II) (22,23).

Obecność struktur typu „szpilki do włosów” na krańcach transkryptu ma znaczenie nie tylko dla zahamowania procesu degradacji (24), ale również — jak postulują niektórzy autorzy — prowadzi do obniżenia zawartości struktur dwuniciowych w pozostałej części cząsteczki mRNA (25). Wytwarzanie tych struktur na początku transkryptu genu odbywa się kosztem wystarczającej ilości energii aby czas dostępności całego miejsca inicjacji podjednostki rybosomu uległ znacznemu wydłużeniu. Ta wielce prawdopodobna hipoteza nie znalazła, jak na razie, potwierdzenia ani tym bardziej zastosowania. Podobnie w celu poprawienia poziomu transkrypcji nie stosuje się metod zwiększania stabilności mRNA (10). Tymczasem sama sekwencja niekodująca również może mieć istotny wpływ na wydajność syntezy transkryptu: sekwencja ograniczająca od strony 5' część kodującą genu pod kontrolą promotora *trp* pozwala trzykrotnie podwyższyć poziom transkrypcji, a ilość aktywnego IFN $\gamma$  — nawet sześciokrotnie, jeśli zostaje poszerzona o odpowiedni fragment z promotora PL. Obie te sekwencje tzn. zarówno poprzedzająca promotor *trp* jak i poprzedzająca promotor PL są wzbogacone w pary (A+T) (26). Delecje segmentu zawierającego dwa takie „bloki” par (A+T) charakterystycznych dla promotora *trp*, doprowadzają do obniżenia jego zdolności wiązania podjednostki rybosomu. Sekwencje te leżą powyżej (w kierunku 5') obszaru (-35) promotora. Pierwsza z nich znajduje się w okolicach pozycji (-50), druga w pobliżu pozycji (-90). W swej części 5' pokrywają się one ze znaną w tych obszarach sekwencją zgodności, charakterystyczną dla natywnych promotorów *E. coli*. Ma ona postać:



Długość 3'-skrajnego odcinka (A+T), który zaznaczono przez podkreślenie, wynosi około 10 – 14 nukleotydów. Mechanizm opisywanego zjawiska nie jest do końca jasny, proponowane wyjaśnienia odnoszą się albo do bezpośredniego oddziaływania tej sekwencji z polimerazą RNA, bądź z bliżej nieokreślonym czynnikiem transkrypcyjnym (27).



Rys. 1. Zestawienie fragmentów początku transkryptu genu: S1 — miejsce wiązania białka S1 (Mertens, 1995), SD — obszar Shine-Dalgarno, or — „obszar rozdzielający”, SP — przykład sekwencji z obszaru kodującego, komplementarnej do 16S rRNA, tu: do fragmentu 1471 – 1482 (Sprengart, 1990), SWAM — sekwencja pierwszych czterech kodonów, komplementarna do fragmentu 1394 – 1399 w 16S rRNA (Chen, 1994), Ω — alternatywny obszar niekodujący, komplementarny do fragmentu 1343 – 1355 w 16S rRNA, mogący służyć jako miejsce SD (Ivanov, 1995), UTR — obszar nie tłumaczony na sekwencję aminokwasową (*untranslated region*), RBS — obszar wiązania rybosomu.

Duża liczba kopii plazmidu obniża niekiedy poziom transkrypcji, ma to miejsce, gdy dochodzi do oddziaływania kompleksu polimerazy RNA z replisomem. Zjawisko to dodatkowo podkreśla znaczenie obecności terminatorów transkrypcji (10). Mimo jednak, że najpopularniejsze typy plazmidów ekspresyjnych to te, które można nazwać pochodnymi pBR322 (10) (kilkadziesiąt kopii na komórkę), stosuje się również plazmidy ekspresyjne o wysokiej liczbie kopii, jak np. plazmidy serii pCZ utworzonej z plazmidu pMIA (2000 kopii na komórkę) (27).

#### 4. Regulacja ekspresji na etapie translacji

Wśród czynników mających wpływ na wydajność ekspresji, a związanych bezpośrednio ze strukturą informacyjnego RNA, można wyróżnić kilka ściśle zdefiniowanych. Są to: sekwencja obszaru Shine-Dalgarno, kodon inicjacji translacji, sekwencja rozdzielająca te dwa obszary — tzw. „spacer region”, sekwencja obszaru leżącego „powyżej” sekwencji SD, a także sekwencja „poniżej” kodonu inicjacji (rys. 1).

W obszarze rozdzielającym (rys. 1) niektórzy autorzy wyróżniają kilkunukleotydowy odcinek od końca 5' oraz trzy- czteronukleotydowy fragment na jego 3' końcu. Innym parametrem, niekiedy silnie zależnym od wymienionych, jest drugorzędowa struktura informacyjnego RNA. W szczególności



podstawowe znaczenie ma obecnořć struktur typu „szpilki do wlosów” w poczatkowej części genu. Istnieją jednak równieŹ przesłanki wynikające z termodynamicznych wlaściwości nici RNA sugerujące zasadniczy wpływ na poziom ekspresji genu struktur drugorzędowych mRNA występujących na całej jego długości. Jedna zamiana G(+35)→C(+35) powoduje zmniejszenie energii swobodnej struktur drugorzędowych o  $\Delta G = 2,8$  kcal/mol i zwiększenie poziomu syntezy IFN $\beta$  z 1 do 28% całkowitego białka komórkowego (28).

#### 4.1. Sekwencja Shine-Dalgarno

Zależnořć ekspresji IFN $\gamma$  od postaci sekwencji Shine-Dalgarno przez długi czas nie stanowiła przedmiotu systematycznej analizy. Praca H. Chen i wsp. stanowi tu raczej wyjątek (29). Ekspresję IFN $\gamma$  prowadzono w komórkach LE392 po umieszczeniu genu pod kontrolą promotora T5P25 w plazmidzie — pochodnej pBR322 (tab. 1). Najwyższy poziom ekspresji uzyskiwano, gdy obszar SD stanowił siedmionukleotydzowy odcinek postaci: T AGGAGGT TT (17% całkowitego białka komórkowego). Podstawienie A →T w pozycji pierwszej obniŹało wydajnořć do około 10%, natomiast podstawienie dwóch kolejnych adenin w miejsca tymin powodowało jedynie nieznaczne (o około 4%), dalsze zahamowanie ekspresji. Zastosowane w tych dořwiadczeniach sekwencje zestawiono w tabeli 2.

TABELA 2  
ZMIANY POZIOMU EKSPRESJI INTERFERONU-GAMMA  
ZALEŹNIE OD DŁUGOŚCI OPTIMALNEGO ODCINKA SHINE-DALGARNO (32)

TGAAGCTACGTT <u>GGAGGT</u> TTAAGCTT.	<b>ATG</b> .TGT.TAC.TGC.CAG.GAC.CC	ca.7%	SD=6nt
TGAAGCTACGT <u>AGGAGGT</u> TTAAGCTT.	<b>ATG</b> .TGT.TAC.TGC.CAG.GAC.CC	ca.15%	SD=7nt
TGAAGCTACG <u>AAGGAGGT</u> TTAAGCTT.	<b>ATG</b> .TGT.TAC.TGC.CAG.GAC.CC	ca.14 %	SD=8nt

Odcinek rozdzielający zawiera miejsce restrykcyjne *HindIII*, nt — liczba nukleotydów.

