Nikolas Paparidis Ferreira dos Santos

Obtenção das quinases dependentes de ciclinas CDK9 e CDK11 humanas
utilizando um sistema bacteriano de expressão (*E. coli*)

Dissertação apresentada ao Instituto de Química de São
Carlos da Universidade de São Paulo como parte dos
requisitos para obtenção do título de Mestre em
Ciências

Área de concentração: Química Orgânica e Biológica

Orientadora: Prof.^a Dr.^a Fernanda Canduri

Exemplar revisado

O exemplar original encontra-se em
acervo reservado na Biblioteca do IQSC-USP

São Carlos

2015

DEDICATÓRIA

*Dedico este trabalho à minha mãe, Tina,
e ao meu avô, Newton, que com seu amor, paciência,
e apoio incondicionais tornaram tudo possível.*

AGRADECIMENTOS

- À minha orientadora, Prof.^a Fernanda Canduri, pela oportunidade de trabalho, paciência, e confiança em mim depositada.
- A todos os colegas do meu grupo e de outros grupos de pesquisa, que contribuíram para a realização deste trabalho, especialmente à Cíntia, Bruna, Nathália, Juliana, Buana e Reetesh pela ajuda e troca de experiências.
- À minha família, por todo o incentivo e apoio com os quais sempre pude contar.
- Aos amigos queridos que, de perto ou de longe, estiveram ao meu lado em todos os momentos da jornada.
- Às agências de fomento CNPq e Fapesp, pela bolsa de auxílio e financiamento do projeto de pesquisa, respectivamente.

EPÍGRAFE

*“Não foi sem intenção que a engenhosa Natureza
uniu a doçura do mel ao ferrão da abelha”*

– Baltasar Gracián, *A Arte da Prudência*, século XVII.

RESUMO

As CDKs 9 e 11 fazem parte da subfamília de CDKs transcricionais, e portanto desempenham papéis de controle na dinâmica atividade de síntese e processamento do RNA mensageiro pela RNA Polimerase II. Devido à sua capacidade de modular individualmente a atividade dos complexos de transcrição da RNAP II, estas quinases assumem uma grande importância para a regulação da expressão gênica em células eucarióticas. Casos de desregulação da atividade de CDKs transcricionais têm sido frequentemente relacionados a diversas patologias humanas graves, incluindo vários tipos de câncer e também a AIDS (em razão do papel essencial desempenhado pela CDK9 na replicação do HIV). O objetivo deste trabalho é a obtenção das CDKs humanas 9 e 11 através da expressão heteróloga em *Escherichia coli*, buscando o aperfeiçoamento de métodos para produzir estas enzimas em quantidade e pureza suficientes para aplicação em estudos estruturais. A utilização de um sistema bacteriano oferece muitas vantagens práticas, como a simplicidade da técnica, baixo custo, altos níveis de expressão, e elevado rendimento na purificação do produto. No entanto, a obtenção de quinases humanas ativas a partir de *E. coli* sempre representou um desafio experimental, devido aos problemas comumente associados à expressão heteróloga. Os resultados apresentados aqui demonstram a possibilidade de se obter a CDK9 humana enzimaticamente ativa pelo reenovelamento *in vitro* da proteína expressa pela bactéria na forma de corpos de inclusão, após etapas de solubilização e purificação em condições desnaturantes. Por outro lado, a CDK11 humana só pôde ser obtida em uma versão encurtada, consistindo apenas no seu domínio quinase, devido à forte inibição que um trecho N-terminal da sequência mostrou exercer sobre a expressão da proteína. Assim, este trabalho fornece exemplos de como é possível superar algumas das adversidades comuns da expressão heteróloga a fim de se obter CDKs humanas ativas empregando o sistema bacteriano.

ABSTRACT

CDK9 and CDK11 are members of the subfamily of transcriptional CDKs and therefore play central roles in the control of the dynamic activity of messenger RNA synthesis and processing by RNA polymerase II. Because of their ability to individually modulate the activity of RNAP II transcription complexes, these kinases are of great importance for the regulation of gene expression in eukaryotic cells. Several cases of deregulation of the transcriptional CDKs have been linked to important human diseases, including various types of cancer and also AIDS (due to the essential role of CDK9 in HIV replication). The objective of this work is to obtain human CDK9 and CDK11 heterologously expressed in *Escherichia coli*, seeking improved methods to produce these enzymes in quantity and purity suitable for structural studies. A bacterial system offers many practical advantages to protein production, such as technical simplicity, low costs, high levels of expression and high purification yield. However, it has always been an experimental challenge to obtain active human kinases from *E. coli*, because of the problems commonly associated with heterologous expression. The results presented here demonstrate the possibility of obtaining the enzymatically active human CDK9 by *in vitro* refolding of the protein expressed as bacterial inclusion bodies, after its solubilization and purification under denaturing conditions. Human CDK11, on the other hand, could only be obtained in a shortened form consisting just of its kinase domain, due to the strong inhibition that an N-terminal stretch exerts on the protein's own expression. Therefore, this work provides examples of how it is possible to overcome some of the common adversities of heterologous expression in order to obtain active human CDKs through the bacterial system.

