



# A produção de insulina humana por ENGENHARIA GENÉTICA

Bactéria transgênica produz insulina humana

**Beatriz Dolabela de Lima**

Professora Adjunta - Departamento de  
Biologia Celular - Universidade de Brasília  
bdlima@unb.br

## A engenharia genética e a biotecnologia moderna

A tecnologia do DNA recombinante possibilita a obtenção de organismos com características novas ou não encontradas na natureza, o que permite uma nova alternativa para o melhoramento genético de espécies de valor biotecnológico. Desse modo, células de bactérias, leveduras e mesmo eucariontes superiores como plantas podem ser programadas com genes exógenos, abrindo a perspectiva de produção nestes organismos de polipeptídeos de interesse, como o interferon, o hormônio de crescimento, a insulina entre outros. A utilização de microrganismos “engenheirados”, capazes de sintetizar proteínas em grande quantidade, apresenta, sob o ponto de vista econômico,

uma vantagem considerável em relação aos processos clássicos de produção.

A extração de proteínas eucarióticas, como a insulina, requer grandes quantidades de matéria-prima (pâncreas suíno e bovino), que nem sempre estão disponíveis e são, geralmente, de elevado custo. Isso torna o processo extrativo cada vez mais oneroso. Nesse contexto, o emprego de técnicas mais eficientes, como a do DNA recombinante, abriu novas perspectivas de produção.

## Diabetes mellitus

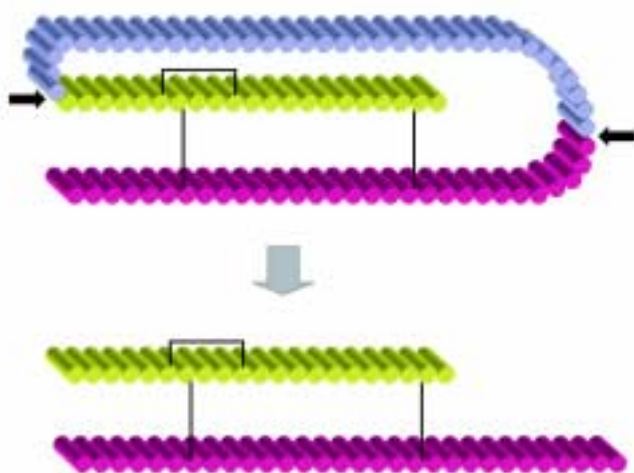
O *diabetes mellitus* é um grupo de doenças causadas pela deficiência na secreção ou na ação do hormônio pancreático insulina, o que produz profundas anormalidades no metabolismo. Há duas classes principais de *diabetes*: o juvenil e o adulto. No primeiro, a doença começa cedo, tornando-se severa e, no último, é lento para se desenvolver, moderado e, frequentemente, não reconhecido.

Os sintomas característicos do *diabetes* são sede excessiva e frequente micção, levando à ingestão de grandes volumes de água. Essas alterações são devidas à excreção de grandes quantidades de glicose na urina (glicosúria). O grande volume de urina no *diabetes* reflete a necessidade do rim de excretar uma certa quantidade de

água junto com a glicose, pois a capacidade do rim de concentrar os solutos na urina tem um limite máximo. Outro sintoma é o nível da glicose sanguínea e como é a resposta à ingestão da glicose. Quando a concentração da glicose no sangue é significativamente alta é chamada de hiperglicemia.

O pH do plasma sanguíneo de pessoas severamente diabéticas é frequentemente menor que o valor normal 7,4, condição esta chamada de acidose. É causada pela superprodução de ácidos metabólicos e o pH do sangue pode cair a 6,8 ou abaixo, e levar a lesões irreparáveis nos tecidos, e à morte. Esse aumento da acidez é um indicativo de alterações profundas no balanço ácido-base do organismo. A acidez aumentada é devido à extensa formação de corpos cetônicos no fígado e à sua liberação no sangue. Como os tecidos não conseguem utilizar a glicose sanguínea, o fígado tenta compensar essa deficiência aumentando a utilização dos ácidos graxos como combustível, mas isso provoca a superprodução de corpos cetônicos, além da capacidade de glicose oxidá-los. A atividade de glicocinase está diminuída no *diabetes* já que é a insulina que estimula a biossíntese dessa enzima. Como consequência, forma-se pouco glicogênio. Como os carboidratos não estão sendo utilizados, as proteínas do organismo são usadas como combustíveis. Os aminoácidos sofrem perda dos seus grupos amino e os acetoácidos formados podem sofrer oxidação em dióxido de carbono e água, em parte pela via do ciclo do ácido cítrico.

