

Przemysław Nuc<sup>1,✉</sup>

Katarzyna Nuc<sup>2</sup>

<sup>1</sup>Uniwersytet im. Adama Mickiewicza, Instytut Biologii Molekularnej i Biotechnologii, Poznań

<sup>2</sup>Akademia Rolnicza im. Augusta Cieszkowskiego, Katedra Biochemii i Biotechnologii, Poznań

✉Uniwersytet im. Adama Mickiewicza, Instytut Biologii Molekularnej i Biotechnologii, ul. Umultowska 89, 61-614 Poznań; e-mail: przem@amu.edu.pl, tel.: (061) 829 59 06

Artykuł otrzymano 26 września 2006 r.

**Słowa kluczowe:** *Escherichia*, nadprodukcja, system, wektor

**Wykaz skrótów:** cAMP – cykliczny adenozynomonofosforan; f-Met – formylometionina; GFP – białko zielonej fluorescencji; IPTG – izopropyl-β-D-1-tiogalaktopiranozyd; IF2 – bakteryjny czynnik inicjujący 2; MBP – białko wiążące maltozę; DnaK – białko szoku termicznego z rodziny Hsp 70

Artykuł sponsorowany przez firmę MERCK Sp. z o.o.



## STRESZCZENIE

Rosnące zapotrzebowanie na wydajną produkcję rekombinowanych białek stawia liczne wyzwania; poczynając od optymalizacji hodowli komórek w warunkach nadekspresji, dostosowania sekwencji wektora, genu i kodowanego białka do systemu biosyntezy białka gospodarza, poprzez poznawanie mechanizmów zwijania białek, konstrukcję białek fuzyjnych i układów koekspresyjnych, po konieczność ulepszania metod oczyszczania i renaturacji białek. Obecnie *Escherichia coli* jest najbardziej zdefiniowanym i najtańszym systemem biosyntezy białka. Z bogactwem dostępnych i wypróbowanych mutantów najlepiej się nadaje do ekonomicznego testowania nowych konstrukcji genowych i podejmowania prób zwiększania skali produkcji rekombinowanych białek.

## WPROWADZENIE

Systemy ekspresji rekombinowanych białek w komórkach pro- i eukariontów opracowuje się by tanio, szybko i bezpiecznie produkować duże ilości białka lub enzymu, który można łatwo oczyścić. Oprócz systemów, których gospodarzem jest *E. coli*, do najpopularniejszych należą: *Bacillus subtilis* oraz *Saccharomyces cerevisiae* i *Pichia pastoris* [1]. Nadprodukcja białek w komórkach ssaków i owadów ogranicza się do skali laboratoryjnej ze względu na drogie podłoża do hodowli. Zwierzęta i rośliny transgeniczne znajdują zastosowanie w produkcji rekombinowanych białek na dużą skalę, lecz początkowy koszt uzyskania, w przypadku zwierząt, a w przypadku roślin spełnienia wymogów bezpieczeństwa, związanych z nierozprzestrzenianiem może być wysoki [2]. Rekombinowane białka warto zatem produkować w systemach eukariotycznych wtedy, gdy nie można ich otrzymać w systemach ekspresyjnych *E. coli* ani w układach drożdżowych.

Systemy ekspresyjne *E. coli* ze względu na dobrze poznaną sekwencję genu, najlepiej zdefiniowany układ transkrypcyjny i translacyjny, duży wybór opisanych promotorów, replikonów oraz mutantów, łatwość ich samodzielnego przygotowania, a także możliwość prowadzenia hodowli w tanich pożywkach i z wykorzystaniem tak prostych czynników selekcyjnych jak antybiotyki, są najczęściej stosowane do nadprodukcji rekombinowanych białek. Trudności nadal sprawiają białka bardzo duże, zawierające liczne mostki dwusiarczkowe, liczne niesparowane grupy tiolowe oraz takie, które wymagają modyfikacji potranslacyjnych. W komórkach *E. coli* z powodzeniem produkuje się interferon, albuminę człowieka, hormony wzrostu i prawie wszystkie rekombinowane enzymy dla diagnostyki i biologii molekularnej [3].

## WARUNKI HODOWLI *ESCHERICHIA COLI*

Od warunków w jakich prowadzi się hodowlę *E. coli* zależy to jaką ilość rekombinowanego białka można uzyskać, w jaki sposób i jakiej jakości produkt będzie można oczyścić. Ponieważ nadprodukcja białek w formie ciał inkluzyjnych, wymaga ich denaturacji i ponownego zwijania, co bywa skomplikowane, a na większą skalę zbyt kosztowne, dąży się do uzyskania produktu w formie rozpuszczalnej. Po doborze warunków hodowli ważne jest również coraz lepsze poznawanie mechanizmów zwijania białek, roli białek opiekuńczych i izomeraz w tym procesie, a także mechanizmów sekrecji, co pozwala na wykorzystanie przydatnych mutantów i zastosowanie koekspresji białek systemu sekrecyjnego. Po opracowaniu systemu wydajnego w skali laboratoryjnej pojawia się zadanie zwiększenia skali produkcji. Tylko na rynku amerykańskim co roku sprzedaje się około 100 ton wołowego hormonu wzrostu, a w przypadku produkcji białka o zastosowaniach terapeutycznych produkcję trzeba rozpocząć od kilkudziesięciu kilogramów rocznie [3,4]. Proces zwiększania skali nie jest prosty i wymaga dokładnego skalkulowania opłacalności, a to często wymusza prowadzenie hodowli o dużej gęstości (HCDC – ang: *high cell density cultures*), przekraczającej 50g suchej masy komórek na litr [g(DCW)l<sup>-1</sup>]

