



SEQUENCIAMENTO DE DNA: DECIFRANDO O MANUAL DE INSTRUÇÕES DOS SERES VIVOS

André Luis Fachini de Souza¹, Liziane Cristina Campos Brusamarello²

1- Universidade do Estado de Santa Catarina – UDESC, Departamento de Solos e Recursos Naturais. Av. Luiz de Camões, 2090, Lages – SC, CEP 88520-000, a2alfs@cav.udesc.br

2- Universidade Estadual de Ponta Grossa – UEPG, Departamento de Biologia Estrutural Molecular e Genética. Av. General Carlos Cavalcanti, 4748, Ponta Grossa – PR, CEP 84010-919, lizianebrusa@yahoo.com.br

a2alfs@cav.udesc.br; lizianebrusa@yahoo.com.br

Palavras-chave: DNA, sequenciamento, dideoxonucleotídeos, pirosequenciamento.

Sequenciamento de DNA

O sequenciamento do ácido desoxiribonucleico (DNA) foi descrito primeiramente por Sanger et al. em 1977 e é considerado um dos feitos científicos mais importantes da ciência. O DNA representa uma espécie de central de informações das células, sejam elas de animais, vegetais ou de microrganismos. É como se fosse uma torre de controle que envia as informações para que a célula execute suas tarefas. A unidade fundamental de informação nos seres vivos é o gene, um segmento de DNA capaz de codificar a informação necessária para a síntese de um produto biologicamente ativo, como uma proteína, que pode estar envolvida em processos de desenvolvimento e metabolismo. Também é atribuído ao DNA a hereditariedade e a variabilidade necessária para o processo evolutivo. A compreensão de como a informação é armazenada e utilizada nas células tem representado um caminho para a solução de problemas relacionados à estrutura e função das células. Um avanço importante para a genética humana foi a descoberta e isolamento do gene relacionado com a fibrose cística, uma doença genética caracterizada pelo funcionamento anormal das glândulas produtoras do muco, lágrima, suor, saliva e suco gástrico. Em 1989 Francis Collins identificou esse gene e analisando a sequência de nucleotídeos e a de aminoácidos por ele codificada, foi capaz de postular que se tratava de uma proteína transmembrana cuja ausência levava à secreção anormal de íons cloro nos tratos respiratórios e gastrointestinal.

A geração de conhecimento em torno dos genes que compõem o genoma dos mais diversos seres vivos é possível por meio da determinação das sequências de nucleotídeos do DNA. Em termos moleculares, a determinação sequencial de nucleotídeos de um gene pode ser

usada para deduzir sua função, comparando-a com as sequências de genes de função conhecida depositadas em bancos públicos.

Para decifrar o genoma humano, os cientistas do Projeto Genoma tinham o enorme desafio de sequenciar os quase 3 bilhões de pares de bases dos cromossomos e organizar os quase 20 mil genes do DNA humano. Com o aumento dos investimentos na área da genômica, o número de genomas sequenciados tem crescido bastante nos últimos anos. Estimativas de janeiro de 2008 indicam cerca de 700 genomas completos sequenciados (www.genomeonline.org). Estes números incluem genomas de procariotos, archaea e eucariotos. Entre os genomas bacterianos, cerca de 50% deles são de bactérias de interesse médico.

Para o sequenciamento de qualquer fragmento de DNA pelas técnicas tradicionais, como o sequenciamento automatizado de DNA baseado no método descrito por Sanger et al (1977), este fragmento precisa ser primeiramente inserido (ou clonado) em um vetor de clonagem, originando um DNA recombinante (Figura 2). O vetor de clonagem é geralmente uma molécula de DNA circular que serve como um veículo que transporta o gene de estudo para o interior de uma célula hospedeira, que é geralmente uma bactéria. O vetor de clonagem possui elementos que permitem sua autorreplicação bem como alguns genes marcadores (resistência a antibióticos) que servem para seleção das bactérias que receberam o vetor. Dentro da célula hospedeira o vetor multiplica-se, produzindo várias cópias idênticas de si próprio e do fragmento de DNA (inserto) ligado a ele. Após várias divisões celulares cada célula contém uma ou mais cópias da molécula de DNA recombinante. Este DNA é isolado e está pronto para ser sequenciado.