A administração de insulina para corrigir a deficiência endócrina e a administração de bicarbonato de sódio para corrigir a perda, tanto do sódio como da capacidade do tampão bicarbonato, podem trazer toda a química do organismo de volta para um balanço quase normal dentro de 12 a 24 horas. Para seguir o curso de tal tratamento, as dosagens de glicose, pH e CO<sub>2</sub> sanguí-



**Figura 1:** Estrutura da insulina humana: pró-insulina (cadeias B, C e A) e insulina (cadeias B e A). Setas pretas indicam o ponto de clivagem para a retirada da cadeia C da pró-insulina, originando a insulina ativa. Cadeia B, em rosa, cadeia C, em azul, e cadeia A, em amarelo

neos são realizadas frequentemente. Desse modo, o *diabetes* juvenil requer terapia com insulina e um cuidadoso controle, por toda a vida, do balanço entre a ingestão de glicose e a dose de insulina injetada (Lehninger, 1984).

### Insulina humana

A insulina humana é produzida nas células β pancreáticas, localizadas dentro dos conjuntos de células de 100 a 200 μm conhecidos como Ilhotas de Langerhans. Essas estão dispersas pelo pâncreas de muitos vertebrados superiores, constituindo cerca de 1% da massa do órgão (Steiner *et al.*, 1985). A insulina tem sido isolada de uma grande variedade de espécies de vertebrados, sendo que, em todas elas, a molécula é composta de duas cadeias polipeptídicas (A e B) ligadas por pontes dissulfídicas.

A insulina humana, como muitos hormônios protéicos, é sintetizada como uma proteína precursora maior, seguida de uma clivagem proteolítica, para gerar o hormônio ativo (Wang & Tsou, 1991). Desse modo, a insulina é produzida sob a forma de um único polipeptídeo, a pré-insulina, com uma cadeia de 110 aminoácidos. Os vinte e quatro primeiros aminoácidos formam o peptídeo sinal ou seqüência pré da proteína e têm a função de facilitar a entrada da mesma no retículo endoperiplasmático. Durante esse processo, o peptídeo sinal é separado da proteína, resultando na formação da pró-insulina (figura 1). Essa molécula resultante, na qual as cadeias A (21 aminoácidos) e B (30 aminoácidos) estão ligadas pelo peptídeo conectante C (35 aminoácidos), é a precursora da insulina. Ela adquire sua conformação com a formação de duas pontes dissulfídicas e é transportada para o aparelho de Golgi, onde vai ser empacotada em grânulos de estoque. Durante a formação e maturação dos grânulos secretórios, a pró-insulina é clivada por enzimas proteolíticas do tipo da tripsina, resultando na liberação do peptídeo C. Desse modo, as duas cadeias A e B estão ligadas entre si por pontes dissulfídicas, tendo uma outra ponte interna na cadeia A, formando a molécula de insulina (figura 1) (Steiner *et al.*, 1985).