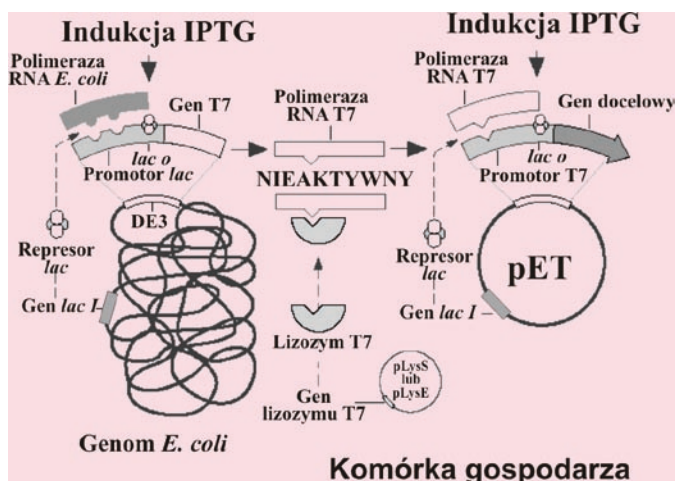
(ang: *dry cell weight*). Żeby planować zwiększanie skali trzeba bardzo dobrze poznać fizjologiczne uwarunkowania dynamiki wzrostu i metabolizmu *E. coli* w warunkach nadprodukcji określonego białka. Mimo ograniczeń, które trzeba przezwyciężyć, takich jak gorsza wymiana gazowa i cieplna, inhibicja substratowa i gromadzenie się substancji hamujących wzrost (koszt dializy bywa za wysoki), takie podejście często się w końcu opłaca. Niektóre składniki pożywek hamują wzrost komórek *E. coli* kiedy występują w zbyt wysokich stężeniach, stąd proste zwiększenie stężenia składników pokarmowych nie daje dobrych efektów. Przykłady granicznych stężeń niektórych składników pożywek: glukoza (50 g/l), amoniak (3 g/l), żelazo (1,15 g/l), magnez (8,7 g/l), fosfor (10 g/l), cynk (0,038 g/l) [4]. Nie jest proste odniesienie warunków produkcji białka w kolbie do fermentora, gdzie w warunkach hodowli o dużej gęstości trzeba uzupełniać składniki pożywki, kontrolować pH, natlenienie i wymianę ciepłą, a jeśli to konieczne, usuwać powstające substancje hamujące wzrost. Jako uzupełniające źródło azotu można stosować np. wodorotlenek amonu by jednocześnie kontrolować pH pożywki. Największe gęstości *E. coli* jakie udało się uzyskać dla szczepów B to 174 g(DCWL)<sup>-1</sup>, w warunkach dializowanej hodowli, gdzie źródłem węgla był glicerol. Zwiększanie gęstości hodowli *E. coli* powyżej 200 g(DCW)L<sup>-1</sup> nie ma już sensu. Przy takich gęstościach następuje gwałtowny wzrost lepkości hodowli. Przy 400 g(DCW)L<sup>-1</sup>, jeśli przyjąć, że średnia cylindryczna komórka ma 3 μm długości i 1 μm średnicy, pożywka stanowi już tylko 25% objętości hodowli. Jednym z największych problemów związanych z kulturami *E. coli* o dużej gęstości jest produkcja octanu, która pojawia się zwykle w warunkach ograniczonego natlenienia lub nadmiaru glukozy. Nadmiar glukozy powoduje, że bakterie produkują octan nawet w warunkach tlenowych na skutek wysycenia szlaku kwasów trikarboksylowych. Stężenie octanu przekraczające 5 g/l w pH = 7,0 już hamuje wzrost komórek *E. coli*. Uprotonowany octan blokuje produkcję ATP zmniejszając udział ΔpH w sile protonomotorycznej. To skutkuje zahamowaniem replikacji, transkrypcji i translacji. Octan powstaje także kiedy tempo wzrostu jest za wysokie – najczęściej rzędu 0,2–0,35 h<sup>-1</sup>, ale zależy to od składu pożywki i od szczepu. Problem hamowania biosyntezy białka przez nadmierną produkcję octanu rozwiązuje się przez ograniczenie źródeł C i N [4,5]. Kiedy źródłem węgla jest glicerol, octan zwykle nie powstaje, prawdopodobnie z powodu powolnego pobierania glicerolu i spowolnienia przepływu węgla przez szlak glikolizy. Czasami wystarczy obniżenie temperatury, co zmniejsza tempo wzrostu i zapotrzebowanie na tlen, a jednocześnie ogranicza powstawanie ciał inkluzyjnych. W niektórych przypadkach, zależnie od nadprodukowanego białka, warto pożywkę uzupełnić odpowiednimi aminokwasami – np. glicyną czy metioniną. Mutacje genów *Ack* (kinazy octanowej) lub *Pta* (fosfotransacetylazy acetylo-CoA), które zaangażowane są w uwalnianie octanu z acetylo-CoA również pozwalają na ograniczenie akumulacji octanu. Mutanta *pta* wykorzystano z powodzeniem w warunkach gęstej hodowli do nadprodukcji ludzkiej interleukiny 2 i lipazy. Można też ograniczyć pobieranie glukozy stosując jej analog – nietoksyczny kompetycyjny inhibitor pobierania glukozy – metylo-α-glukozyd. Czynnikiem ograniczającym wzrost w HDCD bywa także stężenie tlenu. W stanie na-

syconym ilość tlenu rozpuszczonego w wodzie o temperaturze 25°C wynosi 7 mg/litr. Jego ilość w hodowli należy zwiększyć – przez napowietrzanie (zastosowanie czystego tlenu nie jest tanie), szybkie mieszanie pożywki. Czasami stosuje się hodowlę pod zwiększonym ciśnieniem lub koekspresję hemoglobiny. Dobre mieszanie pożywki zapewnia jednorodność produktu – wszystkie komórki powinny mieć jednakowy dostęp do tlenu i do uzupełnianych składników pożywki. Stężenie CO<sub>2</sub> ma wpływ na tempo wzrostu HDCD, ponieważ jego cząstkowe ciśnienie większe niż 0,3 atm zmniejsza tempo wzrostu *E. coli* i powoduje gromadzenie się octanu. Najbardziej kontrolowane warunki wzrostu uzyskuje się stosując wykładnicze uzupełnianie pożywki, dostosowane do logarytmicznego tempa podziałów komórek. W hodowlach o dużych gęstościach problemem staje się nawet metaboliczna produkcja ciepła [4]. Trzeba pamiętać, że już wydajna replikacja plazmidu w komórce jest stresem, a nadprodukcja rekombinowanego białka wprowadza „molekularne zamieszanie”, które objawia się zmianą tempa podziałów, wzrostu i metabolizmu komórki w wyniku gwałtownej utraty energii i zapasu substancji niskocząsteczkowych. Stężenie białka w komórce *E. coli* waha się od 200–300 mg/ml. Nadmierna produkcja heterologicznego białka często kończy się powstawaniem ciał inkluzyjnych lub (i) jego częściową degradacją [6].

## WYBÓR PROMOTORA

Promotor powinien być indukowany i pozwalać na wydajną produkcję rekombinowanego białka, tak by jego ilość w krótkim czasie mogła osiągać 10–30% całkowitego białka komórki. Podstawowa ekspresja z takiego promotora (bez indukcji) powinna być jak najmniejsza. Powinien być łatwo indukowany – fizycznie lub chemicznie, z wykorzystaniem prostego induktora, jak L-arabinoza lub IPTG. Ponieważ IPTG jest toksyczny, nie zawsze można go stosować, np. do produkcji białek o zastosowaniach leczniczych. Indukcja termiczna jest prosta, ale szczep powinien mieć zmutowany gen *rpoH*, inaczej przy zmianie temperatury zostanie wzbudzona ekspresja białek opiekuńczych i proteaz. Najpopularniejsze systemy do ekspresji białek w *E. coli* wykorzystują specyficzną transkrypcję z udziałem polimerazy RNA faga T7, której gen pod kontrolą promotora *lac* można indukować za pomocą IPTG. Najczęściej wykorzystywane są plazmidy pET (Novagen) – aktualnie ponad 40 wersji. W roku 2003 ponad 90% struktur zdeponowanych w PDB (Protein Data Bank) było wynikiem nadprodukcji białek w tym systemie [7].

**Ekspresja w systemie pET** najczęściej polega na tym, że genom gospodarza ma wbudowany fragment genomu faga (ΔE3), zawierający gen polimerazy RNA faga T7, pod kontrolą promotora *lacUV5*. Chociaż promotor *lacUV5* jest mniej wrażliwy na regulację przez cAMP-CRP niż promotor *lac*, dodatek glukozy do pożywki w ilości 1% ogranicza poziom cAMP i wzmacnia jego represję. Białko LacI lub jego 10 razy bardziej skuteczny mutant LacI<sup>q</sup> jest represorem promotora *lacUV5* i hybrydowego promotora *T7 / lac*, znajdującego się na plazmidzie ekspresyjnym. Gen *lacI* naturalnie występuje na chromosomie *E. coli*, a w niektórych systemach także na plazmidzie ekspresyjnym – zawierają go wektory pET i ich pochodne. Kiedy IPTG wiąże się z LacI, tetrameryczny



**Rycina 1.** Indukcja ekspresji genu z promotorem T7 za pomocą IPTG w systemie pET (Novagen).

repressor uwalnia promotor polimerazy RNA T7 i hybrydowy promotor na plazmidzie ekspresyjnym, co uruchamia transkrypcję i translację polimerazy T7. Powstała polimeraza RNA T7 rozpoznaje promotor T7 – unikalną 20-nukleotydową sekwencję (TAATACGACTCACTATAGGG) plazmidu ekspresyjnego i rozpoczyna transkrypcję z szybkością osiągającą 230 nt/s – prawie pięciokrotnie szybciej niż polimeraza RNA *E. coli* (50 nt s<sup>-1</sup>). Śladowy (podstawowy) poziom ekspresji polimerazy RNA T7 jest dodatkowo eliminowany przez koekspresję lizozymu T7 – naturalnego inhibitora polimerazy RNA T7 z plazmidu pLysS (*silent*) lub pLysE (*expressed*). Ramka odczytu lizozymu na tych dwóch plazmidach ma różną orientację – zgodną (pLysE) lub odwrotną (pLysS) w stosunku do promotora *Tc* – indukowanego przez tetracyklinę [7] (Ryc. 1).