Para o sequenciamento do genoma inteiro de um organismo, o DNA precisa ser isolado e “quebrado” em pequenos fragmentos de aproximadamente 1000 pb. Es-

ses fragmentos são inseridos em vetores de clonagem, seguido pela inserção desses vetores em bactérias hospedeiras, constituindo o que chamamos de bibliotecas, onde cada vetor possui um fragmento de DNA diferente e seu conjunto constitui o genoma inteiro. Esse passo é crucial para o sucesso do sequenciamento e é feito manualmente. Após a obtenção da biblioteca, os fragmentos inseridos em cada vetor são sequenciados.

O sequenciamento de DNA pode ser feito por procedimentos químicos (originalmente desenvolvidos por Alan Maxam e Walter Gilbert) ou procedimento enzimático desenvolvido por Frederick Sanger e frequentemente chamado de método dos didesoxinucleotídeos. O princípio geral desta técnica é efetuar a síntese de uma fita de DNA “marcada” complementar à fita da qual se deseja determinar a sequência.

O primeiro passo é a obtenção de fitas simples do DNA que se deseja sequenciar. Essas fitas podem ser separadas pelo aquecimento que quebra as pontes de hidrogênio entre as bases do DNA e que mantém as fitas unidas. Com as fitas separadas, liga-se um oligonucleotídeo sintético (primer) à uma região pré-determinada de uma das fitas do vetor de clonagem (pela complementariedade das bases nitrogenadas, onde A realiza pontes de hidrogênio com T e C com G), próximo ao sítio onde foi inserido o fragmento de DNA em análise (Figura 3A). Esse oligonucleotídeo atua como uma sequência iniciadora, a qual é alongada de maneira complementar à fita (DNA molde) na qual o iniciador está ligado. O alongamento do iniciador é um processo enzimático. A enzima DNA polimerase adiciona didesoxinucleotídeos (dNTPs) sequencialmente na extremidade 3' do iniciador (primer), complementar ao DNA molde (Figura 3A).

Todo esse processo descrito é realizado *in vitro* (fora de um organismo vivo), em um equipamento chamado Termociclador por meio de uma reação chamada polymerase chain reaction – PCR (reação em cadeia da polimerase) (veja o artigo Amplificação de DNA, 01.02, 63-65; 2006; www.sbg.org.br). Na prática, ao tubo de reação, além do DNA a ser sequenciado, adiciona-se a enzima DNA polimerase, o iniciador adequado e os quatro didesoxinucleotídeos (dNTPs):- ATP, TTP, CTP e GTP. A mistura de reação também inclui uma pequena quantidade de didesoxinucleotídeos (ddNTPs) - (Figura 1) - “marcados” com um marcador fluorescente. A marcação fluorescente atribui uma cor diferente para cada um dos quatro didesoxinucleotídeos (Figura 3A).

A DNA polimerase consegue adicionar sequencialmente, de acordo com a complementariedade de bases, os dNTPs na extremidade 3' do iniciador. Quando um análogo didesoxinucleotídeo é incorporado ao poli-

nucleotídeo crescente no lugar de um desoxinucleotídeo normal (dNTP), o crescimento da cadeia é terminada, pois a adição do próximo nucleotídeo requer uma hidroxila (OH) 3' livre, que não está presente nos didesoxinucleotídeos que não possuem hidroxilas nas posições 2' e 3' (Figura 1).

Os vários fragmentos de DNA gerados contendo em uma das extremidades um ddNTP “colorido” são separados de acordo com seus tamanhos por eletroforese em gel (Figura 3B). A eletroforese é uma técnica de separação de moléculas onde partículas de vários tamanhos migrarão diferentemente em um gel quando aplicada uma diferença de potencial. Moléculas de massa menor irão migrar mais rapidamente que as de maior massa. De acordo com a concentração na qual o gel é preparado, é possível a separação de moléculas de DNA que diferem em apenas um nucleotídeo (veja o artigo Eletroforese de ácidos nucléicos: uma prática para o ensino de genética, Genética na Escola 03.01, 43-48, 2008).