|     |      |      |       |      |      |      |      |      |      |      |      |       |     |          |         |     |
|-----|------|------|-------|------|------|------|------|------|------|------|------|-------|-----|----------|---------|-----|
| 1   | TTT  | GTG  | AAC   | CAA  | CAC  | CTG  | TGC  | GGC  | TCA  | CAC  | CTG  | GTG   | GAA | GCT      | CTC     | 45  |
|     | ..C  | ..F  | ...G  | ...  | ...F | ..C  | ..T  | ...A | ...  | ...G |      |       |     |          |         |     |
| 1   | F    | V    | N     | G    | H    | L    | C    | G    | S    | H    | L    | V     | E   | A        | L       | 15  |
| 46  | TAC  | CTA  | GTG   | TGC  | GCG  | GAA  | CGA  | GGC  | TTG  | TTG  | TAC  | ACA   | CCC | AAG      | ACC     | 90  |
|     | ...G | ..T  | ...GT | ...T | ..T  | ...T | ...T | ...C | ..A  | ...  | ..G  | ..A   | ... | ...      | ...     |     |
| 16  | Y    | L    | V     | C    | G    | E    | R    | G    | F    | F    | Y    | T     | P   | K        | T       | 30  |
| 91  | CCC  | CCC  | GAG   | GCA  | GAG  | GAG  | CTG  | CAG  | GTG  | GGG  | CAG  | GTG   | GAG | CTG      | GGC     | 135 |
|     | ..T  | ..T  | ..A   | ..T  | ...A | ...  | ...T | ..T  | ..T  | ...T | ..T  | ..T   | ..A | ...T     | ...     |     |
| 31  | R    | R    | E     | A    | E    | D    | L    | Q    | V    | G    | Q    | Y     | E   | L        | G       | 45  |
| 136 | GGG  | GGC  | CCT   | GAT  | GCA  | GGC  | AGC  | CTG  | CAG  | CCC  | TTG  | GCC   | CTG | GAG      | GGG     | 180 |
|     | ..C  | ..T  | ..G   | ...  | ...T | ...A | ..G  | C..  | ..T  | ...A | ..T  | ...   | ..A | ..T      | ...     |     |
| 46  | G    | G    | P     | G    | A    | G    | S    | L    | Q    | P    | L    | A     | L   | E        | G       | 60  |
| 181 | TCC  | CTG  | CAG   | AAG  | CGT  | GCC  | ATT  | GTG  | GAA  | CAA  | TGC  | TGT   | ACC | AGC      | ATC     | 225 |
|     | ...T | ...G | ...A  | ...T | ..C  | ...T | ..G  | ...C | ..C  | ...C | ..T  | ...TC | ... | ...      | ...     |     |
| 61  | S    | L    | Q     | K    | R    | G    | I    | Y    | E    | G    | C    | T     | S   | I        | 75      |     |
| 226 | TCC  | TCC  | CTC   | TAC  | CAG  | CTG  | GAG  | AAC  | TAC  | TGC  | AAC  | TAG   | 258 | bp       | Hs cDNA |     |
|     | ...G | ...  | ...G  | ...A | ...  | ...A | ...  | ...  | ...A | ...  | ...A | ...   | ... | ...      | Ec gene |     |
| 76  | C    | G    | L     | Y    | Q    | L    | E    | H    | Y    | C    | N    | 60    | aa  | proteína |         |     |

**Figura 2:** Comparação entre a seqüência nucleotídica do gene sintético (Ec gene) e do cDNA para a pró-insulina humana (Hs cDNA). Proteín: seqüência protéica da pró-insulina

### Construção do gene sintético para a insulina humana

Dentre as várias maneiras pelas quais um gene eucariótico pode ser obtido para a expressão em procariotos, a síntese química oferece as seguintes vantagens: fornece diretamente a seqüência exata desejada; as seqüências codificadoras e não codificadoras podem ser desenhadas para a expressão procariótica; sítios de restrição podem ser removidos ou adicionados, íntrons retirados; não há necessidade da etapa de isolamento do mRNA ou DNA genômico e permite a alteração do gene de forma mais simples.

Desse modo, a construção do gene sintético para a pró-insulina humana foi iniciada a partir da seqüência de aminoácidos dessa proteína descrita por Sures *et al.* (1980). Utilizando-se os códons genéticos preferenciais para *Escherichia coli* (De Boer & Katelein, 1986) para a otimização da expressão do gene nessa bactéria, foi montado um gene codificando a proteína humana. A figura 2 mostra a comparação da seqüência de bases na fase de leitura do cDNA da pró-insulina humana (Bell *et al.*, 1979; Sures *et al.*, 1980) com o gene sintético montado nesse trabalho. Esse gene contém 18,6 % de bases substituídas, sendo 16,27%, na terceira base, 1,16%, na segunda base, e 1,16%, na primeira base dos códons. Nenhuma das modificações feitas nos códons alterou o aminoácido de correspondência. Após análise computacional, a presença de seqüências repetitivas, diretas ou inversas, maiores do que oito nucleotídeos não foi encontrada, evitando assim uma possível estrutura secundária local

indesejada nos mRNAs transcritos a serem produzidos a partir destes genes, o que poderia afetar sua expressão.