## WYBÓR SZCZEPU *ESCHERICHIA COLI*

Do nadprodukcji białek w systemach zależnych od polimerazy RNA faga T7 najczęściej wybieramy odpowiednie szczepy typu (DE3). Szczepy *trxB/gor* – zapewniają środowisko dla powstawania mostków dwusiarczkowych w cytoplazmie, np. szczep Origami (Novagen). Szczep *E. coli* Turner (Novagen) ma uszkodzony gen permeazy laktozy – *lacY*, co pozwala na płynną regulację ekspresji rekombinowanych białek w tym systemie stężeniem IPTG. Jest to przydatne gdy obniżenie poziomu produkcji rekombinowanego białka pozwala uzyskać produkt o lepszej rozpuszczalności i aktywności. Szczepy Rosetta i Rosetta-gami (Novagen) z plazmidami pLysRARE pozwalają na ekspresję heterologicznych białek, których ramki odczytu zawierają rzadko wykorzystywane u *E. coli* kodony, np. AGA.

Szczepy C41(DE3) i C43(DE3) to mutanty przeznaczone do ekspresji białek błonowych, które w innych szczepach tworzą ciała inkluzyjne. W warunkach nadprodukcji białka dochodzi raczej do proliferacji błon, aniżeli do powstawania ciał inkluzyjnych. Zwiększoną stabilność mRNA zapewniają szczepy BL21, posiadające mutację *rne131*.

W przypadku białek silnie toksycznych, gen polimerazy T7 wprowadza się infekując szczepy, które nie posiadają

genu polimerazy RNA T7 (np. BL21, BLR) bakteriofagiem λCE6. BLR to pochodna BL21 z uszkodzonym genem *recA* co pozwala na utrzymanie i stabilną ekspresję genu z sekwencjami powtórzonymi, a B834 ma uszkodzone geny proteaz *lon* i *ompT*. Infekcja hodowli fagiem λCE6 wprowadza konstytutywnie ekspresjonowany gen polimerazy RNA faga T7. Nie dochodzi do lizy komórek, a w ciągu 2-3 godzin można uzyskać do 200 mg rekombinowanego białka z 1 litra hodowli [8].

## REPLIKON

Gen służący do nadprodukcji rekombinowanego białka najczęściej włącza się do plazmidu o niskiej liczbie kopii, zwykle zawierający replikon ColE1 lub p15A, aby jego replikacja nie obciążała nadmiernie metabolizmu komórki.

Ścisła kontrola liczby kopii plazmidu ekspresyjnego pozwala także na ograniczenie śladowej ekspresji wprowadzonego genu przed indukcją. Plazmid ekspresyjny pET-coco (Novagen), dzięki zastosowaniu elementów replikonu episomu F – *oriS* i *repF*, a także elementów *cis parABC*, zapewniających rozdział plazmidów między komórki potomne, może być przed indukcją utrzymywany w liczbie 1 kopii na komórkę w pożywce LB zawierającej 0,2% D-glukozy. Przed indukcją ekspresji wprowadzonego genu można zwiększyć liczbę kopii plazmidu przez dodanie do pożywki L-arabinozy (0,01%), co powoduje, że replikator TrfA uruchamia miejsce początku replikacji *oriV* i liczba kopii plazmidu w komórce zwiększa się do 20-50. Zastosowanie takiego wektora w szczepie Turner pozwala na testowanie bardzo szerokiego zakresu poziomu ekspresji wprowadzonego genu [9].

Otrzymanie wieloenzymatycznego kompleksu czy badanie oddziaływań białko-białko może wymagać koekspresji dwóch białek, podobnie, uzyskanie aktywnego, prawidłowo zwiniętego białka może wymagać jego koekspresji z białkiem opiekuńczym. Ramki odczytu umieszcza się albo na jednym wektorze typu Duet (Novagen) (Tab. 1), albo na kilku, których replikony należą do różnych grup niezgodności (Tab. 2). Dzięki możliwości utrzymania w komórce *E. coli* jednocześnie 4 z pięciu wektorów typu Duet możliwa jest koekspresja nawet ośmiu białek [10].

**Tabela 1.** Replikony plazmidów w systemach ekspresyjnych *E. coli*, na podstawie katalogu Novagen 2006/2007.

Plazmid(y)	Replikon (pochodzenie)	Liczba kopii
pET (all) pETDuetTM-1	ColE1 (pBR322)	~40
pACYCDuetTM-1 pLysS pLysE pLacI pRARE	P15A (pACYC184)	10-12
pRSF (all)	RSF1030	> 100
pCDF (all)	CloDF13	20-40
pETBlueTM (all) pTriExTM (all)	ColE1 (pUC)	> 500
pETcoco (all)	Mini-F/RK2 (2) (pBeloBAC11, RK2)	1, amplifikowalny do ~40
pCOLADuetTM-1	ColA	~20-40



Kiedy warunki nadprodukcji i sposób oczyszczania rekombinowanego białka są już dopracowane, warto pomyśleć o trwałym zabezpieczeniu hodowli, której skala ma być powiększona przed pojawianiem się segregantów nie posiadających wprowadzonego genu. Jednym ze sposobów jest zbudowanie takiej konstrukcji genetycznej *cis* (*locus parABC* plazmidu R1, j.w. lub *ccdAB* plazmidu F) lub *trans* (komplementacja uszkodzenia niezbędnego genu na chromosomie supresorem kodowanym przez plazmid), w której utrata plazmidu spowoduje śmierć komórki, a innym – ustabilizowanie szczepu przez wbudowanie jednej lub kilku kopii genu do chromosomu [7].

## STABILNOŚĆ I SEKWENCJA NUKLEOTYDOWA mRNA

Okres półtrwania mRNA w komórkach *E. coli* wynosi od kilku sekund do 20 minut. Wydajne wiązanie rybosomów zwiększa stabilność mRNA, a zmniejsza występowanie struktur drugorzędowych i końca poli-A. W komórkach *E. coli* za poliadenylację odpowiadają dwie poliadenylujące polimerazy RNA: PAPI i PAPII, a za degradację, przede wszystkim dwie egzonukleazy: RNaza II i PNPaza i endoribonukleaza RNaza E. Poliadenylacja ułatwia degradację mRNA z udziałem RNazy II i PNPazy. Koniec 5' mRNA nie powinien zawierać sekwencji tworzących struktury spinki do włosów, co może utrudnić wiązanie rybosomu i przyspieszyć jego degradację. Chociaż udaje się stabilizować koniec 5' mRNA przez wprowadzenie struktury spinki do włosów (przykład *ompA*), nie może ona zawierać niesparowanych nukleotydów po stronie 5'. Strukturę spinki do włosów z pewnością warto umieścić w rejonie 3'-UTR. Na obydwu końcach mRNA można umieścić dodatkowe sekwencje stabilizujące – np. *ompA* na końcu 5' lub GFP na końcu 3'. Nie jest jednak znana uniwersalna sekwencja lub struktura zwiększająca stabilność wszystkich mRNA [6,7].