Conforme os fragmentos de DNA migram diferentemente no gel e são separados, a base terminal de cada fragmento é identificada pela sua fluorescência característica. A fluorescência é identificada assim que o didesoxinucleotídeo passa por um feixe de laser e esta informação é enviada diretamente para um computador que a converte em um gráfico de picos (Figura 3B). Assim, a cor correspondente de cada um dos quatro didesoxinucleotídeos é identificada, atribuindo-se a cada uma delas uma letra (A, T, C ou G). Utilizando detectores de fluorescência controlados por computador, os sistemas automáticos podem identificar aproximadamente 10.000 bases por dia.

Uma vez determinada a sequência de nucleotídeos de um gene, ela precisa ser analisada. Para isso ela é comparada (alinhada) com sequências de genes já estudadas depositadas em bancos públicos que reúnem sequências de DNA e de aminoácidos e estruturas de proteínas. Os principais bancos são o GenBank, o EBI (European Bioinformatics Institute), o DDBJ (DNA Data Bank of Japan) e o PDB (Protein Data Bank). O processo de alinhamento de sequências busca determinar o grau de similaridade entre duas ou mais sequências, ou a similaridade entre fragmentos das mesmas. Esta etapa é realizada através de programas computacionais, específicos para esse tipo de análise, como por exemplo o programa BLAST, disponível em sítios da internet.

Pirosequenciamento

O pirosequenciamento é uma técnica de sequenciamento baseada na detecção de pirofostafa (PPi) libe-

rado durante a síntese de DNA. Este PPI liberado é convertido em uma molécula de Adenosina Trifosfato (ATP) pela enzima **ATP Sulfurilase**, e a enzima **Luciferase** utiliza essa molécula de ATP para oxidar Luciferina e produzir fóton de luz (Figura 4). Os fótons de luz produzidos são detectados por aparelhos e a quantidade de luz transmitida é proporcional ao número de nucleotídeos incorporados. Durante a reação de sequenciamento, os nucleotídeos (A, T, C ou G) são adicionados à cadeia de DNA crescente. Se nenhum fóton de luz for produzido, significa que a base adicionada não corresponde à base complementar da fita de DNA molde a ser sequenciada. Mas, quando o fóton é produzido, sabe-se qual base foi adicionada na reação e a quantidade de luz emitida corresponde ao número de vezes que a base se repete (Ronaghi, 2001; Mardis, 2008). Assim, é obtida a sequência de bases de nucleotídeos do DNA. Uma operação de sequenciamento com duração de 7 horas, utilizando equipamentos específicos, pode produzir um total de 100 milhões de pares de bases utilizando a técnica do pirosequenciamento. Enquanto que, no mesmo intervalo de tempo, o método de didesoxinucleotídeos produz apenas 440 mil pares de bases (Mardis, 2007).

Considerações finais

Novas tecnologias têm sido desenvolvidas para racionalizar e, sobretudo, diminuir o tempo de sequenciamento de genomas. Dentre estas técnicas destaca-se o pirosequenciamento, capaz de sequenciar milhões de pares de bases (inclusive de genomas inteiros) em questão de horas, realizado por apenas uma pessoa e com o auxílio de equipamento específico.