Para facilitar a clonagem, esse gene sintético para a pró-insulina humana foi predito contendo um sítio para a enzima de restrição *Eco* RI no início do gene, seguido de um códon para metionina (ATG), da região codante para a cadeia B, cadeia C e cadeia A, nesta ordem, do códon de término (TAA), e finalmente, de um sítio para a enzima de restrição *Bam* HI. Outros sítios para enzimas de restrição foram adicionados internamente no gene de modo a facilitar modificações posteriores.

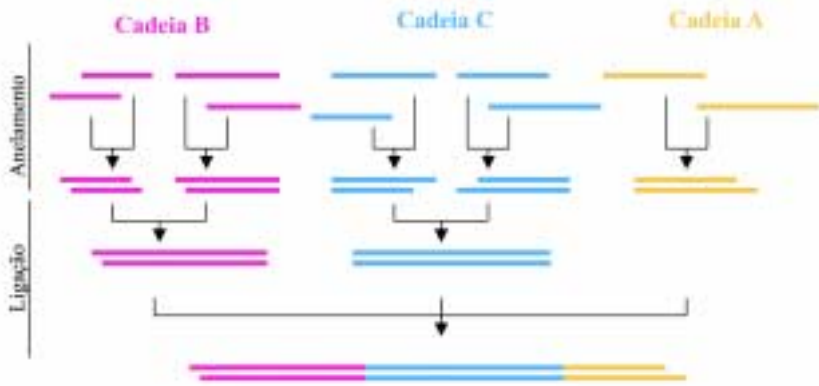
Após a definição da estrutura do gene sintético, foi realizada a sua montagem conforme a estratégia mostrada na figura 3. Quatro oligonucleotídeos foram sintetizados para a cadeia B, quatro para a cadeia C e dois para a cadeia A. Esses oligos foram anelados em pares para a formação da fita dupla de DNA e foram ligados entre si, em duas etapas. Na primeira, foram montadas as cadeias B e C e, na segunda, montado o gene da pró-insulina, sendo que nesta ordem: cadeia B, C e A. Esse gene foi clonado em um vetor para a realização do seqüenciamento de DNA e a confirmação da seqüência nucleotídica correta do gene.

Após a definição da estrutura do gene sintético, foi realizada a sua montagem conforme a estratégia mostrada na figura 3. Quatro oligonucleotídeos foram sintetizados para a cadeia B, quatro para a cadeia C e dois para a cadeia A. Esses oligos foram anelados em pares para a formação da fita dupla de DNA e foram ligados entre si, em duas etapas. Na primeira, foram montadas as cadeias B e C e, na segunda, montado o gene da pró-insulina, sendo que nesta ordem: cadeia B, C e A. Esse gene foi clonado em um vetor para a realização do seqüenciamento de DNA e a confirmação da seqüência nucleotídica correta do gene.

### Construção de um vetor de hiper-expressão para *E. coli*

Nos últimos anos, muitos vetores para *E. coli* têm sido construídos com diversas finalidades, entre essas, a clonagem de cDNAs, de fragmentos de DNA amplificados por PCR, transcrição *in vitro* e para a expressão e produção de proteínas heterólogas. Para cada finalidade, o vetor terá que apresentar determinadas características para otimizar sua utilização. Desse modo, para a construção de um sistema de expressão de proteínas em *E. coli*, seis elementos básicos são necessários:

- uma região necessária para replicação estável e controle do número de cópias;
- um marcador seletivo, como um gene conferindo resistência a antibiótico para a hospedeira;
- um promotor para iniciação da transcrição e seu controle;
- uma região terminadora da transcrição;



**Figura 3:** Estratégia utilizada para a construção dos genes para a pró-insulina humana para a expressão na bactéria *E. coli*

- um sítio de ligação de ribossomas para a iniciação da tradução em uma trinca ATG apropriada;

- e uma região de sítios apropriados para enzimas de restrição, para utilização nas clonagens dos genes a serem expressos.

Seguindo esses requisitos para a construção do vetor de expressão pLMT8.5 (figura 4), utilizamos a origem de replicação do plasmídeo pUC8, o gene de resistência à tetraciclina do plasmídeo RP4, o promotor pL isolado do fago Lambda, a região Shine Dalgarno sintética do fago T7, terminador de transcrição Rho-independente sintético, e uma região de múltiplos sítios únicos para enzimas de restrição (figura 4).