LISTA DE FIGURAS

Figura 1: Estrutura do gene da CDK9 humana	18
Figura 2: Estrutura primária e secundária da CDK9	22
Figura 3: Estrutura cristalográfica da CDK9 complexada ao inibidor AMPPNP (PDB 3BLQ)	24
Figura 4: Ligações de H no reconhecimento da adenina	25
Figura 5: Esquema simplificado do mecanismo da reação de transfosforilação catalisada por Ser/Thr-quinases.....	27
Figura 6: Correlação entre elementos estruturais importantes e a sequência de aminoácidos da CDK9	29
Figura 7: Estrutura das ciclinas que se associam à CDK9	30
Figura 8: Comparação entre as estruturas do DRB e da adenosina.....	32
Figura 9: Função do P-TEFb na fosforilação do CTD, DSIF e NELF.....	35
Figura 10: Recrutamento do P-TEFb pela proteína Tat do HIV	39
Figura 11: Panorama da função celular e regulação do P-TEFb.....	43
Figura 12: Panorama da função celular da CDK11	56
Figura 13: Mapa do vetor bacteriano de expressão pET-28a(+) (Novagen)	71
Figura 14: Etapas para obtenção do pET28a::cdk11(Δ)	73
Figura 15: Sequência nucleotídica da CDK11 e primers utilizados	74
Figura 16: Sequência de aminoácidos da proteína truncada CDK11(Δ)	75
Figura 17: Teste de amplificação em diferentes temperaturas de hibridização	76
Figura 18: Separação do fragmento amplificado a partir do pET28a::cdk11	77
Figura 19: Teste de clivagem dos plasmídeos clonados a partir do produto de recombinação	78
Figura 20: Teste de clivagem comparativo dos plasmídeos contendo a sequência codificante da CDK11 antes e depois da modificação	80

Figura 21: Teste de amplificação com os plasmídeos pET28a::cdk11(Δ) utilizando os primers T7 do vetor	81
Figura 22: Resultado do sequenciamento de DNA evidenciando o sítio de recombinação na sequência codificante da CDK11(Δ)	82
Figura 23: Resultado preliminar para a expressão das CDKs 9 e 11 em <i>E. coli</i>	85
Figura 24: Expressão da proteína CDK11(Δ) em <i>E. coli</i>	87
Figura 25: Avaliação do conteúdo de códons raros na sequência codificante da CDK11	89
Figura 26: Teste de expressão das proteínas CDK11 e CDK11(Δ) em Rosetta (DE3)	91
Figura 27: Ensaio de clivagem comparativo com os plasmídeos pET23a::cdk9 e pET28a::cdk9.	94
Figura 28: Diferença de massa molecular ocasionada pela introdução da cauda de histidina na CDK9.....	96
Figura 29: Dependência do IPTG para expressão da proteína a partir do pET28a::cdk9	98
Figura 30: Comparação entre as regiões de clonagem e expressão dos vetores pET-28a(+) e pET-23 a(+) (Novagen)	99
Figura 31: Expressão da His-tag-CDK9 predominantemente na fração insolúvel do lisado bacteriano	100
Figura 32: Resultado do <i>Western Blot</i> para avaliar a expressão e solubilidade da His-tag-CDK9.....	104
Figura 33: Aplicação da fração solúvel do lisado na resina de níquel (<i>IMAC</i>) em condições não desnaturantes	106
Figura 34: Perfil das amostras eluídas na cromatografia de afinidade ao Ni^{2+} (<i>IMAC</i>) em condições desnaturantes	108
Figura 35: Resultado do teste de atividade com a resina de ATP.....	110
Figura 36: Ciclo catalítico da CDK9	113