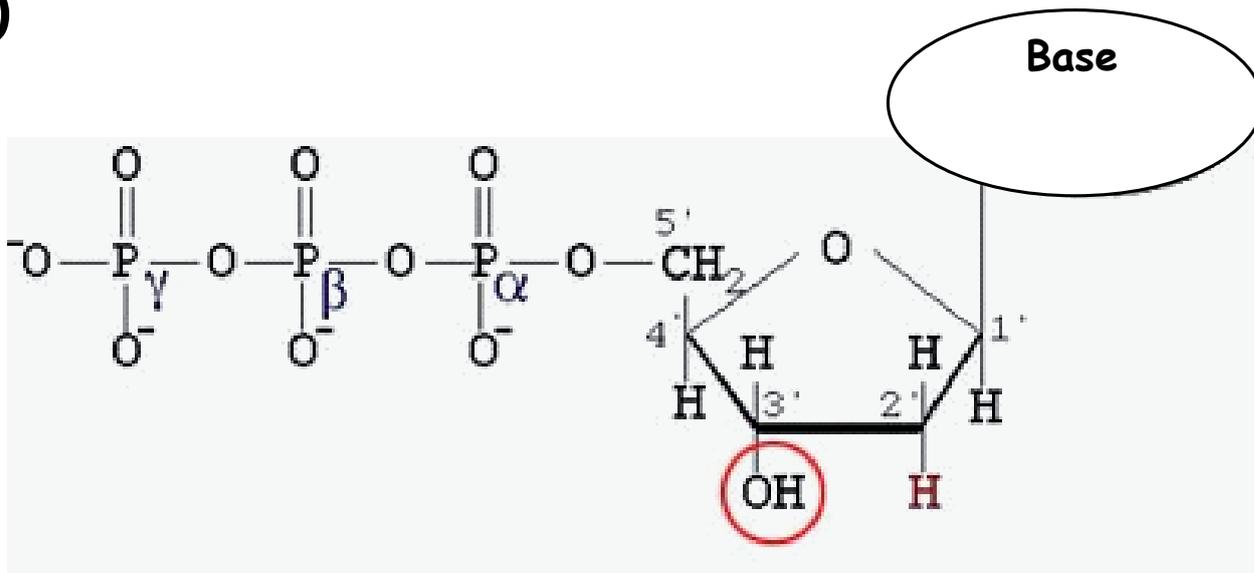
Apesar do domínio das técnicas de sequenciamento de DNA terem revolucionado o estudo da biologia e da medicina, o conhecimento da informação contida no gene não é o bastante para solucionar muitos dos problemas que permeiam essas áreas do conhecimento. Surge então a necessidade da genômica funcional, uma nova

fase na era pós-genômica que, a partir do conhecimento da sequência de um genoma, tem-se a necessidade de interpretar a função desta sequência. Neste sentido, novos conhecimentos estão sendo gerados e novas técnicas têm sido desenvolvidas com o intuito de atender a demanda que a pesquisa científica exige.

Referências

- Glick, B.R. e Pasternak, J.J. (1998) *Molecular biotechnology, principles and Applications of recombinant DNA*. 2 ed. Washington: ASM Press.
- Lehninger, A.L.; Nelson, D.L. e Cox, M.M. (2000) *Princípios de bioquímica*. 2 ed. São Paulo: SARVIER.
- Mardis, R. E. (2007) The impact of next-generation sequencing technology on genetics. *Trends in Genetics*. 24: 133-141
- Mardis, R. E. (2008) Next-Generation DNA Sequencing Methods. *Annual Review of Genomics Human Genetic*. 9: 387-402.
- Old, R.W. e Primrose, S.B. (1994) *Principles of gene manipulation*. 5 ed. Berlin: Blackwell Science.
- Prosdocimi, F.; Cerqueira, G.C.; Binneck, E.; Silva, A.F.; Reis, A.N.; Junqueira, A.C.M.; Santos, A.C.F.; Nbani, A.; Wust, C.I.; Filho, F.C.; Kessedjian, J.L.; Petretski, J.H.; Camargo, L.P.; Ferreira, R.G.M.; Lima, R.P.; Pereira, R.M.; Jardim, S.; Sampaio, V.S. e Flatschart, A.V.F. (2002) *Bioinformática: manual do usuário*. *Biociência e Desenvolvimento*, 29:12-25.
- Ronaghi, M. (2001) Pyrosequencing Sheds Light on DNA Sequencing. *Genome Research*. 11:3-11.
- Sanger, F. Nicklen, S. Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors *Proceedings of the National Academy of Sciences* 74:5463-5467.
- Simpson, A. J.G. (2000) Genome Sequencing networks. *Nature Reviews Genetics* 2:979-983.
- Voet, D.; Voet, J.G. e Pratt, C.W. (2000) *Fundamentos de bioquímica*. Porto Alegre: Artes Médicas Sul.