Para a região de replicação estável e de controle do número de cópias do plasmídeo foi escolhida a origem de replicação do plasmídeo pUC8. Essa origem é derivada do plasmídeo pMB1

(ou ColE1) onde mutações e deleções foram introduzidas de modo a aumentar o número de cópias do plasmídeo dentro da célula. Esses plasmídeos podem estar presentes em mais de 500 cópias por célula.

O gene de resistência à tetraciclina foi escolhido e utilizado porque este permite que, além da seleção das bactérias recombinantes, o antibiótico tetraciclina estará constantemente fazendo pressão seletiva durante a fermentação. A resistência à tetraciclina consiste de uma proteína que bombeia o antibiótico para fora da célula, não permitindo sua atuação na inibição da síntese protéica no interior da bactéria. Isso tem como consequência, em uma fermentação industrial, que se utilizem menores quantidades do antibiótico, pois este não será inativado pela bactéria resistente e estará constantemente presente na fermentação. Assim, a produtividade no bioreator

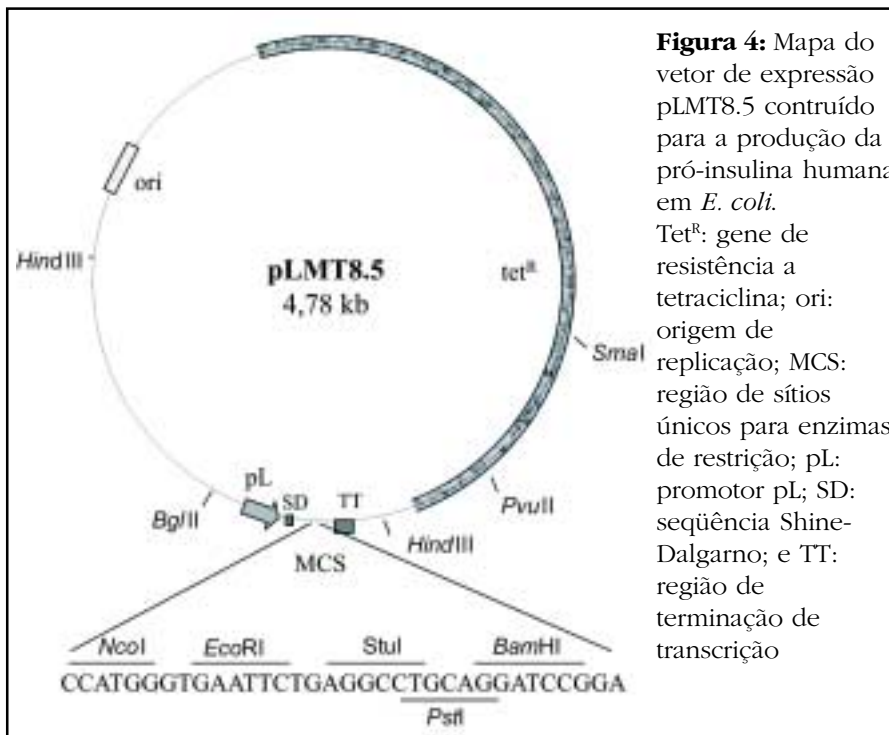
não será afetada devido ao acúmulo de bactérias sem plasmídeos. Esse fenômeno de acúmulo de bactérias sem plasmídeos foi observado por muitos pesquisadores e é responsável por dificuldades no escalonamento requerido para a comercialização de produtos recombinantes (Kumar *et al.*, 1991).

Para a iniciação da transcrição foi escolhido o promotor pL do bacteriófago Lambda. Este é um promotor forte e é regulado negativamente pelo repressor codificado pelo gene cI. A mutação cI<sup>857</sup> tornou o repressor sensível à temperatura, funcional a 28°C e não funcional a 42°C. Assim, a expressão de seqüências codantes sob o controle desses promotores e pelo repressor cI<sup>857</sup> pode ser ativada simplesmente por uma mudança na temperatura da cultura, o que, principalmente em escala industrial, simplifica e barateia o processo.