Wcześniejsza praca E. Remaut i wsp., dotyczy raczej wpływu na poziom ekspresji całego obszaru wiązania rybosomu (20). Opisane w tej pracy wprowadzenie obszaru RBS pochodzącego z genu replikazy RNA faga MS2, z silnym promotorem PL faga  $\lambda$ , pozwala na ekspresję interferonu gamma wynoszącą 3,5% całkowitego białka komórkowego, podczas gdy zastąpienie go sekwencją pochodzącą z atenuatora tryptofanowego podnosi ją do 15%. Wartości te według ostatnio publikowanych danych (Mertens i wsp. 1995) wynoszą odpowiednio 10 i 25% całkowitego białka komórkowego. Natomiast zastosowanie obszaru RBS pochodzącego z faga T7g10 podnosi poziom akumulowanego białka do wartości 52% (30). Obszar wiązania rybosomu (RBS — *Ribosomal Binding Site*) odgrywa w procesie translacji rolę analogiczną do tej jaką odgrywa promotor w trakcie transkrypcji i w tym sensie można go traktować jako „translacyjny odpowiednik promotora” (31). Obszar ten zawiera około 30 – 40 nukleotydów (32). Siła RBS — czyli jego zdolnořć do wiązania podjednostki 30S rybosomu, zależy w głównej mierze od tworzących



się lokalnie struktur drugorzędowych (33). Niezależnie od tego, na skuteczność wiązania wpływa również w sposób istotny sekwencja, w której skład wchodzi wspomniany obszar SD, odcinek rozdzielający, oraz pierwsze nukleotydy tak sekwencji kodującej jak i poprzedzającej bezpośrednio sekwencję SD. Z prac Golda i Stormo znana jest sekwencja tzw. uniwersalnych RBS zawierająca wszystkie wymienione elementy (31). Ma ona umożliwiać ekspresję dowolnego obcego białka. Obok elementów kanonicznych, zawiera liczne miejsca pozwalające na dopasowanie „uniwersalnego” miejsca RBS do konkretnego genu, poprzez wstawienie doń właściwych (optymalnych) nukleotydów. W sekwencji tej korzystne są trakty bogate w pary (A+T) lub nawet sekwencje oligo-A. Sekwencja uniwersalnego RBS nie może być jednak traktowana jako optymalna, gdyż elementy kanoniczne nie mogą uwzględniać jednocześnie wszystkich uwarunkowań strukturalnych dla dowolnego genu. Nie składają się zatem na nią również pozycje o charakterze zachowawczym, czyli utrzymujące jednocześnie podwyższony poziom ekspresji, ale rzadziej ulegające mutacji (34).

#### 4.2. Odcinek rozdzielający

Zmienna częstość mutacji w obszarze rozdzielającym SD i kodon inicjacji obserwowali Mateucci i wsp., a występujące zmiany poziomu ekspresji białka zależały silnie od tych mutacji (34). W miejscach (-9), (-8), (-1) licząc od kodonu inicjującego ATG (A = +1) prawdopodobieństwo wystąpienia mutacji jest jak  $5 \times 10^{-2} - 10^{-1}$ , a w pozycjach (-7), (-6), (-3) jest jak  $2 \times 10^{-2} - 3 \times 10^{-2}$ . Występowanie tyminy w pozycji (-4), jak się wydaje, ma charakter konserwatywny. Natomiast zmiany w pozycjach (-3) i (-2), które są bardziej prawdopodobne, mogą znacząco wpływać na poziom ekspresji. Podstawienie G(-3) → T(-3) obniża ekspresję do 30%, a podstawienie A(-3) → T(-3) zwiększa do 130% wartości początkowej. Zmiany wynikające z mutacji w pozycji (-2) są mniejsze: podstawienie T(-2) → C(-2) podwyższa ekspresję o 10% w stosunku do wartości początkowej, a podstawienie A(-2), G(-2) → T(-2) obniża do 60% tej wartości. Guanina w pozycji (-1), może znacznie zahamować ekspresję. Prawdopodobnie wiąże się to albo z hamowaniem translacji przez powstawanie drugorzędowych struktur (niemożliwych do przewidzenia metodami rachunkowymi) albo z osłabieniem oddziaływania z nukleotydami pętli antykodonu fMet-tRNA, albo też z wytwarzaniem niewłaściwych wiązań pomiędzy mRNA, a z końcem 3' 16S rRNA (35). Dzieje się tak w przypadku tripletu ACC, który jeśli występuje w pobliżu tej pozycji (za kodonem ATG) może zahamować translację, np. białka hirudyny poprzez zbyt duże powinowactwo do otoczenia rejonu ASD (anty-SD) w 16S rRNA (35). Mimo to należy zaznaczyć, że znane są układy ekspresyjne dostarczające znacznych ilości IFN $\gamma$ , a zawierające G w pozycji (-1) (36) (por. rys. 2).

Wpływ sekwencji poprzedzającej bezpośrednio start translacji na ekspresję innych białek badano metodami mutagenyzy kasetowej. Wykazano uprzywilejowaną rolę tyminy w pozycji (-1) i adeniny/tyminy w pozycji (-2) np. UAU i CUU są tripletami wspomagającymi ekspresję genu *lacZ*, wtedy gdy



poprzedzają kodon inicjacji translacji (35). Natomiast w przypadku IFN $\beta$  podstawienie C $\rightarrow$ U w pozycji (-2) lub (-1) obniża wydajność syntezy do odpowiednio 8 i 4% wartości początkowej wynoszącej 18%, zaś jednoczesne podstawienie cytozyny w obu miejscach powoduje spadek syntezy do wartości poniżej 1% (29). Podobnie dzieje się dla wielu rozmaitych białek, jak np. acylazy penicyliny czy IFN $\gamma$ : sekwencja i długość odcinka rozdzielającego ograniczają w analogiczny sposób ich ekspresję (37). Regułą jest, że odległość pomiędzy odcinkiem SD, a kodonem ATG nie powinna być mniejsza niż 7-8 nukleotydów (37,38), a skrócenie jej do 6-7 powoduje spadek aktywności względnej uzyskiwanego IFN $\gamma$  (mierzonej hamowaniem efektu cytopatycznego na komórkach WISH) do 15%. Skrócenie tego odcinka do 5 nukleotydów powoduje spadek ekspresji poniżej 10% aktywności, a do 4 — blokuje ekspresję całkowicie (37). Wydłużenie odcinka rozdzielającego do 11 nukleotydów nie zmienia istotnie wartości ekspresji w porównaniu z układem optymalnym (8 nukleotydów). Również wprowadzenie zmian wewnątrz środkowej części jedenastonukleotydowego odcinka, o ile nie zmienia to struktury wyższego rzędu (por. tab. 3), nie wpływa na ekspresję IFN $\gamma$ . Podobny wynik uzyskano dla łańcucha  $\mu$  mysiej immunoglobuliny, gdzie wydłużenie 7-nukleotydowej sekwencji:

GTGATCA ATG

do postaci:

GTATTGATCA ATG

nie zmieniało poziomu ekspresji o więcej niż o 0,4 ru (*relative units*) (39) (por. tab. 3).

TABELA 3  
WARIANTY ODCINKA ROZDZIELAJĄCEGO SD  
I KODON INICJACJI TRANSLACJI, A POZIOM EKSPRESJI IFN $\gamma$  (41)

AGGA <b>CACAG</b> AUG	ca 0%	4nt
AGGA <b>ACAGAU</b> CU AUG	100%	8nt
AGGA <b>UCAGAUCAUCU</b> .AUG	100%	11nt
AGGA <b>ACAGAUGAUCU</b> .AUG	100%	11nt

Zestawienie sekwencji obszaru rozdzielającego (or) sekwencję SD i kodon inicjacji translacji, użytych w układach ekspresyjnych dla interferonu gamma, gen pod kontrolą promotora PL, w układzie dwucistronowym. Sekwencje dłuższe niż 8 nukleotydów nie zmieniają wydajności ekspresji.