(A)



(B)

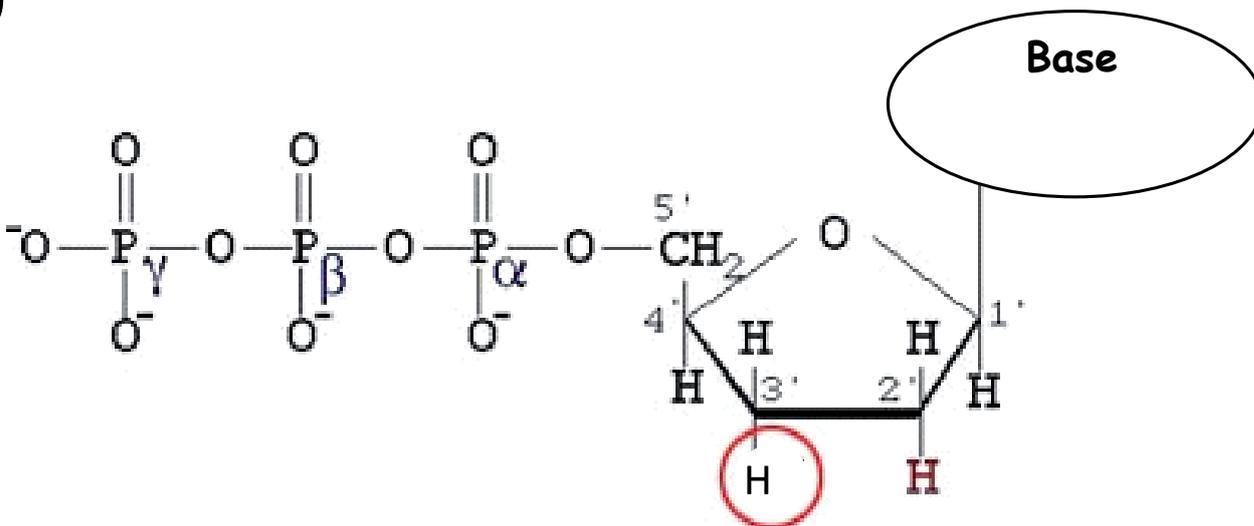


Figura 1. Estrutura dos desoxinucleotídeos e dideoxinucleotídeos. (A) desoxinucleotídeos-dNTPs, não apresentam a hidroxila (OH) no carbono 2' da ribose. (B) dideoxinucleotídeos-ddNTPs, não apresentam as hidroxilas nos carbonos 2' e 3' da ribose.

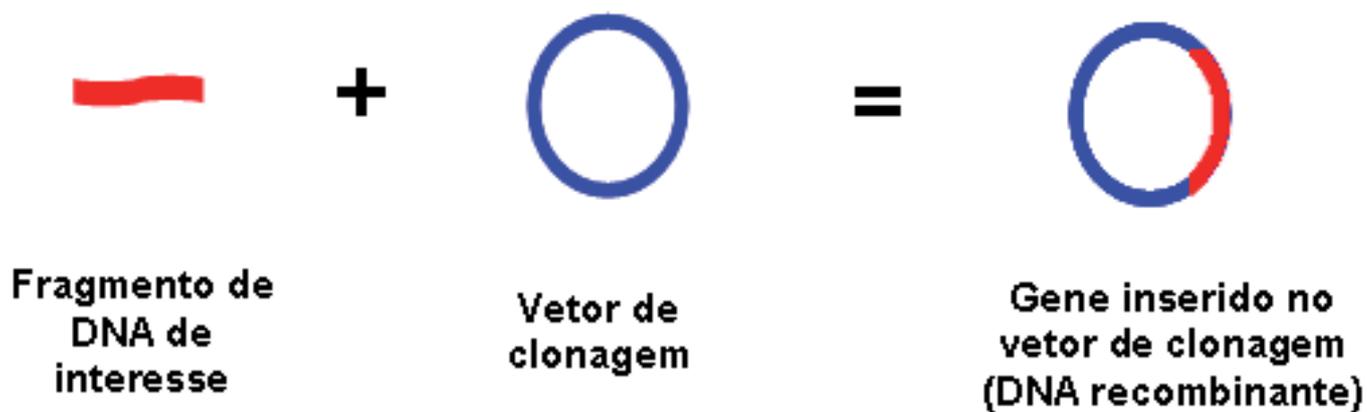


Figura 2. Esquema de inserção do gene a ser sequenciado em um vetor de clonagem. Esse processo ocorre com o auxílio de enzimas de restrição. Essas enzimas reconhecem regiões específicas no DNA e cortam essas regiões que, posteriormente, são ligadas ao vetor de clonagem também cortado com enzimas de restrição.