LISTA DE TABELAS

Tabela 1: Principais eventos no desenvolvimento da pesquisa sobre a CDK9.	5
Tabela 2: Principais eventos no desenvolvimento da pesquisa sobre a CDK11	40
Tabela 3: Cepas bacterianas de clonagem e expressão utilizadas.....	65
Tabela 4: Sequências de DNA relacionadas à CDK11	83
Tabela 5: Sequências de DNA relacionadas à CDK9	95

SUMÁRIO

1 INTRODUÇÃO	12
1.1 Apresentação da CDK9	12
1.2 Genética da CDK9	18
1.3 Estrutura molecular da CDK9 e P-TEFb	21
1.4 Função molecular do P-TEFb	31
1.5 Apresentação e genética da CDK11	47
1.6 Função molecular da CDK11	52
2 JUSTIFICATIVA	57
3 OBJETIVOS	58
4 MATERIAIS E MÉTODOS	59
4.1 Modificação do plasmídeo pET28a::cdk11	59
4.1.1 Amplificação do pET28a::cdk11(Δ) por <i>PCR</i>	59
4.1.2 Eletroforese em gel de agarose	60
4.1.3 Ciclização do DNA plasmidial amplificado	61
4.1.4 Sequenciamento do DNA	63
4.2 Expressão das proteínas e purificação da CDK9	63
4.2.1 Transformação das cepas de expressão	63
4.2.2 Experimentos de indução	64

4.2.3	Lise das células	66
4.2.4	Eletroforese em gel de poliacrilamida (SDS-PAGE)	67
4.2.5	<i>Western Blot</i>	67
4.2.6	Purificação da CDK9 por cromatografia <i>IMAC</i>	68
4.2.7	Reenovelamento <i>in vitro</i> da CDK9	68
4.2.8	Ensaio de atividade enzimática	69
4.2.9	Quantificações de proteínas totais	70
5	RESULTADOS E DISCUSSÃO	71
5.1	Construção do plasmídeo pET28a::cdk11(Δ)	71
5.1.1	Estratégia para deleção interna na sequência da CDK11	72
5.1.2	Obtenção do pET28a::cdk11(Δ) a partir do pET28a::cdk11	76
5.1.3	Ensaio confirmatórios para o produto de recombinação	78
5.2	Obtenção das proteínas CDK9 e CDK11 humanas em <i>E. coli</i>	84
5.2.1	Expressão da CDK11(Δ) no sistema bacteriano	86
5.2.2	Expressão da CDK9 humana no sistema bacteriano	94
5.2.3	Purificação e reenovelamento da CDK9	105
6	CONCLUSÃO	115
	REFERÊNCIAS BIBLIOGRÁFICAS	117

1 INTRODUÇÃO

1.1 Apresentação da CDK9

A proteína que hoje é conhecida como CDK9 foi originalmente descoberta em 1994 por um grupo de pesquisadores da *Temple University*, na Filadélfia, em um estudo que pretendia identificar potenciais reguladores do ciclo celular a partir de uma biblioteca de cDNA humano (GRAÑA et al., 1994). Empregando *primers* degenerados, construídos com base na sequência determinada da CDC2 (hoje designada CDK1), foram isolados clones nos quais a sequência codificante de uma nova proteína relacionada à CDC2 foi reconhecida. A análise do quadro de leitura aberto (*ORF*) revelou que esta possuía 372 resíduos de aminoácidos e massa molecular estimada em aproximadamente 42 kDa, contendo as 11 regiões conservadas características do domínio catalítico das proteínas quinases. Ficou claro se tratar de uma proteína da mesma família da CDC2 e demais CDKs reconhecidas até aquele momento, especialmente pela observação de um motivo no seu domínio N-terminal (PITALRE) que corresponde a um trecho altamente conservado das CDKs. Este motivo é designado *PSTAIR box* em alusão à sequência homóloga original encontrada nas CDKs 1 e 2. Foi então sugerido que a nova proteína ficasse provisoriamente denominada PITALRE enquanto não fosse esclarecida a natureza de sua função (GRAÑA et al., 1994).

Análises do RNA de diversos tecidos humanos mostraram que a PITALRE era profusamente expressa em todos eles. Utilizando anticorpos desenvolvidos contra a PITALRE também foi possível localizar a proteína em escala subcelular, demonstrando sua ocorrência majoritariamente no núcleo, através de análises de *Western blot* com amostras de frações celulares. A PITALRE imunoprecipitada foi então testada em ensaios de atividade enzimática com substratos conhecidos da CDK2, sendo demonstrada sua atividade na fosforilação da caseína, proteína do retinoblastoma (pRb) e proteína básica da mielina (PBM). Além desta caracterização meticulosa alcançada logo no início pelos seus descobridores, foram ainda realizados experimentos de co-imunoprecipitação na tentativa de identificar proteínas que interagissem fisicamente com a PITALRE, embora não tenham avançado a ponto de reconhecer positivamente qualquer ciclina associada à quinase neste estudo (GRAÑA et al., 1994).