W rejonie inicjacji translacji musi się znaleźć miejsce wiązania rybosomu zawierające sekwencję Shine-Dalgarno (SD: UAAGGAGG) położone 7±2 nt powyżej kodonu AUG. Jeśli ma dojść do wydajnej inicjacji translacji rejon ten musi być bogaty w reszty A i T, a pierwszym nukleotydem za kodonem AUG powinna być adenina. Jeśli sekwencja SD wchodzi w skład struktur II-rzędowych na końcu 5' mRNA, silnie zmniejsza się wydajność inicjacji translacji. Zmniejszenie odległości między SD a kodonem AUG do 4 nt lub zwiększenie powyżej 14 nt także znacznie osłabia wydajność inicjacji translacji. Należy pamiętać, że około 8% miejsc inicjacji translacji *E. coli* to kodony GUG, UUG lub AUU [6,7].

Wydajność inicjacji translacji może istotnie wzmocnić sekwencja DB (ang: *downstream box*) położona tuż poniżej kodonu AUG, jeśli jej konsensus: 5'-AUGAAUCACAAAGUG-3' zawiera zasady komplementarne do zasad 1469-1483 16S rRNA. Wzmacniający efekt tej sekwencji jest zależny od właściwego działania sekwencji SD i utrzymuje się w zakresie przesunięć początku DB od pozycji +1 do +6 sekwencji kodującej [6,7,11].

Poniżej sekwencji kodującej powinno się znajdować miejsce terminacji transkrypcji (palindromowa sekwencja tworząca strukturę spinki do włosów), co poprawia stabilność mRNA. Najlepiej jeśli to są dwa kolejne kodony stop zaczynające się tetranukleotydem UAAU ponieważ u *E. coli* najskuteczniejszym kodonem stop jest kodon UAA, zwłaszcza kiedy występuje po nim reszta U. Częstość wykorzystania poszczególnych kodonów w komórce *E. coli* dokładnie odpowiada poziomowi poszczególnych aminoacylo-tRNA w cytozolu [12]. Jeśli zatem ramka odczytu zawiera liczne kodony, które rzadko występują w sekwencjach kodujących genów *E. coli*, jej ekspresja obniża się, ponieważ dochodzi do substytucji (np. arginina jest zastępowana lizyną), przesunięć ramki odczytu, a nawet przeskoku dwóch kolejnych kodonów argininowych AGA albo przedwczesnej terminacji translacji. Ramka odczytu nie powinna też zawierać kolejno po sobie występujących kodonów lizyny AAG, odczytywanych przez tRNA<sup>Lys</sup>UUU [13,14].

Dane dotyczące częstości wykorzystywania poszczególnych kodonów u różnych organizmów można znaleźć na stronie Codon Usage Database: [www.kazusa.or.jp/codon/](http://www.kazusa.or.jp/codon/). Są to dane statystyczne, których istotność zależy od liczby zsekwencionowanych genów organizmu i właściwego określenia sekwencji kodujących.

## STABILNOŚĆ I STRUKTURA PRODUKTU TRANSLACJI

Na stabilność rekombinowanych białek znaczny wpływ ma reszta aminokwasowa występująca bezpośrednio po metioninie (w komórkach *E. coli* – fMet). Od niej zależy wydajność usuwania formylometioniny. W komórkach *E. coli*, takie cięcie jest katalizowane przez aminopeptydazę metioninową. Jeżeli zaraz po f-Met występuje reszta His, Gln, Glu, Phe, Met, Lys, Tyr, Trp lub Arg, cięcie następuje z wydajnością rzędu 16%. W pozostałych przypadkach formylometionina jest odcinana z wydajnością 97%. Jeżeli po odcięciu f-Met pierwszą resztą jest: Arg, Lys, Phe, Leu, Trp albo Tyr, okres półtrwania białka w komórce *E. coli* wynosi zaledwie 2 minuty, podczas gdy w pozostałych przypadkach dochodzi do 10 godzin. Najgorsza sytuacja ma miejsce kiedy za metioniną występuje reszta leucyny, wtedy istnieje wysokie prawdopodobieństwo odcięcia formylometioniny (powyżej), a powstały polipeptyd posiadający na aminowym końcu leucynę, ulega szybkiej proteolizie. By uniknąć takiej sytuacji, należy wziąć pod uwagę możliwość zmiany sekwencji aminokwasowej końca N lub umieszczenie dodatkowej domeny przed Met-Leu [14].

Jeśli produkt translacji rekombinowanego białka ma oryginalny koniec karboksylowy należy zwrócić uwagę by jego sekwencja aminokwasowa nie była zbyt podobna do peptydu ANDENYALAA, syntetyzowanego w komórkach *E. coli* na końcach przedwcześnie terminowanych łańcuchów polipeptydowych, powstałych w wyniku translacji uszkodzonych mRNA, które nie posiadają kodonów stop. Sekwencje prokariotycznych peptydów, będących znacznikami uszkodzonych produktów translacji można znaleźć na stronie tmRNA Database: <http://www.ag.auburn.edu/mirror/tmRDB/>.

## ROZPUSZCZALNOŚĆ NADPRODUKOWANEGO BIAŁKA

To, czy w wyniku nadekspresji można otrzymać duże ilości rozpuszczalnego i dobrze zwiniętego, a zatem aktywnego białka zależy od kilku czynników. Zwykle tym trudniej im rekombinowany polipeptyd jest dłuższy, im więcej zawiera skrętów, reszt proliny, cysteiny i reszt hydrofobowych, a także od tego czy do prawidłowego zwiniania może dojść bez odpowiednich białek opiekuńczych i bez modyfikacji potranslacyjnych i czy w komórce *E. coli* jest podatny na proteolizę. Dla zwiększenia wydajności translacji i prawidłowego zwiniania w *E. coli*, a także po to żeby zwiększyć rozpuszczalność rekombinowanych białek, na końcu N dołącza się białko wiążące maltozę – MBP (40kDa), NusA (54,8 kDa) lub IF2(17,4 kDa). Indeks rozpuszczalności białka w cytosolu *E. coli* można wyliczyć na stronie: [www.biotech.ou.edu](http://www.biotech.ou.edu). Jest to tylko kalkulacja, która sprawdza się dla większości białek [15].

Do nadprodukcji krótkich peptydów, których rozpuszczalne produkty ekspresji są toksyczne lub szybko degradowane, wykorzystuje się N-końcowy, silnie hydrofobowy fragment bakteryjnej izomerazy ketosteroidowej, który w komórkach *E. coli* ulega wydajnej translacji (peptyd KSI; pET-31b, Novagen). Umieszcza się go na końcu N takiego peptydu lub jego tandemowych powtórzeń (wtedy należy stosować szczep BLR, Novagen). Produkt zostanie zdeponowany w postaci ciał inkluzyjnych [16]. Oczyszczanie białek które tworzą ciała inkluzyjne na większą skalę zwykle się nie opłaca, jednak niektóre enzymy deponowane w tych porowatych strukturach zachowują aktywność, a inne po solubilizacji ciał inkluzyjnych w obecności soli chaotropowych, w środowisku redukującym można zrenaturować. Ciała inkluzyjne powstają w wyniku znacznego nagromadzenia białek (>50% całkowitego białka komórki) łatwo wchodzących w „niespecyficzne” oddziaływania hydrofobowe. Zwykle ponad 90% masy ciał inkluzyjnych stanowi nadprodukowane białko, które nie uzyskało właściwej konformacji ze względu na proteolizę, niemożność utworzenia mostków dwusiarczkowych, brak modyfikacji potranslacyjnych czy nieobecność odpowiednich białek opiekuńczych. W skład ciał inkluzyjnych wchodzi także białka opiekuńcze *E. coli*: GroEL, DnaK, IbpA i IbpB, które uczestniczą w ich solubilizacji *in vivo*, w warunkach zahamowanej biosyntezy białka.