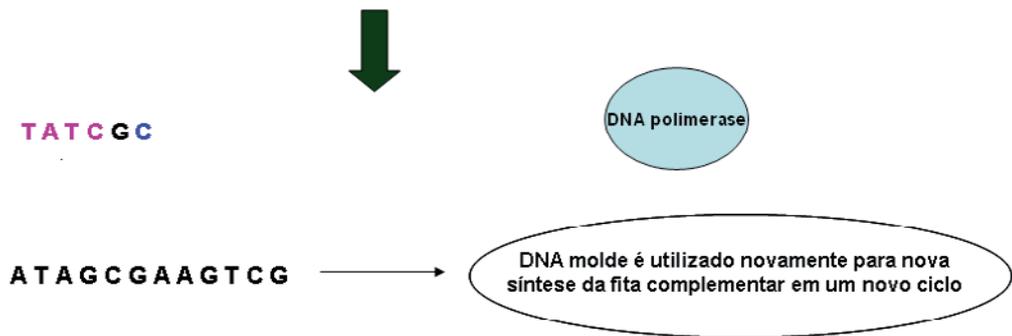
A) Em baixas temperaturas oligonucleotídeo iniciador é ligado ao DNA molde complementar



DNA polimerase inicia a síntese



Aumento da temperatura quebra pontes de H e libera as fitas molde e complementar



Processo acima descrito se repete várias vezes utilizando aparelho de PCR que aumenta e diminui temperatura da reação produzindo vários fragmentos de tamanhos diferentes com base marcada na extremidade

- 5' ATAGC 3'
 - 5' ATAGCG 3'
 - 5' ATAGCGA 3'
 - 5' ATAGCGAA 3'
 - 5' ATAGCGAAG 3'
 - 5' ATAGCGAAGT 3'
 - 5' ATAGCGAAGTC 3'
 - 5' ATAGCGAAGTCG 3'
- ⏟
 Oligonucleotídeo iniciador

B) Fragmentos são submetidos a um aparelho que os separa por tamanho, através da técnica de eletroforese e identifica a fluorescência emitida pela base nitrogenada