A designação “Quinases Dependentes de Ciclinas” (*Cyclin-Dependent Kinases*, ou CDKs) foi oficializada em 1991, três anos antes da descoberta da PITALRE, quando ocorreu o importante Simpósio Sobre o Ciclo Celular de *Cold Spring Harbor* (MALUMBRES et al., 2009). As CDKs são proteínas quinases cuja atividade enzimática depende essencialmente da sua interação com um segundo tipo de proteína, as ciclinas, formando um complexo heterodimérico no qual a CDK representa a subunidade catalítica e a ciclina a subunidade reguladora. Pela grande semelhança observada entre a PITALRE e as CDKs conhecidas, parecia bastante provável a sua interação com alguma proteína relacionada às ciclinas que servisse como subunidade ativadora da quinase. Entretanto, a princípio não fora encontrado nenhum indício de interação entre a PITALRE e as ciclinas conhecidas, havendo apenas a suspeita de um papel relacionado à regulação do ciclo celular pela sua identidade de sequência em relação às demais CDKs (GRAÑA et al., 1994). Como o consenso estabelecido até então era de que uma proteína quinase só poderia ser considerada um membro da família das CDKs depois que fosse demonstrada sua associação e ativação enzimática por uma ciclina, a PITALRE permaneceu sem ser oficialmente reconhecida como uma nova CDK.

Paralelamente à identificação da PITALRE, estudos sobre a transcrição em células de *Drosophila* haviam demonstrado que o papel do *fator positivo de alongação da transcrição b* (P-TEFb) consiste em fosforilar o domínio carboxi-terminal (CTD) de Rpb1, a maior subunidade da RNA Polimerase II (MARSHALL et al., 1996). Já se sabia que o P-TEFb consiste em um heterodímero cuja subunidade menor é responsável pela atividade de proteína quinase. Com o objetivo de aprofundar o conhecimento sobre o P-TEFb e sua função, um grupo de pesquisadores purificou a subunidade menor do P-TEFb (ZHU et al., 1997). A proteína purificada foi então sequenciada, e o resultado utilizado para construir primers que possibilitaram a clonagem do cDNA correspondente. O sequenciamento do cDNA permitiu deduzir a sequência completa de aminoácidos da proteína. Quando a sequência foi analisada, ficou claro que ela apresentava grande semelhança com os membros da família das CDKs. A comparação com um banco de dados revelou que a proteína humana PITALRE era a que apresentava maior identidade de sequência (72%), inclusive o motivo PITALRE encontrava-se conservado em *Drosophila* (ZHU et al., 1997). Portanto, era muito provável se tratar de proteínas ortólogas das duas espécies, o que indicava que a PITALRE deveria ser também encontrada no P-TEFb humano.

O reconhecimento da relação entre a PITALRE e o P-TEFb atraiu novamente o interesse para a proteína quinase, que agora se sabia estar envolvida na transcrição de genes pela RNAP II. No ano seguinte, 1998, a corrida em busca de uma ciclina associada à PITALRE terminou com a sua identificação por dois grupos de pesquisa independentes nos Estados Unidos. O grupo de David Price, em Iowa, havia sido responsável pela identificação da PITALRE como subunidade menor do P-TEFb. Continuando as pesquisas com *Drosophila*, foi possível purificar a subunidade maior do P-TEFb e seguir a mesma estratégia de utilizar o sequenciamento parcial para construir primers que possibilitaram a clonagem do cDNA completo. Com a sequência codificante isolada, foi deduzida a sequência de aminoácidos da proteína e obtida sua expressão em cultura de células (PENG et al., 1998a). Independentemente, na Califórnia, o grupo liderado por Katherine Jones empregou uma estratégia diferente para purificar a mesma proteína. Já havia sido demonstrado pelo grupo de Price que o P-TEFb se associa à proteína Tat do HIV, ativando a transcrição viral nas células infectadas (ZHU et al., 1997). O grupo de Jones utilizou a proteína fusionada GST-Tat para purificar por afinidade e identificar proteínas que interagissem com a Tat em um extrato nuclear de células HeLa. Foi encontrada uma proteína de 87 kDa que, após clonado seu cDNA, foi obtida por tradução *in vitro*. A comparação com um banco de dados mostrou que a proteína possuía, no seu N-terminal, um domínio característico da família das ciclinas – os *boxes* ciclina – e apresentava maior identidade com a sequência da ciclina C humana. Devido à sua propriedade exclusiva de interagir com a Tat (trans-ativadora da transcrição do HIV), foi sugerida a designação ciclina T para este novo membro da família das ciclinas (WEI et al., 1998).