### 4.3. Wzajemne relacje pomiędzy fragmentami obszaru RBS

Odcinek leżący na końcu 3' sekwencji SD — a zatem obszar rozdzielający, może być dodatkowo kompensowany 8-nukleotydowym odcinkiem leżącym na jego 5' końcu. Zmiany w sekwencji tego odcinka modulują ekspresję w zakresie nawet dwóch rzędów wielkości. Utworzenie struktury o ener-



gii  $\Delta G > |-10,7 \text{ kcal/mol}|$  wystarcza aby całkowicie zablokować ekspresję nawet przy „silnej” sekwencji SD postaci GGAG (40). Zatem samo wyeksponowanie miejsca SD bądź kodonu inicjacji translacji nie zawsze wystarcza. Należy raczej przyjąć, że siła miejsca RBS, w znacznej mierze zależy od jego udziału w strukturach dwuniciowych, a ściślej od energii koniecznej do usunięcia tych struktur. Inaczej mówiąc, regulacja wydajności następuje bardziej przez wyeksponowanie całego obszaru inicjacji, niż przez stan łańcucha nukleotydowego poza tym obszarem. Można jednak również mówić o korelacji między wydajnością translacji a wielkością frakcji zrelaksowanych jednoniciowych mRNA. Przy czym przez stan relaksacji należy rozumieć stan optymalnej energetycznie konformacji cząsteczki w danym układzie. Tak zatem, niekiedy, jak w przypadku białka płaszczka faga MS2, okazuje się, że mutacje w pozycjach (-6), (-3), (+6), (+9), (+12), (+13), (+15) licząc od AUG (A = +1) wpływają na ekspresję jedynie, gdy zmieniają stabilność drugorzędowych struktur transkryptu (5). Tworząca się w otoczeniu kodonu inicjacji translacji 12-nukleotydowa struktura „szpilki do włosów” obniża poziom ekspresji tego białka. Przy stopniowym zwiększaniu stabilności tej struktury, na każde kolejne -1,2 kcal/mol przypada 10-krotne obniżenie wartości ekspresji (41). W tym kontekście możliwe jest wyjaśnienie tzw. paradoksu przewagi „prawdziwego” miejsca inicjacji nad „przypadkowym”, czyli miejsca odpowiadającego początkowi translacji określonego genu nad miejscem, jedynie przypominającym swą sekwencją miejsce inicjacji. Wynika ona z energetycznej przewagi struktury „poprawnej” (5). Warto jednak podkreślić, że nie zawsze możliwe jest stwierdzenie czy poziom ekspresji zależy silniej od energii wynikającej z tworzonych wewnątrzcząsteczkowych struktur drugorzędowych czy od samej struktury pierwszorzędowej, która odpowiada za wiązanie podjednostki 30S rybosomu. Wydaje się, że zależność ta zmienia się wraz z położeniem rozpatrywanego obszaru, a wpływać mogą na nią liczne oddziaływania. Przykładowo, obszar TIR (*Translation Initiation Region*) opisany w pracy Sprengarta i wsp. wpływa na poziom ekspresji genu 0,3 faga T7 w sposób zależny od sekwencji odcinka (+9) - (+37), w szczególności zaś od odcinka (+15) - (+26), który to fragment jest komplementarny do sekwencji 1471 - 1482 16S rRNA. Obserwacja ta wykazuje istnienie dodatkowych miejsc oddziaływania rybosomalnych RNA i mRNA wspomagających translację niektórych genów (42). Kilka wybranych sekwencji TIR, zastosowanych w układach ekspresyjnych w *E. coli* dla uzyskania rekombinowanego interferonu gamma, zestawiono na rysunku 2.

Drugorzędowa struktura powstająca na tym obszarze ma charakter konserwatywny tzn. nawet zniszczona poprzez wprowadzenie mutacji punktowych zostaje odtworzona w ciągu kilkuset podziałów komórkowych, drogą spontanicznych mutacji punktowych kompensujących skutki wprowadzonych zmian. Eksperymenty prowadzone nad takimi układami pozwalały zaobserwować w trakcie kolejnych podziałów przywracanie pierwotnych struktur drugorzędowych. Powstające w tych procesach mutacje spontaniczne następowały poza sekwencją SD i nie zmieniały kodowanych aminokwasów. Współzależność postaci obszaru SD i wytwarzanych struktur drugorzędowych



**wektor pJP<sub>1</sub>R<sub>3</sub>-IFN $\gamma$** 

**promotor:** T5P25 +1 SD  
 TATAATAGATTCAAAAATTTGAAGCTAGGAGGTTTAAGCTT. **ATG.TGT.TAC.TGC.CAG.GAC.C**

**wektor pJR<sub>1</sub>-IFN $\gamma$** 

**gen  $\beta$ -lac** SD  
 ACGATGCCTGACTAAGGAGGTTTAAGCTT **ATG.TGT.TAC.TGC.CAG.GAC.C**

**wektor: pDS6****promotor PN25**

+1  
 TATAATAGATTCAAATTTGTGAGCGGATAACAATTTGAATTC----- "RBS"-----**ATG**

**wektor: pIFN $\gamma$ -trp-1/2**

+1 SD  
 AAGTTCACGTAAAAAGGGTATCGCGGAATTCAGGAGGCCTCTAG **ATG.TGT.TAC.TGC.CAG.GAC**

Rys. 2. Sekwencje pogranicza promotora i początkowych fragmentów części kodującej w wybranych układach ekspresji interferonu gamma. Początek transkryptu opisano znakiem +1, obszar Shine-Dalgarno (SD) podkreślono. W części kodującej kolejne kodony genu interferonu gamma oddzielono kropkami. Sekwencja bezpośredniego otoczenia SDII w plazmidzie pDS6, opisana jako „RBS” nie została podana przez autorów. W przypadku plazmidu pJR<sub>1</sub>-IFN $\gamma$  miejsce startu transkrypcji znajduje się przed pierwszym cistronem. Podane układy opisano w tekście.

końca 5' transkryptu jest przedmiotem analizy Olsthoorna i wsp. Wynika z niej, że stopień komplementarności pomiędzy sekwencjami SD i ASD w 16S rRNA oraz wartość energii wytwarzanej struktury „szpilki do włosów” pozostają w równowadze (43).

Stosowane niekiedy powielanie miejsca RBS, prowadzące do powstawania tandemu bądź nawet wielokrotnego miejsca wiązania rybosomu, nie zwiększa w sposób jednoznaczny wydajności translacji interferonu. Istnieje bowiem krytyczna odległość sąsiadujących sekwencji SD. Sąsiedztwo bliższe niż 40 nukleotydów wymusza na tych sekwencjach konkurencję o dostęp do podjednostek 30S rybosomu. Zjawisko to ilustrują eksperymenty nad syntezą hybrydowego białka hIFN $\gamma$ -hCT. Podwojenie poprzedzającej sekwencję kodującą, dwunastonukleotyduj sekwencji „uniwersalnego” SD:

TGG AGA AGG AGG TTTA(A)

powoduje spadek ekspresji aż do 17% wartości początkowej. Trzykrotne powtórzenie tej sekwencji obniża wydajność syntezy białka do 11%. Spadek ekspresji, aczkolwiek wynika również z inhibicji transkrypcji (odpowiednio 60 i 40%), jest jednak zbyt znaczny by mógł być całkowicie w ten sposób wyjaśniony. Utworzenie tetrameru sekwencji SD powoduje niewielki wzrost syntezy RNA i poziomu ekspresji, nie przywracając jednak poziomu pierwot-



nego (44). Powójny układ RBS bywa stosowany w sposób skuteczny i pozwala niekiedy na uzyskanie w *E. coli* dużych ilości IFN $\gamma$  (21).