Powstawanie ciał inkluzyjnych można ograniczyć przez zmianę warunków hodowli (składu pożywki, temperatury, obniżenie poziomu ekspresji – szczep Turner), zastosowanie koekspresji izomeraz prolinowych lub disulfidowych (DsbA i DsbC), a także wysoko lub niskocząsteczkowych białek opiekuńczych. Ograniczenie proteolizy wymaga stosowania szczepów z defektywnymi genami proteaz lub z uszkodzonym systemem odpowiedzi na szok termiczny. W przypadku „trudnych” białek o charakterze hydrofobowym dobre efekty przynosi stosowanie szczepów C41(DE3) lub C43(DE3) pochodzących od BL21(DE3), w których można produkować białka błonowe, ponieważ w wyniku nadekspresji dochodzi raczej do proliferacji wewnętrznych błon aniżeli do powstawania ciał inkluzyjnych [17].

Uzyskanie wydajnej nadprodukcji heterologicznego białka w *E. coli*, w formie rozpuszczalnej, mimo że układ jest najlepiej poznany wymaga zwykle wypróbowania kilku różnych konstrukcji, w różnych warunkach hodowli, z zastosowaniem różnych szczepów albo podjęcia prób solubilizacji renaturacji i doczyszczania białka zdeponowanego w ciałach inkluzyjnych. W tym ostatnim przypadku trzeba podjąć jeszcze więcej prób, optymalizując tempo renaturacji, pH zastosowanego buforu, temperaturę, siłę jonową, zastosowanie detergentu i jego stężenie, soli chaotropowych, czynników redukujących, związków chelatujących, jonów metali czy niskocząsteczkowych dodatków ułatwiających zwinianie, jak np. glicyna albo L-arginina. Ponad 600 protokołów renaturacji białek z solubilizowanych ciał inkluzyjnych można znaleźć w bazie danych REFOLD: <http://refold.med.monash.edu.au>. W oparciu o te dane opracowano 96-dolkowy system iFOLD (Novagen) pozwalający na szybkie wyszukanie warunków renaturacji rekombinowanego białka zbliżonych do optymalnych [18].

## SEKRECJA DO PRZESTRZENI PERIPLAZMATYCZNEJ

Cytosol *E. coli* to środowisko redukujące, gdzie nie dochodzi do powstawania mostków dwusiarczkowych. Zaobserwowano to na przykładzie alkalicznej fosfatazy, która zostaje właściwie zwinięta dopiero w przestrzeni periplazmatycznej. W cytosolu *E. coli* nie są właściwie zwiniane również te białka, które w natywnej konformacji zawierają liczne, niesparowane reszty cysteiny. W takich przypadkach mamy do wyboru albo pozostać przy ekspresji białka w cytosolu i stosować szczepy z uszkodzonym systemem tioredoksyny (*trxB*) (szczep AD494, Novagen) lub także glutaredoksyny (*trxB/gor*) (Origami, Novagen), zastosować koekspresję izomerazy disulfidowej DsbC pozostającej w cytosolu, albo zdecydować się na ekspresję rekombinowanego białka kierowanego do przestrzeni periplazmatycznej. W tym celu, na końcu N dołącza się peptyd sygnałowy rozpoznawany przez translokazę Sec. Podczas translokacji jest on odcinany przez proteazę związaną z wewnętrzną błoną, której centrum aktywne znajduje się po stronie periplazmatycznej. Najczęściej stosowane peptydy sygnałowe pochodzą z genów *ompT*, *ompA*, *pelB* (pET22b), *phoA*, *malE*, *lamB* i z  $\beta$ -laktamazy. Zwykle składają się z 18–30 reszt aminokwasowych: dwóch lub większej liczby reszt zasadowych na końcu aminowym, siedmiu reszt budujących hydrofobowy rdzeń i hydrofilowego końca C, rozpoznawanego przez proteazę. Hydrofilowy koniec zawiera zwykle małe reszty (Ala, Gly lub Ser) w pozycjach od –1 do –3 (jeśli +1 to pierwsza pozycja dojrzałego białka), poprzedzone w pozycji –6 resztą łamiącą helisę (Pro lub Gly) [17,19].

Niektóre białka tworzą ciała inkluzyjne także w periplazmie, stąd, przydatna bywa koekspresja izomeraz disulfidowych, eksportowanych do tego przedziału. Inne zalety to mniejsze ryzyko proteolizy, łatwe uwolnienie produktu z tego przedziału przez szok osmotyczny i mniejsza pula białek na początku oczyszczania. System wykorzystujący sekrecję nadprodukowanego białka do periplazmy ma także wady. Może on wymagać ograniczonej nadekspresji, by system sekrecji nie został zablo-

**Tabela 2.** Zastosowanie wektorów Duet i szczepów do koekspresji od czterech do ośmiu rekombinowanych białek, na podstawie katalogu Novagen 2006/2007.

Odpowiednie kombinacje wektorów				Rekombinowane białka (liczba)	Grupa odpowiednich szczepów
Wektor 1	Wektor 2	Wektor 3	Wektor 4		
pETDuet <sup>TM</sup> -1 (Amp <sup>R</sup> )	pACYCDuet <sup>TM</sup> -1 (Cam <sup>R</sup> )	pRSFDuet <sup>TM</sup> -1 lub pCOLADuet <sup>TM</sup> -1(Kan <sup>R</sup> )	pCDFDuet <sup>TM</sup> -1 (Sm <sup>R</sup> )	8	A
pRSFDuet <sup>TM</sup> -1 lub pCOLADuet-1(Kan <sup>R</sup> )	pCDFDuet-1 (Sm <sup>R</sup> )	pACYCDuet-1 (Cam <sup>R</sup> )	żaden	6	A
pETDuet-1 (Amp <sup>R</sup> )	pACYCDuet-1 (Cam <sup>R</sup> )	pRSFDuet-1 lub pCOLADuet-1(Kan <sup>R</sup> )	żaden	6	A
pACYCDuet-1 (Cam <sup>R</sup> )	pRSFDuet-1 lub pCOLADuet-1(Kan <sup>R</sup> )	żaden	żaden	4	A
pETDuet-1 (Amp <sup>R</sup> )	pACYCDuet-1 (Cam <sup>R</sup> )	pCDFDuet-1	żaden	6	B
pETDuet-1 (Amp <sup>R</sup> )	pACYCDuet-1 (Cam <sup>R</sup> )	żaden	żaden	4	B
pACYCDuet-1 (Cam <sup>R</sup> )	pCDFDuet-1 (Sm <sup>R</sup> )	żaden	żaden	4	B
pETDuet-1 (Amp <sup>R</sup> )	pRSFDuet-1 lub pCOLADuet-1(Kan <sup>R</sup> )	pCDFDuet-1 (Sm <sup>R</sup> )	żaden	6	C
pETDuet-1 (Amp <sup>R</sup> )	pRSFDuet-1 lub pCOLADuet-1(Kan <sup>R</sup> )	żaden	żaden	4	C
pRSFDuet-1 lub pCOLADuet-1(Kan <sup>R</sup> )	pCDFDuet-1 (Sm <sup>R</sup> )	żaden	żaden	4	C
pETDuet-1 (Amp <sup>R</sup> )	pCDFDuet-1 (Sm <sup>R</sup> )	żaden	żaden	4	D

