

Produção de Proteínas Recombinantes em *Escherichia coli*

Prof. Dr. Catarina Akiko Miyamoto¹

Resumo

A produção de proteínas recombinantes para fins terapêuticos, veterinários, e agro-pecuários tem se mostrado bastante eficaz tecnológica e economicamente. O objetivo deste manuscrito é passar uma visão geral de como estas proteínas podem ser produzidas na bactéria *Escherichia coli*. Mostramos aqui desde a “pesca” do gene de interesse do genoma original, clonagem do mesmo em plasmídeos de expressão até a produção da proteína recombinante. Discutimos também a possibilidade de “melhorar” a sequência gênica para que aumente a produção protéica.

Palavras-Chave

proteína recombinante, clonagem, plasmídeo, PCR

Introdução

A tecnologia do DNA recombinante possibilita a produção de proteínas heterólogas em grande quantidade. O sistema de expressão adequado depende da proteína a ser produzida. O entendimento do dogma central da biologia molecular é essencial para a produção de proteínas recombinantes, uma vez que este compreende os três processos principais que a célula utiliza para que a informação genética seja expressa: replicação, transcrição e tradução (Figura 1). Em última análise, a mensagem genética presente no DNA pode ser decifrada e expressa na forma de proteína. Utilizando deste conhecimento, dos controles da expressão gênica e das preferências de códon em *Escherichia coli* (*E. coli*), proteínas de interesse biotecnológico podem ser produzidas em grande escala.

Autor

1 Mestre e Doutor em Ciências (Área de concentração – Bioquímica), USP. Docente das Faculdades Integradas de Três Lagoas – AEMS. Cursos de Biomedicina, Nutrição, Tecnologia em Processos Químicos e Tecnologia em Gestão Ambiental.

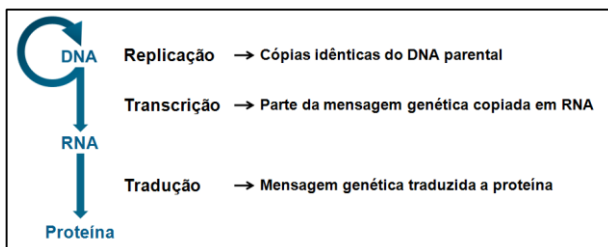


Figura 1. Dogma central da biologia molecular mostrando os três processos principais que a célula utiliza para que a mensagem genética seja expressa. Adaptado de Lehninger Principles of Biochemistry, 5ª Edição, W.H. Freeman and Company, Nova York, EUA.

Atualmente, empresas de biotecnologia sintetizam genes com códons preferenciais para vários sistemas de expressão. Inicialmente, nos ateremos à metodologia tradicional. Posteriormente, mostraremos como uma sequência de DNA pode ser melhorada para que haja aumento na produção de proteínas.

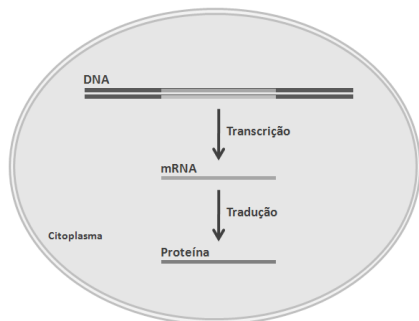
As etapas principais para a produção de uma proteína recombinante compreendem: (1) pesca do gene de interesse, (2) escolha do sistema de expressão, (3) clonagem, e (4) expressão. A seguir, descreveremos cada um destes passos com detalhe.

1. Pesca do gene de interesse

A primeira pergunta que surge é: Qual é o organismo original da proteína de interesse? De uma célula procariótica, um eucarioto inferior ou um eucarioto superior?

A sequência do RNA mensageiro (mRNA) de células procarióticas e de eucariotos inferiores é igual à do DNA genômico. A figura 2A mostra o dogma central de uma célula procariótica e de um eucarioto inferior. Deve-

A.



B.