#### 4.4. Sekwencja końca 5' genu

Na wydajność syntezy mają również istotny wpływ nukleotydy sekwencji kodującej. Sekwencja ta rozpoczyna się kodonem ATG, najczęstszym naturalnym kodonem inicjacji w *E. coli*. Inne: GTG, TTG, ATC (45), znacznie mniej ważą na procesie translacji. Pierwsze kodony decydują również o hamowaniu translacji na skutek powstawania struktur drugorzędowych na ich obszarze. Struktury te mogą nie tylko przesłaniać kodon inicjacji, ale również wiązać koniec 5', sekwencji kodującej. Zależność pomiędzy wydajnością translacji, a stabilnością struktur dwuniciowych na tym obszarze ilustruje przykład IFN $\gamma$ . Jeśli energia potrzebna do wytworzenia takich struktur wynosi -4,4 kcal/mol, wówczas synteza IFN $\gamma$  osiąga poziom około 19% całkowitego białka komórkowego. Przy wzroście wartości tej energii do -6,6 kcal/mol, poziom syntezy IFN $\gamma$  obniża się do 10%. Nie jest jednak regułą, że mniejsza wartość energii zawartej w tworzących się strukturach dwuniciowych przesądza o większej wartości ekspresji niesionego genu (37) (por. tab. 4). Nie decydują o tym także użyte kodony nawet jeśli są one optymalne dla *E. coli*, co wyraźnie podkreśla zestawienie częstości występowania zastosowanych tripletów. Być może, przykład ten należy wiązać z obserwacją de Boera dotyczącą znaczenia otoczenia pozycji (+10) (35). Najbliższe otoczenie tej pozycji powoduje bowiem prawie 30-krotny wzrost poziomu ekspresji IFN $\gamma$  wyłącznie na skutek mutacji w pozycji (+12), nie powodując jednocześnie zmian strukturalnych transkryptu, możliwych do wykrycia metodami analizy numerycznej (46). Pewne uzupełnienie mogą tu stanowić dwa kolejne przykłady z cytowanej pracy L.H.Tessier i wsp. — ciche mutacje dokonane w pozycji (+9) i (+12) zmniejszają średnią częstość występowania zastosowanych kodonów o około 0,5‰ (por. tab. 5) i nie zmieniają stanu konformacji łańcucha RNA. W wyniku tej operacji następuje wzrost ekspresji o 8%. (por. tab. 5).

TABELA 4  
ZMIANY POZIOMU EKSPRESJI IFN $\gamma$  W ZALEŻNOŚCI OD POSTACI KOŃCA 5' GENU (41)

		<v>	$\Delta G$	%T	ru
T1	TGC TAC TGC 6.2 13.4 6.2	8.6	+1,2 kcal/mol	24%	20%
T4	TGC TAT TGT 6.2 14.9 4.7	8.6	-1,2 kcal/mol	33%	100%



TABELA 5  
ZMIANY POZIOMU EKSPRESJI IFN $\gamma$  W ZALEŻNOŚCI OD POSTACI KOŃCA 5' GENU (41)

		<v>	$\Delta G$	%T	ru
T2	TGC TAT TGC				
	6.2 14.9 6.2	9.1	-1.2 kcal/mol	25%	20%
T4	TGC TAT TGT				
	6.2 14.9 4.7	8.6	-1.2 kcal/mol	33%	100%

Tabele przedstawiają wartości energii wynikające z tworzenia się struktury drugorzędowej na początku genu, oraz odpowiadające im wartości wydajności ekspresji podane zarówno jako procent całkowitego białka komórkowego jak i w wartościach względnej aktywności otrzymanego IFN $\gamma$ .

<v> — uśredniona wartość częstości stosowanych kodonów na tym obszarze,  $\Delta G$  — energia potrzebna do wytworzenia struktur drugorzędowych na jednoniciowym fragmencie mRNA, %T — procent całkowitego białka komórkowego, ru — względna aktywność produktu (*relative units*); pod każdym tripletem nukleotydowym podano częstość jego występowania w genomie *E. coli*, przypadającą na 1000 nukleotydów chromosomalnego DNA. Wartości częstości „v” podano za: (6).

Prowadzona przez H. Chen'a (29) analiza zależności wydajności ekspresji IFN $\gamma$  od mutacji w obszarze 5'-końca samego genu uwzględnia całkowitą zamianę pierwszych 11 – 12 nukleotydów genu IFN $\gamma$  wchodzących w skład krańca 3' miejsca RBS. Sekwencje bogate w pary (A+T) znacznie podwyższają wydajność ekspresji. Dzięki temu wprowadzenie sekwencji odpowiadającej czterem pierwszym kodonom acetylotransferazy chloramfenikolu:

GAG AAA AAA ATC

lub sekwencji:

AAA AAA ATT AA

znacznie zwiększa (3-5-krotnie) poziom ekspresji tak zmienionego interferonu-gamma (29). Wydaje się, że można wspomniany efekt wiązać z obserwacjami zespołów Scherera i wsp. oraz Schneidera i wsp., którzy podobnie jak i Dreyfus odnotowywali tendencję do zwiększonego udziału adenin w tej części RBS (29). Ustalono jej optymalną sekwencję SWAM (Scherer, Walkins-haw, Arnott, Morre) o postaci AAAAAAATTAA. Wykazano (29), że również sekwencje podobne do SWAM wzmacniają ekspresję w obecności funkcjonalnego 5' końca RBS, co więcej umożliwiają ekspresję, gdy koniec 5' miejsca RBS nie ma zdolności wiązania rybosomu. Ponieważ, jak się wydaje, na tym obszarze nie występują żadne drugorzędowe struktury, zaproponowano (29) istnienie dodatkowej domeny oddziałującej z rybosomem w trakcie tworzenia kompleksu inicjacji translacji. Proponowanym miejscem oddziaływania jest albo odcinek 1394 – 1399 16S rRNA (odpowiadający miejscu P rybosomu), lub jego końca 5'. Koniec ten posiada miejsca komplementarne do fragmentów końca 5' licznych genów *E. coli*. Wydaje się, że wystarczyć mogą nawet



dwa triplety elementu SWAM aby wystąpiło oddziaływanie podobne do oddziaływania SD/ASD. Podobne, A+T bogate sekwencje, gdy poprzedzają kodon inicjacji wystarczają niekiedy do niezależnego rozpoczęcia procesu translacji. Na przykład sekwencje pochodzące z sekwencji  $\Omega$  wirusa mozaiki tytoniowej zawierające fragment:

UUACAAUUACUA

i będące komplementarnymi do odcinka 1343 -1355 16S rRNA, są w stanie zastępować odcinek SD ze skutecznością rzędu 20% normalnego poziomu ekspresji (47).

Początek genu, dokładniej pierwszy triplet nukleotydowy następujący za kodonem metioniny, podwyższa niekiedy poziom translacji. Następuje to wtedy, gdy dzięki obecności puryny bezpośrednio za kodonem ATG zwiększa się stopień komplementarności pomiędzy mRNA, a pętlą antykodonu fMet tRNA (48). Natomiast usunięcie 9 pierwszych nukleotydów w genie IFN $\gamma$  powoduje 10-krotny wzrost poziomu ekspresji do wartości 7,5% całkowitego białka komórkowego. Delecja ta, w przypadku układu ekspresyjnego zawierającego promotor *tac* powoduje, że powstająca w otoczeniu kodonu inicjacji translacji struktura dwuniciowa jest o wiele mniej stabilna. Obliczenia dokonane na obszarze 60 pierwszych nukleotydów informacyjnego RNA wykazywały tworzenie się tzw. „struktury szpilki do włosów” o wartości odpowiednio -9,9 kcal/mol i -5,7 kcal/mol dla pełnego i zmodyfikowanego przez delecję genu IFN $\gamma$  (49).

#### 4.5. Sekwencja 5'- flankująca SD

Odrębnym, regulującym elementem jest sekwencja ograniczająca od strony 5' obszar SD. Odcinek ten, stanowiący jednocześnie ochronę przed degradacją egzonukleolityczną, bierze również udział w wiązaniu rybosomu. Tworzone na tym obszarze struktury mają znaczenie regulacyjne. Dla zajścia inicjacji translacji niezbędne jest przynajmniej około 10 nukleotydów poprzedzających obszar SD, a wydłużenie tego odcinka o dalsze 8 zasad znacznie zwiększa ekspresję (35). Zazwyczaj w syntetycznym układzie ekspresyjnym sekwencja ta jest krótka i nie powinna przekraczać 40 – 60 nukleotydów.