Com a identificação da ciclina T, a PITALRE foi definitivamente reconhecida como uma CDK e renomeada CDK9, de acordo com a ordem de descobrimento na família. Ainda em 1998, o grupo de Price identificou mais dois tipos de ciclina que se associavam à CDK9 em células humanas, as quais foram designadas ciclinas T2a e T2b, enquanto a ciclina T original passou a ser referida como ciclina T1. As ciclinas T2a e T2b são produtos de *splicing* alternativo do mesmo gene, distinto daquele da ciclina T1 (ironicamente seria logo reconhecido que as ciclinas T2a e T2b, ao contrário da ciclina T1, não são capazes de interagir com a proteína Tat). Todas as três ciclinas T podem ativar a CDK9 quando ligadas e assim hiper-fosforilar o domínio CTD da RNAP II, promovendo a elongação transcricional

(PENG et al., 1998b). Em seguida, a descoberta de uma nova ciclina que interage com a CDK9 – a ciclina K – expandiu as possibilidades do cenário funcional a ser desvendado.

A ciclina K havia sido recentemente identificada em uma biblioteca de cDNA humano, através de um estudo de recuperação da progressão do ciclo celular no qual esta ciclina se mostrou capaz de resgatar células de *S. cerevisiae* da letalidade produzida pela deleção dos genes das ciclinas G₁ (EDWARDS et al., 1998). A escolha da letra “K” para designá-la foi sequencial, por ter sido antecedida pela descoberta de uma ciclina de *Drosophila* que fora denominada ciclina J. A interação entre a CDK9 e a ciclina K foi detectada através de um experimento de duplo-híbrido em levedura (FU et al., 1999). A ciclina K é menor que as ciclinas T e não possui um motivo rico em histidina, embora seu *box* apresente uma alta identidade de sequência em relação às ciclinas T. A ciclina K mostrou se associar à CDK9 formando um heterodímero estável e enzimaticamente ativo. O complexo CDK9/ciclina K foi capaz de fosforilar o CTD da RNAP II em ensaios *in vitro*, com cerca de metade da atividade em comparação com CDK9/ciclina T (FU et al., 1999).

À medida que a família das CDKs se expandia e o conhecimento de suas funções se aprofundava, parecia que nem todas as CDKs possuíam um papel relacionado diretamente ao ciclo celular. As CDKs 1, 2, 3, 4 e 6 apresentam uma variação periódica de atividade enzimática em sincronia com o ciclo celular, e participam de processos relacionados à progressão do ciclo e divisão celular em fases específicas. Outras CDKs possuem atividade independente do ciclo celular e participam de processos que não estão relacionados à divisão, possuindo em comum o fato de estarem envolvidas em diversos mecanismos de regulação da transcrição pela RNAP II. Assim, quanto à função biológica, a família das CDKs é dividida classicamente em dois grupos: As CDKs do ciclo celular e as reguladoras da transcrição. Esta separação não é muito rigorosa, pois alguns membros apresentam função tanto no ciclo celular quanto na regulação transcricional (como é o caso das CDKs 10 e 11). A CDK9 foi tradicionalmente considerada um exemplo inequívoco de CDK reguladora da transcrição, devido aos seus níveis constantes de expressão e atividade enzimática ao longo do ciclo celular (GRAÑA et al., 1994; GARRIGA et al., 2003). Entretanto, estudos recentes parecem ter encontrado uma função importante para o P-TEFb na progressão da fase G₁ do ciclo celular (YANG et al., 2008).

Tabela 1 – Principais eventos no desenvolvimento da pesquisa sobre a CDK9. A Tabela resume as principais contribuições para a evolução do conhecimento a respeito da CDK9, incluindo eventos anteriores à identificação da PITALRE (em 1994), que correspondem a descobertas iniciais relacionadas à família das CDKs.