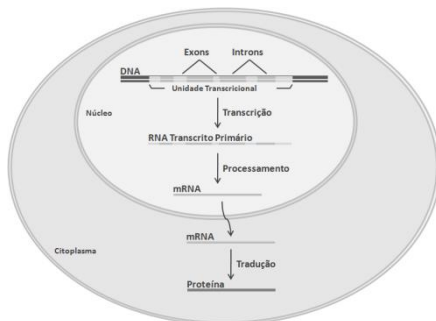


Figura 2. Dogma central da biologia molecular. (A) Procariotos – A sequência do mRNA é igual à sequência do DNA genômico. **(B) Eucariotos superiores** – O transcrito primário sofre processamento (perda dos introns), e somente os exons estão presentes no mRNA.

se ressaltar que o material genético do último está envolto pela membrana nuclear. A pesca do gene de interesse destes organismos pode então ser direta de seu DNA genômico, através da técnica da reação em cadeia da polimerase (PCR) (Figura 3A). Esta técnica, descrita abaixo, permite a pesca seguida da amplificação de qualquer sequência de DNA.

1.1. Reação em cadeia da polimerase (PCR)

Esta técnica permite a pesca direta do gene de interesse de procariotos e eucariotos inferiores, seguida da amplificação do mesmo. Consiste de três etapas: desnaturação do DNA, anelamento com os oligonucleotídeos iniciadores (*primers*), e amplificação do gene de interesse (Figura 3A).

A etapa da desnaturação do DNA, ou rompimento das ligações de hidrogênio entre as bases nitrogenadas permite que, na etapa seguinte, *primers* específicos se anelem às fitas individuais de DNA. A temperatura normalmente utilizada é de 94-95 °C por um período de 1-3 minutos.

A etapa do anelamento dos *primers* é importante uma vez que os mesmos devem se hibridizar de forma precisa na sequência original do DNA. Para se obter a temperatura ideal para este fim, testes com várias temperaturas de anelamento são realizados. A temperatura inicial é de 5 °C abaixo da temperatura de fusão (T_m) dos *primers*. A temperatura ideal pode ser maior ou menor. Termocicladores com gradiente de temperaturas são ideais para a realização destes testes, onde a temperatura de -5 °C abaixo das T_m s dos *primers* é locada no meio do gradiente.

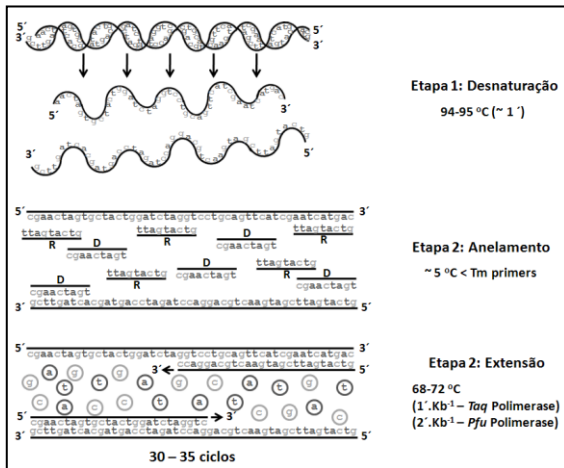
A etapa da amplificação da região codificadora do gene de interesse é realizada na presença da mistura dos quatro nucleotídeos e DNA polimerase proveniente de bactérias termofílicas. A temperatura ideal para estas enzimas realizarem a reação de extensão da segunda fita está entre 68-72 °C. O tempo de extensão varia de uma enzima para outra. As três etapas, acima descritas, são repetidas 30-35 vezes. Deste modo, a região flanqueada pelos dois *primers* é amplificada exponencialmente (Figura 3B).

Várias DNA polimerases estão disponíveis comercialmente, sendo as mais utilizadas, *Taq*- e *Pfu* polimerases. A primeira é proveniente de *Thermus aquaticus*, e a segunda de *Pyrococcus furiosus*. Deve-se ressaltar que a *Pfu* polimerase é uma enzima de alta fidelidade (1), sendo então recomendada para a pesca de genes de interesse.