Oddziaływania angażujące nukleotydy końca 5' transkryptu są w stanie zapewniać relaksację całego łańcucha nukleotydowego (50). Oddziaływania te prowadzą do tworzenia struktur dwuniciowych typu „szpilki do włosów”. W niektórych genach występujących w naturze, jak w przypadku białka płaszczka faga MS2 powstawanie tych struktur oddziałuje na kilkusetnukleotydowe fragmenty łańcucha. Identyczny efekt relaksacji można osiągnąć przy znacznie krótszych flankach końca 5' genu zapewniając im właściwą sekwencję. Podobne zjawisko zaobserwowano badając wpływ 21-nukleotydowego fragmentu poprzedzającego obszar RBS, na ekspresję genu reduktazy dihydrofolanu (24). W badanym układzie ekspresyjnym obszar wiązania rybosomu (pochodzący z genu 1 bakteriofaga T7) podlega wyeksponowaniu dzięki wy-



twarzaniu na krańcu 5' transkryptu silnej struktury dwuniciowej. W ten sposób wspomniany 21-nukleotydowy fragment (pochodzący z genu reduktazy), powoduje podniesienie ekspresji o rząd wielkości (24). Jednocześnie struktury formujące się dzięki oddziaływaniu obszaru najbliższego otoczenia sekwencji SD (tzw. obszar periSD) z fragmentami sekwencji leżącymi dalej w kierunku 3', w głębi RBS, mogą prowadzić niekiedy do 100-krotnego zahamowania ekspresji, jak ma to miejsce w wypadku interferonu-gamma (44).

W cytowanej wcześniej pracy H. Chen i wsp. (29) przebadano szereg krótkich sekwencji poprzedzających optymalną strukturę Shine-Dalgarno wraz z ich wpływem na poziom ekspresji IFN $\gamma$ . Odcinek poprzedzający kodon startu stanowiła sekwencja rozpoznawana przez endonukleazę *HindIII*, poprzedzona dwiema tymidynami. Śladową ekspresję (<2%) białka uzyskiwano, gdy odcinek flankujący był dłuższy niż 9 nukleotydów, optymalną gdy miał on długość 16 nukleotydów (około 16,8% całkowitego białka komórkowego), a sekwencja flankująca miała postać: ATAAATTTG; wstawienie dodatkowych czterech nukleotydów TCGA nie podnosi dalej poziomu ekspresji, lecz zmniejsza go do około 7% (por. tab. 6).

TABELA 6

ZMIANY POZIOMU EKSPRESJI INTERFERONU GAMMA LICZONEGO JAKO PROCENT CAŁKOWITEGO BIAŁKA KOMÓRKOWEGO W ZALEŻNOŚCI OD DŁUGOŚCI ODCINKA POPRZEDZAJĄCEGO OBSZAR SD (32)

ATAAATTTG	GTAGGAGGTTTAAGCTT.ATG	<2%	9nt
ATAAATTTGAAGCTAC	GTAGGAGGTTTAAGCTT.ATG	ca17%	16nt
ATAAATTTGAAGCTACTCGA	GTAGGAGGTTTAAGCTT.ATG	ca7%	20nt

Odcinek rozdzielający (or) zawiera miejsce restrykcyjne *HindIII*.

#### 4.6. Układy dwucistronowe

Innym sposobem, prowadzącym do wydajnej ekspresji heterologicznego białka jest tzw. układ dwucistronowy. Pierwszy taki system oparty na plazmidzie pMIA, charakteryzującym się wysoką liczbą kopii, został skonstruowany dla ekspresji bydłęcego hormonu wzrostu bGH pod kontrolą promotorów *trp* i *lpp*. W tym układzie pierwszy, krótki peptyd podlega wydajnej ekspresji, a dwa miejsca SD dobrane są tak, by umożliwić optymalne wiązanie rybosomu. Kodon stopu dla pierwszego peptydu umieszczony jest za drugim miejscem SD, powodując wyeksponowanie tego obszaru (51).

Dwie sekwencje służące ekspresji bGH oparte na tym systemie przedstawiono poniżej:

**ATGTATCGATTAAATAAGGAGGAATAACATATG ...**  
**ATGTTCCATTGGAGGATGATTAATG...**

Zadaniem pierwszego cistronu jest zapewnienie optymalnej inicjacji translacji (52). Jego dobór zależy w wysokim stopniu od wybranego promotora







tw. ciałek inkluzyjnych. Skład i struktura agregatów nazywanych ciałkami inkluzyjnymi zależy od przedziału komórkowego w jakim powstają. O ile w cytozolu przybierają one formę bardziej regularną — quasi-krystaliczną — i w przypadku IFN $\gamma$  są to obiekty w kształcie zbliżonym do walca o wymiarach 1,26  $\mu\text{m}$  na 0,81  $\mu\text{m}$ , o tyle w periplazmie mają postać amorficzną. Miejsce tworzenia inkluzji określa jej morfologię, a w rezultacie i jej rozpuszczalność *in vitro* (60). Samo powstawanie inkluzji zależne jest od temperatury hodowli, składu podłoża (np. obecności nie metabolizowalnych cukrów), szczepu bakteryjnego, jak i wreszcie od stopnia modyfikacji sekwencji samego białka.

Uważa się, że istnieje więcej niż jeden mechanizm powstawania inkluzji, a równowagę agregacji są w stanie przesunąć nawet pojedyncze aminokwasy podstawione w określonej pozycji łańcucha peptydowego.

Interferon gamma — jako stosunkowo małe białko — stanowi dobry układ modelowy do badań nad zależnością agregacji od składu aminokwasowego i struktury pierwszorzędowej. Z badań krystalograficznych wiadomo, że koniec C cząsteczki nie wchodzi w struktury wyższego rzędu i nawet daleko idące zmiany pozostają bez znaczenia dla jego właściwości immunologicznych (61). Obserwowano zatem znaczenie zmian w sekwencji końca C dla procesów agregacji i tworzenia ciałek inkluzyjnych. 90% natywnego IFN $\gamma$  produkowanego przez *E. coli* JM 101 w 37°C akumuluje się w postaci inkluzji (62). Następuje to bez znaczącego spadku poziomu ekspresji, który wynosi 40-50% całkowitego białka wewnątrzkomórkowego (62). Gen interferonu gamma modyfikowano wzdłuż całej jego długości (63), dzieląc go na sześć odcinków. Pierwszy (od końca 5') — o długości 78 nukleotydów, ostatni z nich — o długości 72 nukleotydów. Najwyższy stopień agregacji osiągały te cząsteczki, które powstawały w wyniku mutacji w genie w miejscach odpowiadającym obszarom 26aa-56aa i 95aa-123aa. Mutacje polegały na wprowadzeniu w miejsce każdego z tych obszarów, fragmentów „przypadkowych” sekwencji DNA. Po sklonowaniu, transformanty *E. coli* charakteryzujące się wysoką ekspresją zmutowanego genu IFN $\gamma$ , stanowiły dla każdej z wymienionych mutacji odpowiednio 79 i 65% uzyskanych klonów. Jednocześnie obie te grupy charakteryzowały się najwyższym poziomem agregacji uzyskiwanego białka — stanowiącym odpowiednio 64 i 37%. Mutacje fragmentu genu odpowiadające pozycjom 119aa-143aa ograniczały poziom ekspresji do wartości albo istotnie niskich, albo zbliżonych do poziomu białka nie modyfikowanego (89% klonów).

Większość z opisywanych mutein (czyli białek posiadających zmodyfikowaną sekwencję aminokwasową) interferonu gamma to molekuly zawierające więcej niż jeden zmieniony aminokwas, co stwarza pewną trudność w przypisaniu właściwości agregujących tylko niektórym z nich. Niewątpliwie pozycja Pro(122) jest takim miejscem kluczowym. Zamiana Pro(122)-->Ser(122) powoduje znaczne zwiększenie rozpuszczalności, podstawienie Pro(122) -->Lys(122) wiąże się natomiast z wysokim udziałem formy inkluzyjnej uzyskiwanego białka. Wypada jednocześnie podkreślić, że na stopień rozpuszczalności białka nie wpływa sama tylko zawartość proliny.