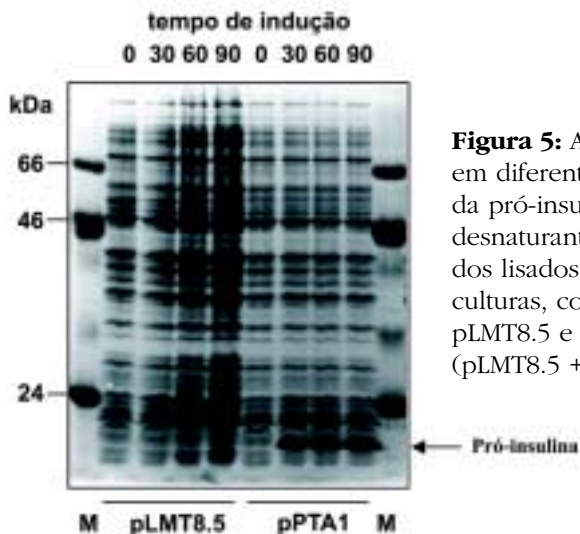
A partir de um promotor e de um início de transcrição, a RNA polimerase transcreve o RNA até o reconhecimento de um sítio de terminação, que funciona como um sinal para o término da transcrição. Dois tipos diferentes de terminadores são encontrados: terminadores simples ou Rho-independentes e terminadores Rho-dependentes (dependentes da proteína Rho). Os terminadores Rho-independentes possuem duas características: uma estrutura em forma de grampo, gerada por pareamento entre repetições invertidas, contendo uma região rica em pares G-C na base do grampo, separadas por uma curta distância, e outra, subsequente da primeira, que contém uma região de, aproximadamente, seis resíduos de uridina no final da unidade. Esses terminadores parecem ter alguma função na proteção do mRNA, pela sua capacidade de formar uma estrutura em grampo, provavelmente fornecendo uma barreira contra a ação de uma exonuclease 3' (Brawerman, 1987), e não dependem de um fator protéico, proporcionando uma efetiva finalização da transcrição no vetor de expressão.

A etapa mais limitante na síntese protéica é a ligação dos ribossomas às moléculas de mRNA. Desde que o número de ribossomas na célula exceda a classe de mensageiros, uma maneira de aumentar a expressão de um gene clonado é aumentar o número dos transcritos correspondentes. Para tal, a maneira mais simples é clonar o gene de interesse em um plasmídeo com grande número de cópias.

As duas características dos mRNAs de *E. coli*, que determinam onde e qual a eficiência do processo de reconheci-



**Figura 4:** Mapa do vetor de expressão pLMT8.5 construído para a produção da pró-insulina humana em *E. coli*. Tet<sup>R</sup>: gene de resistência a tetraciclina; ori: origem de replicação; MCS: região de sítios únicos para enzimas de restrição; pL: promotor pL; SD: seqüência Shine-Dalgarno; e TT: região de terminação de transcrição



**Figura 5:** Análise da expressão em diferentes tempos de indução da pró-insulina em *E. coli*. Gel desnaturante (SDS-PAGE) a 15% dos lisados protéicos totais das culturas, contendo o plasmídeo pLMT8.5 e o plasmídeo pPTA1 (pLMT8.5 + gene da pró-insulina)

mento ribossoma-mRNA, são o códon de iniciação, geralmente AUG ou GUG, e a região Shine Dalgarno (SD), ou sítio de ligação de ribossomas (RBS). Essa região é complementar ao final 3' do RNA ribossomal 16S (Shine & Dalgarno, 1974; Shultzaberger *et al.*, 2001). O SD tem tipicamente de três a seis nucleotídeos de tamanho e está localizado quatro a quinze nucleotídeos antes do códon de iniciação. Essa região tem a função de guiar o ribossoma para o ponto de início correto (Dreyfus, 1988). Olins *et al.* (1988) descreveram uma região SD a montante do gene 10 do bacteriófago T7 (g 10-L). Esse gene codifica para a proteína da capa, que é a proteína mais sintetizada após a infecção do fago T7. Essa região SD possui um grande potencial para otimizar a eficiência da tradução de genes homólogos e heterólogos clonados em *E. coli*.

Para facilitar a clonagem de genes para a expressão no vetor, uma região apropriada de sítios únicos para enzimas de restrição foi adicionada entre a região promotora e o SD, e a região terminadora da transcrição, onde temos o sítio para a enzima de restrição *NcoI* contendo o códon ATG de início de tradução (figura 4).

### Produção da pró-insulina humana

Para a produção da pró-insulina, o gene sintético da pró-insulina foi subclonado no vetor de expressão pLMT8.5 e o plasmídeo recombinante foi denominado de pPTA1. Esse plasmídeo foi utilizado para a transformação da cepa *E. coli* N4830-1, pois essa possui o repressor *ci* termossensível.