1.2. Reação da transcriptase reversa seguida de PCR (RT-PCR)

A sequência de mRNAs de células eucarióticas superiores é diferente da do DNA genômico, desde que o produto primário da transcrição gênica sofre processamento para se transformar em mRNA (Figura 2B). Assim, pesca de genes destas células deve ser realizada através da reação da transcriptase reversa seguida de PCR. Esta enzima

A.



B.

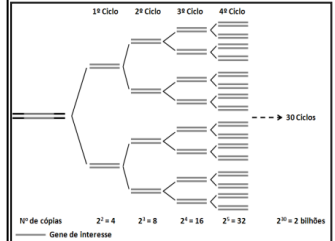


Figura 3. Reação em cadeia da polimerase (PCR). (A) As etapas da reação: Etapa 1 – DNA é submetido a altas temperaturas para que ocorra a desnaturação; Etapa 2 – DNA desnaturado está apto a se anelar com os *primers* diretos (D) e reversos (R) iniciadores; Etapa 3 – DNA polimerase sintetiza a segunda fita de DNA na presença dos desoxi-ribonucleosídeos 5'-trifosfatos dTTP (**t**), dATP (**a**), dCTP (**c**) e dGTP (**g**). (B) **Amplificação exponencial** da sequência flanqueada pelos *primers*.

catalisa a síntese de uma fita de DNA complementar (cDNA) à do RNA, degrada a fita de RNA do híbrido DNA-RNA, e em seguida sintetiza a segunda fita do DNA (2) (Figura 4).

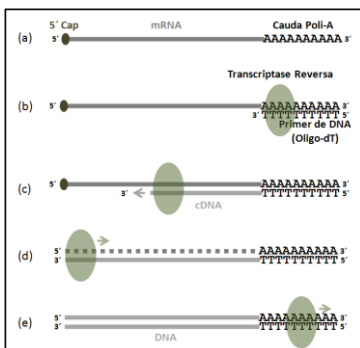


Figura 4. Reação da transcriptase reversa. Esta enzima utiliza a fita de RNA como molde e sintetiza uma de DNA complementar à primeira. Em seguida, degrada a fita de RNA do híbrido DNA-RNA e por final, sintetiza a segunda fita de DNA.

1.3. Oligonucleotídeos iniciadores (*primers*)

A pesca e amplificação do gene de interesse por PCR e RT-PCR exigem *primers* iniciadores para as enzimas sintetizarem a fita de DNA. As DNA polimerases utilizam *primers* direto e reverso que abrangem as regiões contíguas e as terminações 5' da fita direta e da reversa, respectivamente, do gene (Figura 5A).

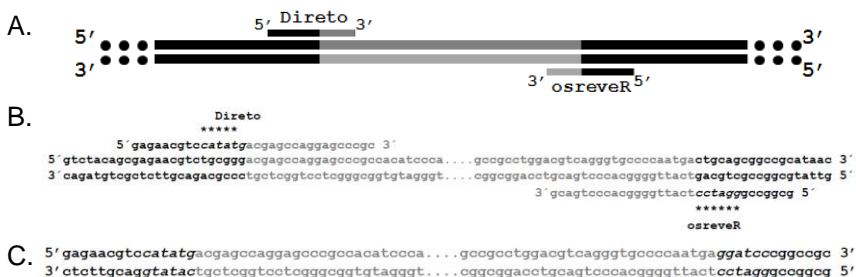


Figura 5. Primers iniciadores. A. *Primers* direto e reverso correspondem a cópias idênticas das terminações 5' das duas fitas do gene de interesse. B. *Primers* com sítios de restrição (em itálico) para posterior clonagem em plasmídeos de expressão. As bases que não pareiam com a fita complementar estão salientadas com asteriscos. C. **Produto de PCR (amplicon)** da região do DNA mostrada em B.