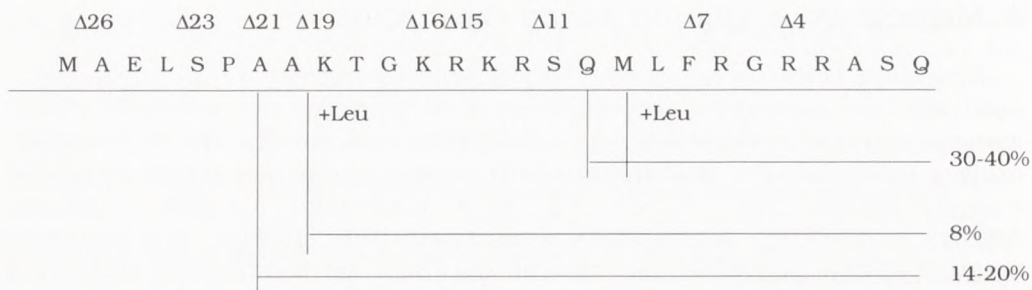


## 6. Mutacje delecyjne na końcu C-terminalnym

Struktura C-końca cząsteczki IFN $\gamma$  determinuje zarówno jego rozpuszczalność jak i — pośrednio — podatność na degradację proteolityczną. Niska wartość efektywnej ekspresji wielu mutantów delecyjnych wynika niekoniecznie z niewydolności układu translacji, często natomiast można ją wiązać z niestabilnością tych molekuł w *E. coli* (64). Obecność tetrameru Arg-Lys-Arg-Lys na końcu C prowadzi do wytwarzania IFN $\gamma$  głównie w postaci nierozpuszczalnej. Dzieje się tak zarówno dla natywnej postaci tego białka jak i dla jego mutantów delecyjnych pozbawionych do 11 aminokwasów (63). Frakcje rozpuszczalne tych białek są podatne na degradację proteolityczną. Mutanty delecyjne IFN $\gamma$  pozbawione 14-20 aminokwasów, a zatem również tego tetrameru stanowią głównie frakcję rozpuszczalną (65). Wydaje się, że niestabilność IFN $\gamma$  pozbawionego więcej niż 21 ostatnich aminokwasów, wynika ze zmian strukturalnych. Poniżej tej liczby aminokwasów delecje nie powodują zmian w strukturze, przynajmniej obserwowalnych techniką dichroizmu kołowego (63). Rola poszczególnych aminokwasów w procesie agregacji białka i utrzymywania stabilności jego molekuly w strukturze dimeru pozostaje nie do końca wyjaśniona. Istotna na przykład dla aktywności antywirusowej utrata seryny w pozycji 135 i pozostałych aminokwasów końca C-terminalnego nie wystarcza dla utrzymania dimeru (66). Dalsze delecje w pozycji 134, 133, 132 nie powodują wprawdzie zmian strukturalnych widocznych techniką jądrowego rezonansu magnetycznego (NMR), zmniejszają jednak radykalnie aktywność cząsteczki. Ewidentne zmiany następują dopiero przy większych delecjach końca C, sięgających do pozycji 133 i dalej — aż do pozycji 125. Delecje 19 – 23 aminokwasów prowadzą do zmian struktury, widocznych techniką dichroizmu kołowego, a także w testach na aktywność antywirusową.

Pomiędzy wielkością delecji a sposobem utrzymywania się białka w komórce zachodzi wyraźny związek. Przy delecji genu powodującej pozbawienie białka 7 ostatnich aminokwasów jedynie 20% IFN $\gamma$  (24  $\mu\text{g}/\text{ml}$ ) pozostaje we frakcji rozpuszczalnej. Przy delecji powodującej utratę ostatnich 16 aminokwasów udział frakcji rozpuszczalnej pozostaje na tym samym poziomie, chociaż przy wyższej całkowitej wartości wydajności ekspresji (45  $\mu\text{g}/\text{ml}$ ). Delecja nukleotydów kodujących kolejne trzy aminokwasy zwiększa stopień rozpuszczalności produktu, przy jednoczesnym utrzymaniu całkowitej jego ilości na poziomie 27  $\mu\text{g}/\text{ml}$ . Przyłączenie dodatkowego kodonu leucyny na końcu genu powoduje zwiększenie rozpuszczalności do 50% (63). Aminokwasy końca C-terminalnego mają udział również w utrzymywaniu stabilności cząsteczki. Utrata trzech N-terminalnych aminokwasów połączona z delecją jedenaście C-terminalnych wystarcza dla utraty aktywności interferonu gamma. Jest prawdopodobne, że następuje to w wyniku zmian w strukturze dimerycznej, naturalnej dla aktywnego interferonu (64). Sama utrata trzech początkowych aminokwasów nie wpływa istotnie ani na strukturę ani na zdolność do agregacji w ciała inkluzyjne. Obie znajdujące się tam cysteiny są jedynymi w całym białku i stąd wydaje się, że mostki dwusiarczkowe nie





Rys. 4. Sekwencja aminokwasowa C-końca interferonu gamma. Na rysunku pokazano przebadane mutanty delecyjne oraz odpowiadające im wartości ekspresji podane jako procent całkowitego białka komórkowego. Kolejne mutanty delecyjne opisano symbolem  $\Delta$  i liczbą usuniętych aminokwasów. Delecje mniejsze niż 11 aminokwasów nie powodują istotnych zmian w poziomie ekspresji w porównaniu z poziomem ekspresji białka natywnego. Niektóre z mutacji doprowadziły do włączenia dodatkowego aminokwasu — leucyny na końcu cząsteczki. Opis szczegółowy umieszczono w tekście.

odgrywają istotnej roli w procesie agregacji (49). Na tworzenie inkluzji wpływają w rezultacie przede wszystkim zmiany punktu izoelektrycznego (pI) wynikające z delecji. Zmiany sekwencji końca C-terminalnego prowadzą do znacznych przesunięć pI, który dla białka natywnego wynosi  $pI_{WT} = 10,4$ , a dla białka pozbawionego czternastu C-końcowych aminokwasów  $pI_{\Delta C} = 8,0$  (63). Delecje końca C-terminalnego  $IFN\gamma$ , a właściwie końca 3' jego genu w znacznym stopniu wpływają na poziom samej ekspresji. Na przykład, o ile utrata ostatnich 10-11 aminokwasów nie zmniejsza istotnie poziomu białka heterologicznego, który osiąga wartość około 30 – 40% całego białka komórkowego, to delecja 14-20, 21 aminokwasów obniża ją do mniej niż 21%, a nawet do 14% (rys. 4). Pozbawienie  $IFN\gamma$  19 aminokwasów prowadzi do 8% poziomu ekspresji (63), a dalsza redukcja — powyżej 21 aminokwasów powoduje 50-100-krotne zredukowanie ilości otrzymywanego białka (65).

## 7. Fuzje białkowe z udziałem $IFN\gamma$

Białka fuzyjne są to białka hybrydowe, na które składają się dwa lub więcej elementów o znanych właściwościach. Białka fuzyjne występują również w obu formach — zarówno rozpuszczalnej jak i ciałek inkluzyjnych. Konstrukcja białka fuzyjnego przesądza o jego postaci w roztworze. Białko fuzyjne w postaci rozpuszczalnej zachowuje swą aktywność biologiczną. Zaletą produktu precypitującego w ciałka inkluzyjne jest natomiast ochrona przed proteolizą i stosunkowa łatwość w jego oczyszczaniu.

Zasadniczo rozróżnia się dwie strategie konstruowania białek fuzyjnych: (a) białko stanowiące przedmiot analizy podłączane jest na końcu C, albo (b) na końcu N białka, którego gen stanowi część plazmidu ekspresyjnego.