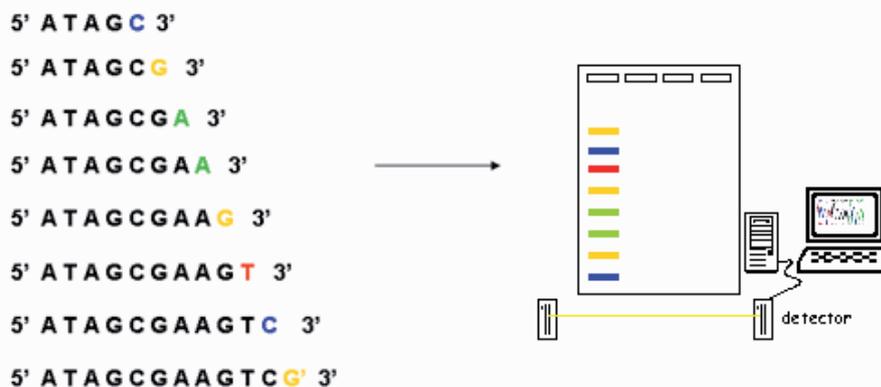


Figura 3. Método de sequenciamento usando didesoxinucleotídeos coloridos (ddNTPs). (A) Reação de sequenciamento. Primeiramente as fitas de DNA são separadas pelo aquecimento. Com as fitas separadas, o oligonucleotídeo iniciador (primer) se liga a uma região específica em uma das fitas. A enzima DNA polimerase alonga o iniciador inserindo bases complementares à fita na qual o iniciador está ligado. Sempre que a DNA polimerase inserir um didesoxinucleotídeo (ddNTP) colorido, a reação para porque a enzima é incapaz de adicionar um outro nucleotídeo na sequência. (B) Sequenciamento utilizando aparelho que realiza eletroforese e detecção de didesoxinucleotídeos marcados. Uma vez formadas várias fitas de DNA com diferença de um nucleotídeo de uma para a outra, essas fitas são separadas por eletroforese. As fitas menores migram mais rapidamente (de cima para baixo) e, assim que as fitas contendo em uma das extremidades um didesoxinucleotídeo colorido, a cor é identificada e cada cor corresponde a uma das quatro bases nitrogenadas que caracterizam o DNA.

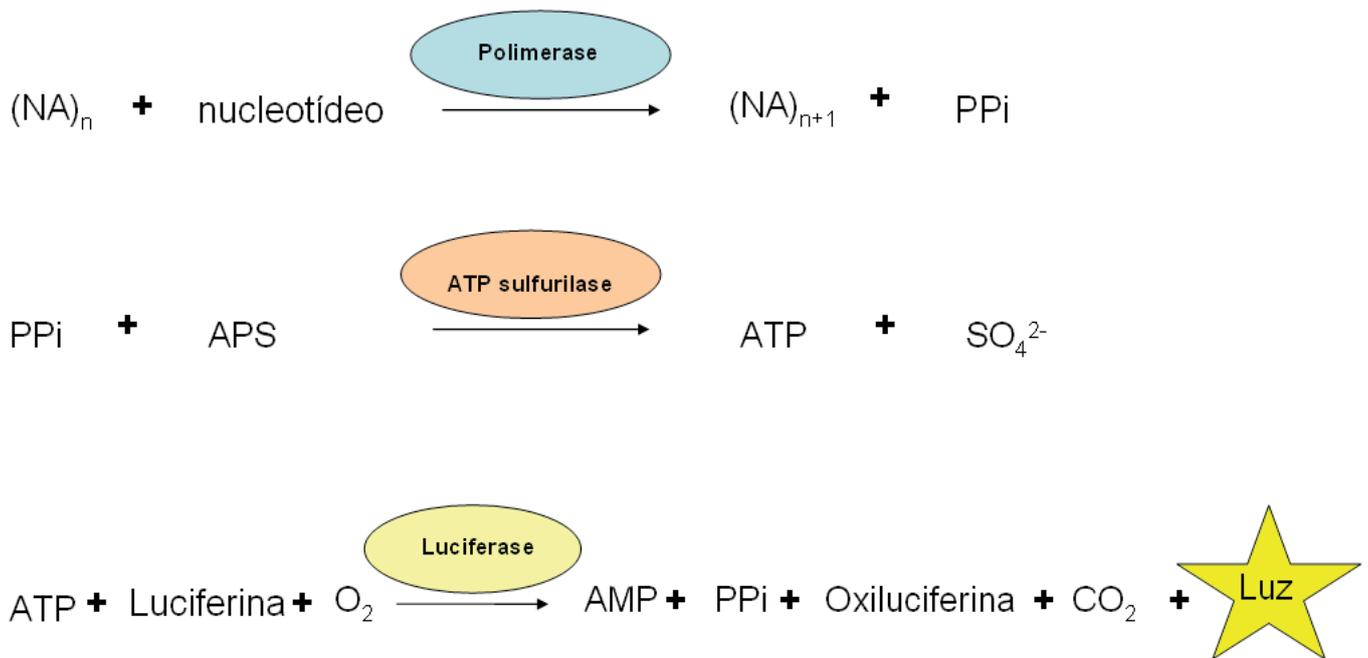


Figura 4. Princípio geral das reações do pirosequenciamento. A enzima Polimerase catalisa a reação de adição de um nucleotídeo a uma cadeia de ácido ribonucléico (RNA). Uma molécula de pirofosfato (PPi) é liberada como produto dessa reação e, logo em seguida, convertida em ATP pela união com adenosina fosfossulfato (APS) através da atividade de ATP sulfurilase. A oxidação de luciferina pela Luciferase produz fóton de luz.