Grupy szczepów				
Grupa A	Grupa B	Grupa C	Grupa D	
B834(DE3)	B834(DE3)	B834(DE3)	B834(DE3)	Rosetta-gami (DE3)*
BL21(DE3)	BL21(DE3)	B834(DE3)pLysS <sup>†</sup>	B834(DE3)pLysS	Rosetta-gami(DE3)pLysS*
BLR(DE3)	BLR(DE3)	BL21(DE3)	BL21(DE3)	Rosetta-gami 2(DE3)*
HMS174(DE3)	HMS174(DE3)	BL21(DE3)pLysS <sup>†</sup>	BL21(DE3)pLysS	Rosetta-gami 2(DE3)pLysS*
NovaBlue(DE3)	NovaBlue(DE3)	BLR(DE3)	BLR(DE3)	Rosetta-gami B(DE3)
Origami <sup>TM</sup> 2(DE3)*	Origami(DE3) *	BLR(DE3)pLysS <sup>†</sup>	BLR(DE3)pLysS	Rosetta-gami B(DE3)pLysS
Tuner <sup>TM</sup> (DE3)	Origami 2(DE3)*	HMS174(DE3)	HMS174(DE3)	Tuner(DE3)
	Origami B(DE3)	HMS174(DE3)pLysS <sup>†</sup>	HMS174(DE3)pLysS	Tuner(DE3)pLysS
	Tuner(DE3)	NovaBlue(DE3)	NovaBlue(DE3)	
		Origami 2(DE3)*	Origami (DE3)*	
		Origami 2(DE3) pLysS <sup>†</sup>	Origami (DE3) pLysS*	
		Rosetta <sup>TM</sup> (DE3)	Origami 2(DE3)*	
		Rosetta(DE3)pLysS <sup>†</sup>	Origami 2(DE3)pLysS*	
		Rosetta 2(DE3)	Origami B(DE3)	
		Rosetta 2(DE3)pLysS <sup>†</sup>	Origami B(DE3)pLysS	
		RosettaBlue <sup>TM</sup> (DE3)	Rosetta(DE3)	
		RosettaBlue(DE3)pLysS <sup>†</sup>	Rosetta(DE3)pLysS	
		Rosetta-gami <sup>TM</sup> 2(DE3)*	Rosetta 2(DE3)	
		Rosetta-gami 2(DE3) pLysS <sup>†</sup>	Rosetta 2(DE3)pLysS	
		Tuner(DE3)	RosettaBlue(DE3)	
		Tuner(DE3)pLysS <sup>†</sup>	RosettaBlue(DE3)pLysS	

\*Szczepy posiadające mutację rpsL, dającą oporność na streptomycynę; stąd do selekcji rekombinantów pCDFDuet trzeba używać spektynomycyny.

<sup>†</sup>Szczepy zawierające plazmid z replikonem P15A; przy stosowaniu do koekspresji plazmidów pETDuet-1 i pRSFDuet-1, może wystąpić znaczne obniżenie poziomu ekspresji ramek odczytu wklonowanych do pRSFDuet-1. Taki efekt nie występuje, kiedy zamiast pRSFDuet-1 stosuje się wektor pCOLADuet-1. Oznaczenia oporności na antybiotyki: Amp<sup>R</sup>, ampicylina/karbenicylina; Kan<sup>R</sup>, kanamycyna; Cam<sup>R</sup>, chloramfenikol; Sm<sup>R</sup>, streptomycyna/spektynomycyna. Informacje dotyczące szczepów można znaleźć na stronie: [www.novagen.com](http://www.novagen.com)

kowany. Można zdecydować się na koekspresję składników systemu Sec-SecY i SecE na poziomie zbliżonym do ekspresji genu rekombinowanego białka. Inny problem to łatwe przedostawanie się białek z przestrzeni periplazmatycznej przez zewnętrzną błonę do pożywki. System nadekspresji z sekrecją do periplazmy pozwala czasem uzyskać aktywną formę „trudnego” enzymu, lecz konieczność ograniczenia nadprodukcji i problemy z utrzymaniem produktu w tym przedziale czynią system sekrecyjny mało przydatnym do produkcji rekombinowanych białek na dużą skalę. Ważne jest uzyskanie szczepów o bardzo sprawnym systemie sekrecji i optymalizacja wa-

runków ich hodowli na przykład w warunkach szoku osmotycznego, np. z zastosowaniem takich dodatków jak sorbitol i glicylobetaina [17,19].

## ZNACZNIKI UŁATWIAJĄCE OCZYSZCZANIE REKOMBINOWANYCH BIAŁEK

Produkcja rekombinowanych białek to często jedyna lub najtańsza metoda otrzymania homogennego preparatu enzymu, białka, peptydu, pozwalającego na zbadanie jego struktury i funkcji, często jedyny sposób, żeby enzym produkowany na potrzeby medycyny był dobrze



scharakteryzowany. Jeśli rekombinowane białko ma być narzędziem w przemyśle, rolnictwie, przetwarzaniu żywności, biologii molekularnej – umożliwia wprowadzenie pożądaných zmian struktury (unieruchomienie na wybranym podłożu, poprawienie odporności na wymagane warunki jak pH, temperatura, siła jonowa) i funkcji (zmiana specyficzności substratowej czy wprowadzenie nowych możliwości aktywacji lub blokowania aktywności itp.).

Żeby osiągnąć powyższe cele trzeba wyprodukowane w kontrolowanych warunkach białko oczyścić jak najszybciej i z jak największą wydajnością. Dlatego ramki odczytu rekombinowanych białek łączy się z sekwencjami kodującymi krótkie peptydy lub całe domeny innych białek, które pozwalają na oczyszczenie rekombinowanego białka w wyniku jednostopniowej chromatografii powinowactwa, przeprowadzonej niemal natychmiast po jego uwolnieniu z komórki.

Dołączony peptyd nie powinien zmieniać aktywności enzymu ani zaburzać jego struktury i powinien się nadawać do oczyszczania wielu różnych białek z wykorzystaniem tego samego, niedrogiego złoża (nośnik - łącznik - ligand). Jeśli trzeba dodany peptyd usunąć, rekombinowane białko musi być tak zaprojektowane, by po lub w trakcie oczyszczania można go było odciąć. Oczyszczanie rekombinowanych białek na złożu powinno być możliwe w różnych warunkach, czyli przy różnym stężeniu soli, pH, w obecności detergentów, czynników chelatujących, redukujących i związków chaotropowych. Elucja białka ze złoża powinna przebiegać w łagodnych warunkach tak, by nie dochodziło do jego denaturacji na tym etapie. Złoże powinno być trwałe: odporne na modyfikacje chemiczne, proteolizę i powinno się nadawać do wielokrotnego wykorzystania – łatwe i skuteczne usuwanie pozostałości oraz regeneracja.

Najbardziej wypróbowanym peptydem dołączanym do rekombinowanych białek jest heksapeptyd His<sub>(6)</sub>, który można umieszczać na końcu N lub C rekombinowanych enzymów, rzadko hamuje ich aktywność i jest z powodzeniem stosowany w pro- i eukariotycznych układach ekspresyjnych.