Estes *primers* podem ser desenhados de modo que contenham os sítios de restrição que serão utilizados posteriormente para clonagem da sequência codificadora de interesse em plasmídeos de expressão. Os mesmos devem também ter *Tms* semelhantes para facilitar a obtenção de temperatura de anelamento adequada. A figura 5B mostra os *primers* (direto e reverso) utilizados para a clonagem de uma sequência de DNA. Os sítios de restrição *Nde* I (*catatg*) e *Bam* HI (*ggatcc*) flanqueiam as terminações 5' e 3' do gene, respectivamente. Note que algumas bases não pareiam com a fita complementar (salientados com asteriscos). A figura 5C mostra o produto de PCR (ou amplicon) obtido utilizando estes *primers*.

A enzima transcriptase reversa necessita de um *primer* complementar ao RNA para a síntese de cDNA. Poli-dT é utilizada no caso de sequências codificadoras desconhecidas (Figura 4). Por outro lado, usa-se *primers* específicos para regiões codificadoras conhecidas.

2. Escolha do sistema de expressão

Os plasmídeos de expressão contêm um promotor forte e um sítio de múltipla clonagem, além da origem de replicação (OR), do marcador genético (MG) e outras sequências (Figura 6).

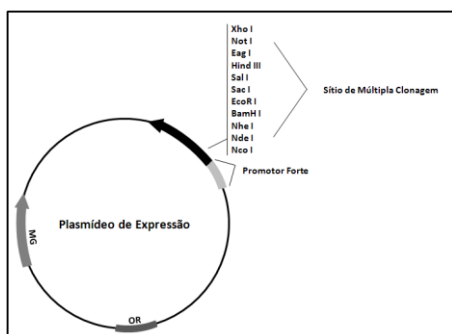
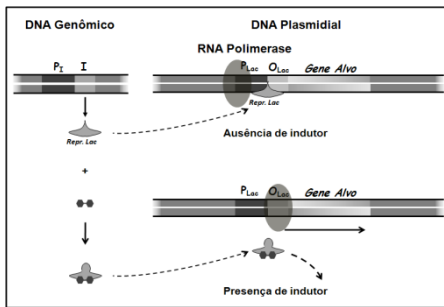


Figura 6. Desenho esquemático de um plasmídeo de expressão mostrando as principais características: promotor forte, sítio de múltipla clonagem (SMC), origem de replicação (OR) e marcador genético (MG).

Os principais plasmídeos de expressão utilizam os promotores Lac, Tac e T7. O promotor Lac é semelhante ao utilizado pelas cepas tipo selvagem de *E. coli*, sendo então reprimido pela proteína repressora *lacI* na ausência de indutor (lactose ou isopropiltiogalactosídeo - IPTG) e induzido em sua presença (Figura 7A). O promotor Tac contém a região promotora (onde a RNA polimerase se liga) do *Operon Trp* e a região operadora (onde a proteína repressora se liga) do *Operon Lac* (Figura 7B). Assim, ambos promotores são induzidos por lactose ou IPTG e utilizam a RNA polimerase bacteriana para transcrever o gene de interesse (Figuras 7A e 7B).

A.



B.

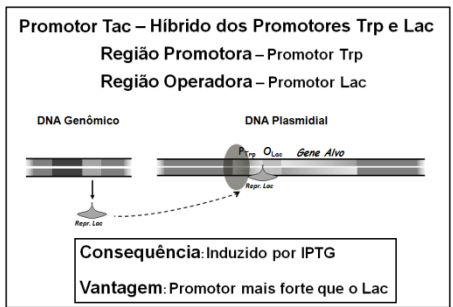


Figura 7. Comparação dos promotores Lac e Tac. A. Promotor Lac – A indução da expressão do gene alvo acontece somente na presença do indutor (IPTG). **B. Promotor Tac** – por ter a região operadora do promotor Lac, a indução ocorre também na presença de IPTG.