Ano	Acontecimento
1970	*Leland Hartwell publica suas primeiras observações sobre os genes <i>cdc</i> em <i>S. cerevisiae</i> , reconhecendo mutantes com fenótipo condicional de interrupção em pontos específicos do ciclo celular (HARTWELL et al., 1974).
1971	*Yoshio Masui e Clement Markert demonstram a existência de um fator citoplasmático (MPF) responsável pela maturação de ovócitos de sapo (MASUI et al., 1971).
1983	*Tymothy Hunt descreve o padrão periódico de expressão das ciclinas ao longo do ciclo celular em ovos fertilizados de ouriços-do-mar, nos quais as primeiras divisões celulares ocorrem sincronizadamente (EVANS et al., 1983).
1987	*Paul Nurse identifica um gene humano homólogo ao <i>cdc2</i> de <i>S. pombe</i> , capaz de resgatar da letalidade mutantes com perda da função de <i>cdc2</i> . A proteína codificada por este gene é conhecida hoje como CDK1 (LEE et al., 1987).
1988/ 1989	*O isolamento do fator MPF mostrou que seu componente principal é um heterodímero, cuja subunidade menor é homóloga ao produto de <i>cdc2</i> em leveduras e possui atividade de proteína quinase. Em seguida, a subunidade maior foi identificada como uma ciclina (DUNPHY et al., 1988; GAUTIER et al., 1988; DRAETTA et al., 1989).
1991	*O Simpósio sobre o Ciclo Celular de Cold Spring Harbor oficializa o uso do termo “quinasas dependentes de ciclinas” e o reconhecimento das primeiras CDKs, de 1 a 6 (MALUMBRES et al., 2009).
1992	*Nicholas Marshall e David Price formulam um modelo para as fases iniciais da transcrição pela RNA Polimerase II que conjectura a existência de um fator positivo de elongação (P-TEF) com papel determinante na regulação transcrricional (MARSHALL et al., 1992).
1994	*Descoberta da CDK9 humana (originalmente PITALRE) durante uma varredura que procurava em uma biblioteca de cDNA novos reguladores do ciclo celular de mamíferos por homologia de sequência com as CDKs conhecidas (GRAÑA et al., 1994).
1996	*Revelada a atividade de proteína quinase do P-TEFb na fosforilação do CTD da RNA Polimerase II (MARSHALL et al., 1996).
1997	*A PITALRE foi reconhecida como a subunidade menor do P-TEFb pelo mesmo estudo que identificou sua interação com a Tat do HIV (ZHU et al., 1997).
1998	*A subunidade maior do P-TEFb é reconhecida como uma nova ciclina pelos grupos de David Price e Katherine Jones. A nova ciclina é designada ciclina T e logo são identificados os subtipos T1, T2a e T2b (PENG et al., 1998a; PENG et al., 1998b; WEI et al., 1998).
1999	*Demonstrada a associação e ativação enzimática da CDK9 pela ciclina K, de modo semelhante à interação com as ciclinas T (FU et al., 1999).

(Continua)

Tabela 1 – Principais eventos no desenvolvimento da pesquisa sobre a CDK9 (Continuação)

Ano	Acontecimento
2001	<p>*Hartwell, Nurse e Hunt recebem o prêmio Nobel de Fisiologia ou Medicina “por suas descobertas de reguladores-chave do ciclo celular”.</p> <p>*Descoberto que a subunidade Spt5 do fator negativo de alongação DSIF também constitui um alvo de fosforilação do P-TEFb (KIM et al., 2001).</p>
2003	<p>*Desvendada a interação do P-TEFb com o complexo inibidor formado pelo snRNA 7SK e HEXIM (MICHELS et al., 2003).</p> <p>*Identificada uma nova isoforma da CDK9, codificada pelo mesmo gene, que possui 55 kDa e também forma complexos com as ciclinas T e K (SHORE et al., 2003).</p>
2004	<p>*A subunidade RD do fator negativo de alongação NELF é reconhecida como mais um importante alvo de fosforilação do P-TEFb (FUJINAGA et al., 2004).</p>
2005	<p>*É descoberta a interação do P-TEFb com a proteína de bromodomínio Brd4 e seu efeito positivo sobre a transcrição (JANG et al., 2005; YANG et al., 2005).</p> <p>*Observada a atividade das proteínas fosfatases PP1 e PP2A sobre a CDK9 auto-fosforilada em diferentes contextos (AMMOSOVA et al., 2005).</p>
2007	<p>*Demonstrada a regulação da atividade do P-TEFb por acetilação da CDK9 (FU et al., 2007).</p>
2008	<p>*Elucidação da estrutura cristalográfica da CDK9 nas suas formas livre (inativa) e em associação com a ciclina T1 (ativa), complexada com inibidores (BAUMLI et al., 2008).</p> <p>*Revelada a participação do P-TEFb na transição G₁/S do ciclo celular através da sua interação com Brd4 (YANG et al., 2008).</p>
2009	<p>*A fosforilação da CDK9 em Thr29 é observada como resultado da associação do P-TEFb com Brd4 (ZHOU et al., 2009).</p>
2010	<p>*Os efeitos da perda de função da CDK9 são observados em células humanas com a superexpressão de uma forma dominante negativa da CDK, revelando um resultado diferente do esperado sobre a expressão de RNAm (GARRIGA et al., 2010).</p> <p>*Elucidada a estrutura cristalográfica do P-TEFb em complexo com a proteína transativadora (Tat) do HIV (TAHIROV et al., 2010).</p>
2012	<p>*Desvendado o papel da cauda C-terminal da CDK9 no mecanismo enzimático e sua implicação na ordem de ligação e dissociação dos substratos (BAUMLI et al., 2012).</p> <p>*Descoberta a fosforilação da CDK9 no resíduo Ser90 pela CDK2 (BREUER et al., 2012).</p> <p>*Evidências experimentais diretas revelam que a CDK9 fosforila a posição Ser5 do CTD, e não Ser2 como se acreditou durante muito tempo (CZUDNOCHOWSKI et al., 2012).</p>