Wydaje się, że sposób pierwszy ma zdecydowaną przewagę, gdy celem jest uzyskanie znacznych ilości białka o znanych właściwościach. Pierwsze białko, zapewnia efektywność układu transkrypcji i translacji, a także może zabezpieczać przed degradacją. Skonstruowano szereg białek fuzyjnych z udziałem IFN $\gamma$ . Są to hybrydy z IL-2, TNF $\alpha$ , TNF $\beta$ , IFN $\alpha$ -A, MCSF. Białka te zazwyczaj zachowują aktywność każdej ze składowych, a celem ich tworzenia bywa uzyskanie wyższej aktywności charakterystycznej dla jednej bądź drugiej ze składowych uzyskiwanego białka hybrydowego. Poszukiwania takie uzasadniał m. in. synergizm działania IFN $\gamma$  i TNF. Połączone tetrapeptydem Pro-Val-Gly-Pro tworzyły one 31kDa białko zachowujące zarówno aktywność antywirusową jak i cytotoksyczną. Białko to kodowane przez gen poprzedzony promotorem T7 jest rozpuszczalne (67). Łącznik aminokwasowy między oboma polipeptydami zapewnia elastyczność całej molekuly i może mieć też postać — Gly-Asp-Pro-Leu-Glu-Ser (68).

Białko fuzyjne IFN $\gamma$  i TNF $\beta$  (gen pod kontrolą promotora *trp*) również posiada aktywność zwiększoną w porównaniu do każdej z części składowych, przynajmniej w odniesieniu do procesów proliferacji komórek. Podobny efekt uzyskano dla fuzji IFN $\gamma$  z EGF (69). Niekiedy konstrukcja fuzji może mieć na celu modyfikację chemiczną samego hybrydu, np. fosforylację, jak w przypadku IFN $\gamma$  połączonego z HuIFN $\alpha$  (69). Stosuje się nadto konstrukcje mające na celu ułatwienie procesu oczyszczania IFN $\gamma$ , jak w przypadkach fuzji z somatostatyną (70), czy też czynnikiem uwalniania hGH (71).

Konstruuje się też fuzje białkowe w celu przemieszczenia produktu białkowego poza cytoplazmę, ponieważ upraszcza to proces oczyszczania: zawieszony w podłożu produkt można strącać unikając lizy komórek wytwarzających białko. Próby uzyskania wydzielanego do podłoża białka heterologicznego prowadzi się tworząc hybrydy z natywnymi prekursorami białek sekrecyjnych *E. coli*. O ile jednak metody takie bywały skuteczne w przypadku różnych białek, to zastosowane do wytwarzania interferonu gamma nie przyniosło pozytywnych rezultatów. Hybryd tworzony przez peptyd sygnałowy enterotoksyny *E. coli* połączonej z interferonem gamma był całkowicie degradowany w trakcie sekrecji (72). Problem wydzielania na zewnątrz *E. coli* białek heterologicznych wykracza poza ramy wyznaczone tematem pracy.

Podsumowując, w artykule podjęto próbę krytycznego przedstawienia uwarunkowań decydujących o wydajności ekspresji białek heterologicznych w *E. coli* rozpatrywanych na przykładzie IFN $\gamma$ . Omówiono czynniki wpływające na wydajność ekspresji na etapach transkrypcji, translacji i modyfikacji posttranslacyjnych, wynikające zwłaszcza z danych strukturalnych. Należy stwierdzić, że o wydajnej ekspresji decydują: silny promotor umieszczony w plazmidzie o liczbie kopii zbliżonej do około 30 kopii/komórkę i silny terminator transkrypcji. Obszar Shine-Dalgarno powinien mieć postać bliską optymalnej — tzn. sekwencję AGGAGG, a odcinek poprzedzający kodon inicjacji translacji nie powinien być krótszy niż 7-8 nukleotydów i charakteryzować się raczej zwiększonym udziałem par A+T. Cały obszar wiązania rybosomu i sekwencja flankująca SD od strony 5' powinny mieć taką postać aby na obszarze inicjacji translacji nie powstawały struktury dwuniciowe,



których usunięcie wymaga energii większej niż 10,7 kcal/mol. Takie struktury są natomiast wskazane w sekwencjach poprzedzających RBS. Dla wydajności ekspresji ważne też jest aby początek sekwencji kodującej był bogaty w pary A+T. Struktura aminokwasowa powinna natomiast spełniać warunki wynikające z reguł ograniczających degradację proteolityczną.

Za merytoryczne uwagi, krytyczną ocenę i moralne wsparcie dziękuję Panom Profesorom: Leonowi Sedlaczkowi, Adamowi Jaworskiemu i Wiesławowi Kacy, bez życzliwego wsparcia których praca ta nie mogłaby powstać.

## Literatura

1. Grajek W., (1995), *Biotechnologia*, 1 (28), 91-102.
2. Goeddel D., (1990), *Methods in Enzymology*, 185, 3-7.
3. Varshavsky A., (1992), *Cell*, 69, 725-735.
4. Shrader T.E., Tobias J.W., Varshavsky A., (1993), *J.Bacteriol.*, 175(14), 4364-4374.
5. de Smit M.H., van Duin J., (1990), *Proc. Natl. Acad. Sci.USA*, 87, 7668-7672.
6. Wada Ken-Nosuke, Wada Yoshiko, Doi Hirofumi, Ishibashi Fumia, Gojobori Takashi, Ikemura Toshimichi, (1991), *Nucl. Acid Res.*, 19(suppl.), 1981-1985.
7. Robinson M., Lilley R., Little S., Emtage J.S., Yarranton G., Stephens P., Millican A., Eaton M., Humphreys G., (1984), *Nucl. Acid Res.*, 12(17), 6663-6671.
8. Nakamura T., Suyama A., Wada A., (1991), *FEBS Letters*, 289(1), 123-125.
9. Faxen M., Plumbridge J., Isaksson L.A., (1991), *Nucl. Acid Res.*, 19(19), 5247-5251.
10. Balbas P., Bolivar F., (1990), *Methods in Enzymology*, 185, 14-37.
11. Shi-Hsiang Shen, (1984), *Proc. Natl. Acad. Sci. USA*, 81, 4627-4631.
12. Gray P.W., Goeddel D., (1982), *Nature*, 298, 859-863.
13. Tanaka S., Oshima T., Ohsuye K., Ono Teiichi, Mizono A., Ueno A., Nakazato H., Tsujimoto M., Higashi N., Noguchi T., (1983), *Nucl. Acid Res.*, 1(6), 1707-1723.
14. Jay E., MacKnight D., Lutze-Wallace C., Harrison D., Wishart P., Whi-Yu Liu, Asundi V., Pomeroy-Cloney L., Rommens J., Englington L., Pawlak J., Jay F., (1984), *J.Biol.Chem.*, 259 (10), 6311-6317.
15. Shirai A., Tanouchi M., Todd Ch.W., Wallace R.B., (1987), *Gene*, 57, 11-19.
16. Jay E., Rommens J., Pomeroy-Cloney L., MacKnight D., Lutze-Wallace C., Wishart P., Harrison D., Whi-Yu Liu, Dawood M., Jay F., (1984), *Proc. Natl. Acad. Sci.USA*, 81, 2290-2294.
17. Perez L., Vega J., Chuay C., Menendez A., Ubieta R., Montero M., (1990), *Appl. Microbiol. Biotechnol.*, 33(4), 429-434.
18. Podhajska A., Hasan N., Szybalski W., (1985), *Gene*, 40(1), 163-168.
19. Wulfig C., Pluckthun A., (1993), *Gene*, 136(1-2), 199-203.
20. Remaut E., Marmenout A., Simons G., Fiers W., (1987), *Methods in Enzymology*, 153, 416-431.
21. Bujard H., Gentz R., Lanzer M., Steuber D., Mueller M., Ibrahim I., Haeuptle M.T., Dobberstein B., (1987), *Methods in Enzymology*, 155, 417-433.
22. Newbury S., Smith N.H., Robinson E.C., Hiles I.D., Higgins C.F., (1987), *Cell*, 48, 297-310.
23. Bouvet P., Belasco J.G., (1992), *Nature*, 360 (6403), 488-491.
24. Helke A., Geisen R.M., Vollmer M., Sprengart M.L, Fuchs E., (1993), *Nucl. Acid Res.*, 21(24), 5705 - 5711.
25. de Smit M.H., van Duin J., (1990), *Proc. Natl. Acad. Sci.USA*, 87, 7668-7672.
26. Nishi T., Itoh S., (1986), *Gene*, 44(1), 29-36.