A cepa *E. coli* N4830-1 transformada

com o plasmídeo pPTA1, foi utilizada para a indução térmica da Pró-insulina. Como controle, foi utilizada cultura da cepa contendo o plasmídeo pLMT8.5 sem o gene clonado. As proteínas induzidas estão mostradas na figura 5. Verifica-se a indução de uma proteína de, aproximadamente, 10.000 daltos, que corresponde à Pró-insulina na cultura do recombinante pPTA1. Essa proteína corresponde a 20% das proteínas totais da bactéria. Nas amostras com o plasmídeo pLMT8.5, não observamos essa proteína induzida. Podemos concluir portanto, que ocorreu uma hiperexpressão da Pró-insulina a partir do gene clonado no plasmídeo pLMT8.5 construído.

Verificou-se também que ocorreu formação de corpos de inclusão da proteína recombinante produzida, o que auxilia a proteção contra proteólise e facilita sua purificação. Eles são formados como agregados citoplasmáticos que podem ser purificados após a lise da célula, centrifugação e solubilização das proteínas com, por exemplo, uréia ou guanidina.

A perspectiva de produção a partir dessas bactérias recombinantes, obtidas nesse trabalho, nas condições de cultura utilizadas, é de 1 g de pró-insulina para cada 17 litros de cultura, com uma densidade celular de  $10^9$  células por ml.

### Considerações finais

Esse trabalho foi realizado dentro do Projeto "Produção de insulina humana através de precursores recombinantes em *Escherichia coli*", coordenado pelo Prof. Spartaco Astolfi Filho, com financiamento da Biobrás e PADCT/FINEP, dentro do Convênio UnB/Biobrás. Esse

sistema de produção de insulina está patenteado nos Estados Unidos (patentes nos. 6068993, 30/05/2000, e 6281329B1, 28/08/2001) e com pedido de patente no INPI, Brasil.

### Bibliografia

- Bell, G.I.; Swain, W.F.; Pictet, R.; Cordell, B.; Tischer, E. & Goodman, H.M. 1979. Nucleotide sequence of a cDNA clone encoding human preproinsulin. *Nature*, 282, 525-7.
- Brawerman, G. 1987. Determinants of messenger RNA stability. *Cell*, 48: 5-6.
- De Boer, H.A. & Kastelein, R.A. 1986. Biased codon usage: an exploration of its role in optimization of translation. In: *Maximizing gene expression*; Reznikoff, W. & Gold, L.; eds. Butterworth Publishers, Stoneham, USA.
- Dreyfus, M. 1988. What constitutes the signal for the initiation of protein synthesis on *Escherichia coli* mRNAs? *J. Mol. Biol.*, 204: 79-94.
- Kumar, P.K.P.; Maschke, H.E.; Friehs, K. & Schugerl, K. 1991. Strategies for improving plasmid stability in genetically modified bacteria in bioreactors. *Tibtech*, 9: 279-84.
- Lehninger, A.L. 1984. Princípios de Bioquímica. São Paulo, Savier.
- Olins, P.O.; Devine, C.S.; Rangwala, S.H. & Kavka, K.S. 1988. The T7 phage gene 10 leader RNA, a ribosome-binding site that dramatically enhances the expression of foreign genes in *Escherichia coli*. *Gene*, 73: 227-35.
- Shine, J. & Dalgarno, L. 1974. The 3'-terminal sequence of *E. coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proc. Natl. Acad. Sci. USA*, 71: 1342-6.
- Shultzaberger R.K., Bucheimer R.E., Rudd K.E. & Schneider T.D. 2001. Anatomy of *Escherichia coli* Ribosome Binding Sites. *Journal of Molecular Biology*, 313: 215-28.
- Steiner, D.F.; Chan, S.J.; Welsh, J.M. & Kwoh, S.C.M. 1985. Structure and evolution of the insulin gene. *Ann. Rev. Genet.*, 19: 463-84.
- Sures, I.; Goeddel, D.V.; Gray, A. & Ullrich, A. 1980. Nucleotide sequence of human preproinsulin complementary DNA. *Science*, 208: 57-9.
- Wang, C.C. & Tsou, C.-L. 1991. The insulin A and B chains contain sufficient structural information to form the native molecule. *TIBS*, 16: 279-81.