O promotor T7, por outro lado, utiliza a RNA polimerase do fago T7. A vantagem deste promotor é que esta RNA polimerase é 20x mais ativa do que a bacteriana, e conseqüentemente, o nível de expressão protéica é bem maior. Entre os plasmídeos que utilizam a T7 RNA polimerase, estão os do sistema pET de expressão (3; Manual pET System, Novagen). Neste sistema, a indução da expressão da proteína de interesse é indireta. A indução direta é do gene da T7 RNA polimerase, lisogenizada no DNA genômico, que está sob controle do promotor Lac (Manual pET System, Novagen). Ao ser traduzida, a proteína T7 RNA polimerase se liga em sua região promotora presente no plasmídeo pET, e então, o gene de interesse é transcrito e a seguir, a proteína é expressa (Figura 8, Manual pET System, Novagen). Gerações mais recentes de vetores pET apresentam também a região operadora do *operon Lac* (Figura 8, Manual pET System, Novagen); com isto a expressão basal de T7 RNA polimerase não é suficiente para que ocorra também a expressão basal da proteína de interesse. Ou seja, esta só será expressa após

indução com o indutor. Isto é importante porque proteínas tóxicas às bactérias não são expressas após indução quando há vazamento de expressão de forma basal.

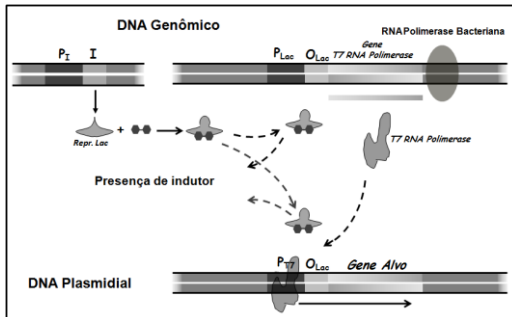


Figura 8. Esquema de indução do sistema pET. Na presença de indutor (●●), o repressor Lac não se liga em *O_{Lac}* e a T7 RNA polimerase é expressa. Esta então se liga em seu promotor presente no DNA plasmidial.

3. Clonagem do gene de interesse

O gene de interesse pode ser clonado em um único sítio de restrição ou em dois diferentes. Utiliza-se uma ou outra estratégia dependendo da finalidade do procedimento. A clonagem do gene para fins de produção protéica é normalmente realizada em dois diferentes sítios de restrição para que a sequência se insira no plasmídeo de forma direcionada. Focalizaremos nossa discussão para esta finalidade.

Inicialmente, um estudo do mapa de restrição do gene de interesse deve ser realizado para em seguida selecionar os sítios que flanquearão o mesmo. Os sítios ausentes ou os que estão presentes em menor número são os preferenciais. Conjuntamente, os sítios de restrição presentes e suas posições relativas no sítio de múltipla clonagem (smc) devem ser observados. Os sítios mais a montante no smc devem ser escolhidos para flanquear a terminação 5' do gene de interesse, enquanto os posicionados mais a jusante, a 3'.

Sequências codificadoras podem ser clonadas de modo que as proteínas resultantes sejam expressas de forma autêntica ou quimérica. A

primeira contém apenas os aminoácidos presentes na proteína original. As quimeras podem conter caudas com aminoácidos extras tanto na porção N- como na C-terminal da proteína resultante. Estas caudas são úteis para o processo de purificação da mesma. Caudas de histidinas, da enzima tioredoxina, entre outras são amplamente utilizadas.

Para a expressão de proteína autêntica, a terminação 5' de sua sequência codificadora deve conter o códon de iniciação da tradução (ATG). Inserção de sítio de restrição contendo este códon é conveniente para esta finalidade. As enzimas *Nde* I e *Nco* I podem ser então utilizadas, uma vez que seus sítios (CATATG e CCATGG, respectivamente) contêm este códon. A terminação 3' deve conter um dos códons de terminação (TGA, TAA ou TAG) seguido a jusante de um sítio de restrição presente no smc do plasmídeo selecionado (Figura 6). Um exemplo de amplicon utilizando esta estratégia está mostrado na Figura 5C. A sequência de interesse está flanqueada pelos sítios de restrição *Nde* I e *Bam* HI. O amplicon obtido e o plasmídeo onde o mesmo será clonado são digeridos com as mesmas enzimas, e a seguir, ambos são ligados pela enzima DNA ligase. Um exemplo de plasmídeo recombinante assim obtido está mostrado na figura 9.