1.2 Genética da CDK9

O gene que codifica a CDK9 humana está localizado em 9q34.1, próximo à extremidade telomérica no braço longo do cromossomo 9. Existem duas formas alternativas da proteína CDK9 – designadas CDK9₄₂ e CDK9₅₅ – que correspondem a isoformas codificadas pelo mesmo gene, com massas moleculares aproximadas de 42 e 55 kDa. As duas isoformas existem devido à ocorrência de dois promotores distintos no gene da CDK9, os quais encontram-se separados por mais de 500 pb (SHORE et al., 2005), como mostra a Figura 1.

Figura 1 – Estrutura do gene da CDK9 humana. O esquema ilustra a organização do gene da CDK9 em sequências intrônicas (não codificantes) e exônicas (codificantes). Os blocos representam os éxons, destacando a sequência codificante da CDK9₄₂ em verde, e o segmento adicional de 351 pb (exclusivo da CDK9₅₅) em azul claro. As setas roxas indicam os pontos alternativos de início da transcrição do RNA da CDK9 nos promotores. As setas em laranja indicam as posições dos códons de iniciação (ATG) para cada uma das isoformas da CDK9, a partir dos quais os ribossomos iniciam a tradução da sequência em proteína. São indicados os tamanhos (em pb) de éxons e íntrons, mas os blocos não estão em escala.

Fonte: Adaptado de Liu et al. (2000, p. 54) e Shore et al. (2005, p. 54).

A CDK9₅₅ é transcrita a partir de um promotor contendo elemento TATA, e possui 351 pb adicionais na sua sequência codificante em relação à CDK9₄₂. A CDK9₄₂, por sua vez, é transcrita a partir de um promotor mais adiante, rico em pares GC e sem um elemento TATA funcional. O segmento de 351 pb adicionado à extremidade 5' da sequência codificante na

CDK9₅₅ é contínuo com o códon de iniciação da CDK9₄₂, e não possui nenhuma sequência intrônica. Assim, ambas as sequências codificantes da CDK9 são intercaladas pelos mesmos seis íntrons (LIU et al., 2000; SHORE et al., 2005). O promotor da CDK9₄₂ já foi bem caracterizado, de maneira que sabemos que a transcrição é iniciada em torno de 79 pb antes do seu códon de iniciação. De acordo com o diagrama da Figura 1, é evidente que a sequência promotora da CDK9₄₂ encontra-se no interior (na extremidade 5') da sequência codificante da CDK9₅₅. Antes da identificação da isoforma CDK9₅₅ (SHORE et al., 2003), esta sequência a 5' do códon de iniciação da CDK9₄₂ era reconhecida apenas como sequência regulatória (por conter o promotor da CDK9₄₂), mas não como sequência codificante. O promotor da CDK9₄₂ é pequeno e possui elevada atividade constitutiva, indicando que seja consideravelmente mais ativo que o promotor da CDK9₅₅ (SHORE et al., 2005). Variações substanciais podem ser observadas na abundância relativa das duas isoformas entre diversos tecidos humanos e de camundongo. Por exemplo, a CDK9₄₂ é predominantemente expressa nos testículos e no baço, enquanto a CDK9₅₅ predomina no cérebro e no fígado de murinos. Estas variações na quantidade relativa das isoformas ocorrem também entre diferentes linhagens de células cultivadas (SHORE et al., 2005).