27. Hsiung M.H., MacKellar W.C., (1987), *Methods in Enzymology*, 153, 390-401.
28. Gross G., Mielke C., Hollatz I., Blocker H., Frank R., (1990), *J. Biol. Chem.*, 265 (29), 17627-17636.
29. Chen H., Pomeroy-Cloney L., Bjerkens M., Tam J., Jay E., (1994), *J. Mol. Biol.*, 240(1), 20-27.
30. Mertens N., Remaut E., Fiers W., (1995), *Gene*, 164, 9-15.
31. Gold L., Stormo G.D., (1990), *Methods in Enzymology*, 185, 89-93.
32. de Boer H., Hui A.S., (1990), *Methods in Enzymology*, 185, 103-114.
33. de Smit M.H., van Duin J., (1990), *Proc. Natl. Acad. Sci. USA*, 87, 7668-7672.
34. Mateucci M.D., Heyneker H.L., (1983), *Nucl. Acid Res.*, 11(10), 3113-3121.
35. Degryse E., (1990), *FEBS Lett.*, 269(1), 244-246.
36. Sverdlov E.D., Tsarev S.A., Krykbaev R.A., Chernov I.P., Rostapshov V.M., (1987), *FEBS Lett.*, 212(2), 233-236.
37. Keilman C., Wanner G., Bock A., (1993), *Biol. Chem. Hoppe-Seyler* 374(10), 938-992.
38. Tessier L.H., Sondermeyer P., Faure T., Dreyer D., Bonavente A., Villeval D., Courtney M., Lecocq J.P., (1984), *Nucl. Acid Res.*, 12(20), 7663-7675.
39. Wood C.R., Boss M.A., Patel T.P., Emtage J.S., (1984), *Nucl. Acid Res.*, 12(9), 3937-3950.
40. Bucheler U.S., Werner D., Schrimmer R.H., (1992), *Nucl. Acid Res.*, 20(12), 3127-3133.
41. Olsthoorn R.C., Lics N., van Duin J., (1994), *EMBO J.*, 13, (11), 2660-2668.
42. Sprengart M.L., Fatscher H.P., Fuchs E., (1990), *Nucl. Acid Res.*, 18 (7), 1719-1723.
43. Olsthoorn R.C.L., Zoong S., van Duin J., (1995), *Mol. Microbiol.*, 15(2), 333-339.
44. Alexciev K., Uscheva A., Pavlova M., Yavachev L., Ivanov I., (1987), *Int. J. Biochem.*, 21(9), 987-996.
45. Chalut C.H., Egly J.M., (1995), *Gene*, 156, 43-45.
46. Leung D. za: de Boer H., Hui A., (1990), *Methods in Enzymology*, 185, 109.
47. Ivanov I.G., Alexandrova B.A., Dragulev B.P., Abou Hadair M.G., (1995), *Gene*, 160, 75-79.
48. Morelle G., Frank R., Meyerhans A., (1991), *Biochem. Biophys. Acta*, 1089(3), 320-324.
49. Lee S.G., Ricca G.A., Crumley G., Lloyd R.S., Drohan W., (1988), *Biochem. Biophys. Res. Com.*, 151(1), 598-607.
50. Smit M.H., van Duin J., (1993), *Mol. Microbiol.*, 9(5), 1079-1088.
51. Spanjaard R.A., van Dijk M.C., Turion A.J., van Duin J., (1989), *Gene* 80(2), 345-351.
52. Schoener B., Belagaje R.M., Schoener R.G., (1990), *Methods in Enzymology*, 153, 401-416.
53. Makoff A.J., Smallwood A.E., (1990), *Nucl. Acid Res.*, 18(7), 1711-1718.
54. Lee N., Zhang S.Q., Cozzitorto J., Yang J.S., Testa D., (1987), *Gene*, 58(1), 77-86.
55. Seno M., Hinuma S., Onda H., Igarashi K., (1986), *FEBS Lett.*, 199(2), 187-192.
56. Schauder B., McCarthy J.E., (1989), *Gene*, 78(1), 59-72
57. Looman A.C., Bodlaender J., de Gruyter M., Vogelaar A., van Knippenberg P.H., (1986), *Nucl. Acid Res.*, 14(13), 5481-5497.
58. Schoner B.E., Hsiung H.M., Belagaje R.M., Mayne N.G., Schoner R.G., (1984), *Proc. Natl. Acad. Sci.*, 81, 5403-5407.
59. Brink M.F., Verbeet M.Ph., de Boer H.A., (1995), *Gene*, 156, 215-222.
60. Bowden G.A., Paredes A.M., Georgiou G., (1991), *Bio/technology*, 9, 725-730.
61. Wetzel R., Perry L.J., Veilleux C., (1991), *Prot. Eng.*, 3(4), 301.
62. Wetzel R., Perry L.J., Veilleux C., (1991), *Bio/technology*, 9, 731-737.
63. Slodowski O., Boehm J., Schoene B., Otto B., (1991), *Eur. J. Biochem.*, 202, 1133-1140.
64. S.Ka-Shing Luk, Jay E., Jay F.J., (1990), *Biol. Chem.*, 11314-13319.
65. Schoene B., Boehm J., Roggenthien J., Slodowski O., Otto B., (1989), *J. Interferon Res.*, 9 Suppl., 2, S180.
66. Lundell D., Lunn C., Dalgarno D., Fossetta J., Greenberg R., Reim R., Grace M., Narula S., (1991), *Prot. Eng.*, 4(3), 335-341.
67. Korobko V.G., Davydov I.V., Shinogradova L.N., Filipov S.A., Esipov D.S., Berkova N.P., (1991), *Bioorg. Khim.*, 17(2), 189-196.
68. Hou Y., Zhang D., Wu S., Zhi G.J., (1989), *J. Interferon. Res.*, 9 Suppl., 2, S149.



69. Zhao X.X., Li B.-L., Lenger J.A., van Ripper G., Pestka S., (1989), *Anal. Biochem.*, 178(2), 342-347.
70. Notani J., Saitoh Y., Terada M., Yamada H., Ishii Y., Niwa M., (1989), *Biotechnol. Prod.*, 5(1), 40-55.
71. Basu M., Dharm E., Levine J.F., Kramer R.A., Crowl R.M., Campbell R.M., (1991), *Arch. Biochem. Biophys.*, 286(2), 638-634.
72. Gentz R., Kuys Y., Zwieb C., Taatjes D., Taatjes H., Bannwarth W., Steuber D., Ibrahimi I., (1990), *J. Bacteriol.*, 170(5), 2212-2220

## Heterologous genes expression in *E. coli* ; interferon gamma as the model system

### Summary

Interferon gamma is a molecule which may serve as a good object for studying and explaining the role of expression limiting conditions.

Three levels of limiting factors modifications, i. e. operon choice and structure, changes in the secondary mRNA structure, and the influence of post-translational modulation processes are described. The relationship between the operon structure and expression level for other heterologous proteins is compared with that of IFN-gamma. In particular, the shape of ribosome binding site (RBS) region and its affected different nucleotide substitutions are presented.

The rule of N-terminal and C-terminal end changes for protein stability and inclusion body formation is also shown.

### Key words:

*Escherichia coli*, heterologous expression, mRNA structure, translation initiation, ribosome binding site, inclusion, IFN $\gamma$ .

### Adres do korespondencji:

Tomasz Cieřlikowski, Centrum Mikrobiologii i Wirusologii PAN, ul. Lodo-  
wa 106, 93-232 Łódź.