Para a produção de proteína quimérica com cauda na porção N-terminal, sua região codificadora deve ser clonada em fase com a sequência da cauda extra de aminoácidos. Para este fim, insere-se convenientemente um sítio de restrição adequado na terminação 5' do gene. Sua terminação 3' deve conter um dos códons de terminação, seguido a jusante por outro sítio de restrição. Por outro lado, na construção do plasmídeo de expressão de uma proteína com cauda na porção C-terminal, o gene de interesse deve conter o códon de iniciação de tradução, o códon de terminação, caso presente, deve ser cancelado, e a

seqüência da cauda deve estar em fase com a do primeiro. Plasmídeos para estes fins são comercialmente disponíveis.

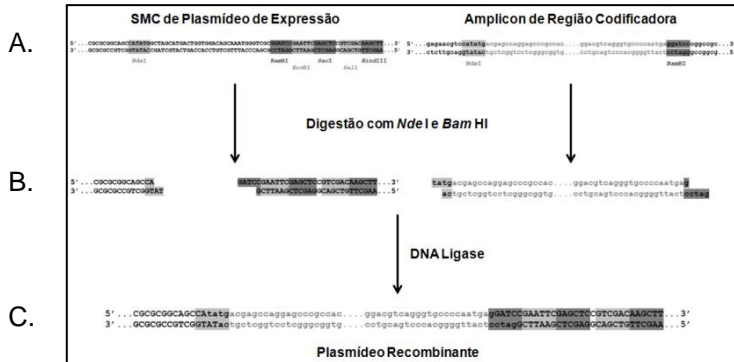


Figura 9. Produção de plasmídeo recombinante. A. O plasmídeo de expressão e amplicon são digeridos com as mesmas enzimas de restrição (*Nde I* e *Bam HI*). B. Produtos das digestões. C. Plasmídeo recombinante formado pela ligação dos produtos das digestões pela enzima DNA ligase.

4. Expressão das proteínas recombinantes

Os plasmídeos recombinantes de expressão obtidos, conforme acima descrito, são então introduzidos em células de cepas de *E. coli* convenientemente engenheiradas para que as mesmas produzam a proteína de interesse. Estas células “transformadas” com os plasmídeos são plaqueadas em meio de cultura semi-sólido e incubadas (37 °C, 12-14 hs). Algumas colônias são inoculadas em meio de cultura líquido rico, até densidade óptica a 600 nm (DO₆₀₀) adequada para cada sistema e então induzidas com indutor (lactose ou IPTG). Em sistemas que utilizam a RNA polimerase bacteriana, a cultura bacteriana é induzida na fase midi-log de crescimento (DO₆₀₀ ~ 0,4-0,5); os que utilizam a T7 RNA polimerase, a indução é normalmente no final da fase logarítmica (DO₆₀₀ ~ 0,7-0,8). Após indução, as culturas são crescidas por mais 5-6 horas (RNA polimerase

bacteriana) ou 3-4 horas (T7 RNA polimerase). A seguir, as células bacterianas são coletadas por centrifugação, e o nível de expressão protéica é verificado por eletroforese em gel de poliacrilamida-SDS (4) de uma pequena amostra de bactérias induzidas. Um esquema do procedimento para expressão de proteínas recombinantes está mostrado na figura 10, e o nível de expressão na figura 11.