W roku 1975 po raz pierwszy opisano możliwość wiązania i oczyszczania białek zawierających reszty histydyny, cysteiny i tryptofanu na złożach chromatograficznych z unieruchomionymi jonami metali przejściowych, takich jak Co<sup>2+</sup>, Ni<sup>2+</sup>, Cu<sup>2+</sup> i Zn<sup>2+</sup> (IMAC, ang: *immobilized metal ion chromatography*). Pierwszym użytym w tym celu ligandem chelatującym jon metalu był kwas iminodioctowy. Okazało się, że najwydajniej można w ten sposób oczyszczać białka zawierające liczne, występujące po sobie reszty histydyny. Pierścienie imidazolowe histydyny są w pH fizjologicznym donorami elektronów tworząc koordynacyjne wiązania z unieruchomionymi jonami metali. Związane białka zawierające trakty reszt histydyny można wymyć ze złoża obniżając pH (4,5–5,9) lub stosując imidazol (50–500 mM) [20].

W 1987 roku Hochuli opracował sposób wiązania Ni<sup>2+</sup> do stałego podłoża za pomocą kwasu nitrylotrioctowego (Ni<sup>2+</sup>-

NTA). Czterowartościowe wiązanie chelatujące jon niklu, pozostawia dwie wolne wartościowości, które mogą być wykorzystane do wiązania biopolimerów. Białka z motywem His<sub>(3)</sub> wiążą się wydajnie do takiego złoża tylko w warunkach natywnych, np. w buforach fosforanowych, a His<sub>(6)</sub> także w wysokim stężeniu soli i w obecności związków chaotropowych, takich jak mocznik czy chlorowodorek guanidyny. W złożach TALON jon kobaltu jest nieco silniej związany ze stałym podłożem – np. sefarozą za pośrednictwem karboksymetyloasparagianinu (Co<sup>2+</sup>-CMA), co zmniejsza ryzyko wymycia jonów kobaltu razem z oczyszczanym białkiem. Złoża Co<sup>2+</sup>-CMA pozwalają na elucję związanych białek w łagodniejszych warunkach, a jednocześnie powstaje mniej wiązań niespecyficznych niż na złożach Ni<sup>2+</sup>-NTA. Pojemność złożów Ni<sup>2+</sup>-NTA i Co<sup>2+</sup>-CMA zwykle mieści się w zakresie od 5 do 15 mg białka na 1g złoża.

Jednostopniowa chromatografia na takich złożach pozwala na ponad 95% oczyszczenie białek ze znacznikiem His<sub>(6)</sub>. Wiązanie białek przeprowadza się zwykle w buforach zawierających niewielkie stężenia imidazolu – zwykle 5–20 mM, co zabezpiecza przed niespecyficznym wiązaniem białek zawierających mniejszą ilość kolejno występujących reszt histydyny. W komórkach *E. coli* nie występują białka z dłuższymi traktami reszt histydyny. Żeby uniknąć izolacji białek, które mogą wiązać się z rekombinowanym białkiem-His<sub>(6)</sub> poprzez mostki dwusiarczkowe w buforach do lizy komórek stosuje się dodatek β-merkaptotetanolu w stężeniu do 20 mM w przypadku złożów Ni<sup>2+</sup>-NTA, do 10mM w przypadku Co<sup>2+</sup>-CMA. Dla uniknięcia niespecyficznych oddziaływań oczyszczanego białka-His<sub>(6)</sub> z innymi białkami i z kwasami nukleinowymi stosuje się detergenty (0,1–1% Triton X-100 lub 0,5% sarkozynian sodowy), wysokie stężenie soli – nawet 2 M NaCl, a w celu ograniczenia oddziaływań hydrofobowych – dodatek etanolu lub glicerolu (do 30%) [20,21].

Na złożach Co<sup>2+</sup>-CMA, w podobnych warunkach można oczyszczać białka zawierające polihistydynowy znacznik naturalnego pochodzenia – tzw. HAT-tag. HAT to 19 aminokwasowa sekwencja pochodząca z końca N kurzej, mięśniowej dehydrogenazy mleczanowej, która ma mniejszy ładunek i mniej wpływa na rozpuszczalność rekombinowanych białek niż His<sub>(6)</sub>. Mocznik osłabia wiązanie peptydu HAT do Co<sup>2+</sup>-CMA bardziej niż chlorowodorek guanidyny [15,21].

**Strep-tag** to nonapeptyd (AWRHPQFGG), wiążący się do streptawidyny w kieszeni, w której przyłącza się biotyna, a zatem biotyna lub jej pochodne są kompetytorem wiązania tego peptydu ze streptawidyną. Strep-tag można umieszczać tylko na końcu C rekombinowanych białek, aby było możliwe utrzymanie mostka solnego między wolną grupą karboksylową, a resztą Arg streptawidyny. W wyniku poszukiwań uniwersalnego peptydu nadającego się także do umieszczenia na końcu N rekombinowanych białek wyselekcjonowano strep-tagII (IBA GmbH) o sekwencji NWSHPQFEK, który można skrócić do WSH-PQFEK (pET 51, Novagen). Taki peptyd wykazuje obniżone powinowactwo do streptawidyny ( $K_D = 13 \mu M$ ). W wyniku losowej mutagenazy pętli (pozycje od 44-52) znajdującej się w pobliżu kieszeni streptawidyny wiążącej bio-

tyne udało się wyselekcjonować mutanty silniej wiążące strep-tagII ( $K_D = 1 \mu\text{M}$ ), mutant 2-44 Glu-Ile, 45 Ser-Gly, 47 Val-Arg. Tak zmodyfikowana streptawidyna występuje pod handlową nazwą Strep-Tacin (IBA GmbH). Strep-Tacin wiąże strep-tagII około 100 razy silniej niż streptawidyna. Strep-tag pozwala na jednostopniowe oczyszczenie rekombinowanego białka w 99%, a ze względu na neutralny skład reszt aminokwasowych zwykle nie wpływa na aktywność rekombinowanych enzymów i zwykle nie ma potrzeby usuwania tego znacznika. W przeciwieństwie do His-tagu nie powoduje agregacji i wytrącania się białek w obecności jonów metali ciężkich, a sposób oczyszczania nie stwarza zagrożenia występowaniem paramagnetycznych zanieczyszczeń uniemożliwiających zastosowanie NMR do badania struktury oczyszczonego białka. Możliwość zastosowania łagodnych warunków wiązania i elucji (2,5 mM detiobiotyny) pozwala na oczyszczanie aktywnych białek i kompleksów wieloenzymatycznych. Strep-tagII jest także przydatny do oczyszczania rekombinowanych białek błonowych, a największą jego zaletą jest jego „komplementarność” z His-tagiem, a mianowicie możliwość stosowania związków chelatujących (EDTA, EGTA) i redukujących (DTT i  $\beta$ -merkaptotetanolu – do 50 mM), a także umiarkowanych stężeń soli chaotropowych (GuHCl do 1 M), detergentów. 6M mocznik niszczy wiązanie strep-tag – Strep-Tacin, lecz nie denaturuje streptawidyny [22].

Do złożeń Strep-Tacin wiążą się endogenne biotynylowane białka, lecz można temu zaradzić stosując bufony ekstrakcyjne zawierające awidynę. Kolejną zaletą takiego znacznika jest uniwersalny sposób detekcji rekombinowanego białka z wykorzystaniem streptawidyny i biotynylowanego przeciwciała sprzężonego z fosfatazą lub peroksydazą [23].