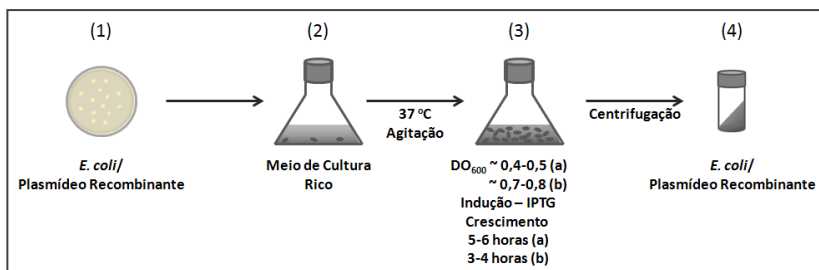


Figura 10. Esquema do procedimento para expressão de proteínas recombinantes. 1. Cultura de *E. coli*, transformadas com plasmídeo de expressão, em meio semi-sólido, 2. Inoculação e crescimento (37 °C), sob agitação, de algumas colônias em meio líquido rico, 3. Indução com IPTG ou lactose em DO_{600} ideal (a) – sistemas que utilizam RNA polimerase bacteriana, e (b) – sistemas que utilizam T7 RNA polimerase. 4. Coleta das bactérias por centrifugação.

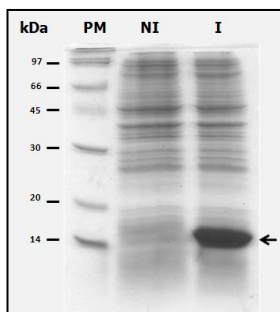


Figura 11. Eletroforese em gel de poliacrilamida-SDS de amostras de bactérias transformadas com plasmídeos de expressão. PM – Marcadores de Peso Molecular (mostradas ao lado em kDa), NI – Bactérias não induzidas, I – bactérias induzidas. Verifica-se a presença de uma banda (indicada por uma flecha) no poço I que não está presente no poço NI

5. Otimização das sequências codificadoras

A sequência gênica de interesse pode conter códons que não são muito utilizados pela *E. coli* (5, 6). O grau de frequência destes códons

pode diminuir, ou mesmo não haver expressão da proteína heteróloga. Os mesmos devem ser substituídos pelos códons sinônimos mais utilizados pela bactéria (índice de adaptação de códons, ou do inglês “codon adaptation index, CAI”) (7). Sequências gênicas podem ser analisadas e melhoradas através de “softwares” disponíveis em sítios na rede internacional de informática (8). Companhias de biotecnologia sintetizam genes com códons ideais para várias células hospedeiras de expressão.

6. Considerações Finais

Muitas proteínas têm atividades biológicas somente após sofrerem modificações pós-traducionais. Estas proteínas devem ser então produzidas em sistemas de expressão de eucariotos, como por exemplo, células de mamíferos, de insetos, entre outros.

7. Referências Bibliográficas

1. Lundberg, KS; Shoemaker, DD; Adams, MW; Short, JM; Sorge, JA; Mathur, EJ. 1991. High-fidelity amplification using a thermostable DNA polymerase isolated from *Pyrococcus furiosus*. *Gene*, **108**:1-6.
2. Temin, HM. 1976. The DNA provirus hypothesis. *Science*, **192**: 1075-80.
3. Studier, FW; Rosenberg, AH; Dunn, JJ; Dubendorff, JW. 1990. Use of T7 RNA polymerase to direct expression of cloned genes. *Meth. Enzymol.*, **185**:60-89.
4. Laemmli, UK. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**:680-85
5. Makrides, SC. 1996. Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microb. Review*, **60**:512-38.

6. Wada, K-n; Aota, S-i; Tsuchia, R; Ishibashi, F; Gojobori, T; Ikemura, T. 1990. Codon usage tabulated from the GenBank genetic sequence data. *Nucl. Ac. Res.*, **18**:2367-411
7. Sharp, PM and Li, W-H. 1987. The codon adaptation index – a measure of directional synonymous codon usage bias, and its potential applications. *Nucl. Ac. Res.*, 15:1281-95.
8. <http://genomes.urv.es/OPTIMIZER/>