**GFP** – znajduje zastosowanie do produkcji rekombinowanych białek błonowych, ułatwia ich oczyszczanie poprzez chromatografię oddziaływań hydrofobowych, pozwala na monitorowanie ich ekspresji [24].

**PKA** – motyw o sekwencji RRASV (pET33, Novagen), fosforylowany przez kinazę białkową A, pozwala na wyznaczenie oczyszczonego białka fosforanem  $^{32}\text{P}$  lub  $^{33}\text{P}_\gamma\text{ATP}$  w obecności kinazy z mięśnia sercowego zależnej od cAMP, a następnie detekcję rekombinowanego białka, produktów jego modyfikacji i proteolizy w badanym ekstrakcie, a także oddziaływań z innymi białkami [25].

**S-tag** – to 15-aminokwasowy fragment rybonukleazy A, odcinany przez subtylizynę, który silnie wiąże ( $K_d = 10^{-9}\text{M}$ ) pozostałą część RNazy A – tzw. białko S. Komplementacja przywraca aktywność RNazy A, a sztuczny substrat ArU-AA z fluoroforem na końcu 5' i z wygaszaczem na końcu 3' pozwala na technicznie proste wykrycie ilości rekombinowanego białka mniejszych niż 1 fmol oraz określenie poziomu ekspresji badanego genu w surowych ekstraktach komórkowych [23].

## ZAKOŃCZENIE

Systemy nadekspresji genów heterologicznych białek w komórkach *E. coli* są wciąż udoskonalane. Hodowle laboratoryjne mogą być automatyzowane. Opracowano pożywki

zapewniające autoindukcję (OvernightExpress, Novagen). Nadekspresję można zatem szybko optymalizować w coraz większej liczbie wariantów. Udało się przezwyciężyć „przeciekanie” ekspresji promotora T7 (pETcoco, Novagen), a także opracować technologie wykorzystania wielu, również nie wymienionych tu peptydów do detekcji i oczyszczania rekombinowanych białek. Kłopotliwą pozostaje konieczność stosowania licznych antybiotyków i niestabilność wielokrotnych transformantów – stąd trudności w przygotowaniu zamrożonych zaszczepów o identycznym mianie. Istotne jest dopracowanie niezawodnych sposobów odcinania znaczników i dalsze gromadzenie danych dotyczących zwijania i renaturacji poszczególnych białek.

## PIŚMIENNICTWO

1. Fisher R, Drossard J, Emans N, Commandeur U, Hellwig S (1999) Towards molecular farming in the future: *Pichia pastoris*-based production of single chain antibody fragments. *Biotechnol Appl Biochem* 30: 117-120
2. Peterson RKD, Arntzen CJ (2004) On risk and plant-based biopharmaceuticals. *Trends Biotech* 22: 64-66
3. Demian AL (2000) Microbial Biotechnology. *Trends Biotech* 18: 26-31
4. Lee SY (1996) High cell density culture of *Escherichia coli*. *Trends Biotech* 14: 98-105
5. Korz DJ, Rinas U, Hellmuth K, Sanders EA, Deckwer W-D (1994) Simple fed-batch technique for high cell density cultivation of *Escherichia coli*. *J Biotechnol* 39: 59-65
6. Hannig G, Makrides SC (1998) Strategies for optimizing heterologous protein expression in *Escherichia coli*. *Trends Biotech* 16: 54-60
7. Sorensen HP, Mortensen KK (2005) Advanced genetic strategies for recombinant protein expression in *Escherichia coli* *J Biotechnol* 115: 113-128
8. Novagen (2006) pET System Manual. Wydanie 11
9. Sektas M, Szybalski W (2002) Novel single-copy pETcoco vector with dual controls for amplification and expression. *in* *Novations* 14: 6-8
10. Held D, Yaeger K, Novy R (2003) New coexpression vectors for expanded compatibilities in *E. coli*. *in* *Novations* 18: 4-6
11. Kozak M (1999) Initiation of translation in prokaryotes and eucaryotes. *Gene* 234: 187-208
12. Deutscher MP (1994) tRNA Processing Nucleases w tRNA, Structure, Biosynthesis and Function, W: Söll D, Rajbhandary UL (red) ASM Press, Washington, str. 51-65
13. Gustafsson C, Govindarajan S, Minshul J (2004) Codon bias in heterologous protein expression. *Trends Biotech* 22: 346-353
14. Makrides S (1996) Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiol Reviews* 60: 512-538
15. Terpe K (2003) Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol* 60: 523-533
16. Morreale G, Lee EG, Jones DB, Middelberg APJ (2004) Bioprocess-centered molecular design (BMD) for the efficient production of an internally active peptide. *Biotechnol Bioeng* 87: 912-923
17. Ventura S, Villaverde A (2006) Protein quality in bacterial inclusion bodies. *Trends Biotech* 24: 179-185
18. Leland P, Rane J, Grabski A (2006) Convenient optimization of protein refolding conditions with iFOLD Protein Refolding System1. *in* *Novations* 23: 3-8
19. Cornelis P (2000) Expressing genes in different *Escherichia coli* compartments. *Curr Opin Biotech* 11: 450-454
20. Qiagen (1999) NiNTA Magnetic Agarose Beads Handbook
21. www.clontech.com
22. Korndorfer IP, Skerra A (2002) Improved affinity of engineered streptavidin for the Strep-tag II peptide is due to a fixed open conformation of the lid-like loop at the binding site. *Protein Sci* 11: 883-893



23. Novagen (2006) Protein Purification and Detection Tools.
24. Drew D, Lerch M, Kunji E, Slotboom DJ, de Gier JW (2006) Optimization of membrane protein overexpression and purification using GFP fusions. *Nature Methods* 3: 303-313

25. de Arruda M, Burgess RR (1995) pET-33b(+): a pET vector that contains a protein kinase A recognition sequence. *inNovations* 4: 7-8

## Recombinant protein production in *Escherichia coli*

Przemysław Nuc<sup>1,✉</sup>, Katarzyna Nuc<sup>2</sup>

<sup>1</sup>Adam Mickiewicz University, Institute of Molecular Biology and Biotechnology, 89 Umultowska St., 61-614 Poznań, Poland

<sup>2</sup>August Cieszkowski University of Agriculture, Department of Biochemistry and Biotechnology, 35 Wołyńska St., 60-637 Poznań, Poland

✉e-mail: przem@amu.edu.pl

**Key words:** *Escherichia*, overexpression, system, vector

### ABSTRACT

Growing needs for efficient recombinant production pose new challenges; starting from cell growth optimization under overexpression conditions, improving vectors, gene and protein sequence to suit them to protein biosynthesis machinery of the host, through extending the knowledge of protein folding, fusion protein construction, and coexpression systems, to improvements in protein purification and renaturation technologies. Hitherto *Escherichia coli* is the most defined and the cheapest protein biosynthesis system. With its wealth of available mutants tested is the best suited to economically test new gene constructs and to scale up the recombinant protein production.

## PROTEIN EXTRACTION

from Novagen

# BugBuster™

protein extraction reagent

## Plus Benzonase®

nuclease

...lets you lyse *E. coli*,  
extract soluble protein,  
and reduce nucleic acid  
viscosity - fast.

Compatible with 2-D  
electrophoresis and other  
proteomics applications

Available in a variety of configurations:

- Ready-to-use reagent
- 10X concentrate
- HT (premixed with Benzonase)
- Primary Amine-free
- Complete Purification Kits with Resins



A brand of Cytoskeleton, Inc.  
an affiliate of Merck KGaA, Darmstadt, Germany