As observações sobre a variabilidade da proporção celular entre as isoformas da CDK9 indicam que ambas têm a sua expressão diferencialmente regulada de modo tecido-específico (ROMANO et al., 2008). Uma consequência importante desta regulação independente é a variação significativa que pode ocorrer na proporção celular das duas isoformas da CDK9 em resposta a um estímulo externo. Por exemplo, foi observado em macrófagos humanos cultivados que a CDK9₅₅ é a isoforma predominante enquanto as células estão inativas. Entretanto, após o tratamento com a endotoxina LPS (lipopolissacarídeo), ocorre uma mudança no padrão de expressão gênica e a CDK9₄₂ passa a predominar (SHORE et al., 2005). O mesmo foi observado em uma cultura de hepatócitos de rato. *In vivo*, os hepatócitos são células quiescentes típicas e expressam a maior proporção relativa de CDK9₅₅ (68% do total). Porém, após remoção do tecido e incubação em condições apropriadas, eles passam a se dividir e podem originar uma cultura primária. Quando uma cultura de hepatócitos é estabelecida, a quantidade relativa de CDK9₄₂ aumenta com o passar do tempo até que ela passa a ser a forma predominante após 60h de cultura, representando cerca de 70% da CDK9 celular total (SHORE et al., 2005). De fato, uma

indução na expressão da CDK9₄₂ durante a ativação de linfócitos-T humanos já havia sido demonstrada antes do reconhecimento da isoforma CDK9₅₅ (HERMANN et al., 1998).

A família das CDKs no genoma humano compreende 21 genes identificados até agora que codificam CDKs verdadeiras, e mais 5 genes que codificam proteínas relacionadas às CDKs, as quinases CDKLs (*CDK-like*) (MALUMBRES et al., 2009). Esta família, por sua vez, se insere como um pequeno subgrupo na grande família das proteínas quinases, uma das maiores famílias de proteínas dos eucariotos superiores. No genoma humano existem mais de 500 genes que codificam proteínas quinases, o que corresponde a cerca de 1,7% de todos os genes humanos (MANNING et al., 2002). O conjunto de todas as proteínas quinases de um organismo é designado quinoma. O desenvolvimento notável dos quinomas em eucariotos superiores reflete a ampla variedade de funções celulares realizadas pelas proteínas que são alvos das quinases: mais de um terço das proteínas em uma célula de mamífero típica são fosforiladas em algum contexto para regular a sua função (ALBERTS et al., 2010).

A CDK9 faz parte do grupo de CDKs reguladoras transcricionais e parece ser filogeneticamente mais próxima das CDKs 12 e 13, com as quais apresenta maior identidade de sequência (entre 44% e 48% no *Homo sapiens*). Desde o descobrimento da PITALRE, a CDK13 - então denominada CHED quinase - foi reconhecida como a proteína mais proximamente relacionada à CDK9 (GRAÑA et al., 1994). Entretanto, recentemente foi esclarecido que o único ortólogo autêntico da CDK9 em *S. cerevisiae* é Bur1, enquanto as CDKs 12 e 13 são os ortólogos da Ctk1 de levedura (BARTKOWIAK et al., 2010). Em animais, os ortólogos da CDK9 são designados pelo mesmo nome. Estas quinases já foram estudadas em uma variedade de organismos modelo e sua função parece ser amplamente conservada entre os diversos grupos. Embora as CDKs sejam mais conhecidas e pesquisadas por suas funções nas células animais, elas também estão presentes e realizam funções importantes no Reino Vegetal. Enquanto as classes de CDKs animais são designadas por números, as CDKs vegetais são designadas por uma combinação de números e letras. Assim, as CDKCs são os ortólogos vegetais da CDK9 animal. Em *Arabidopsis*, CDKC1 e CDKC2 possuem um motivo PITAIRE e parecem interagir com homólogos das ciclinas T e K de animais. (BARRÔCO et al, 2003).

1.3 Estrutura molecular da CDK9 e P-TEFb

Foi mencionado no item anterior que as CDKs pertencem à grande família das proteínas quinases (PKs) eucarióticas. Estas enzimas promovem a transferência de um grupo fosforila ($-\text{PO}_3^{2-}$) do fosfato γ do ATP - ou mais raramente GTP - para hidroxilas nas cadeias laterais dos resíduos de Ser, Thr ou Tyr de seus substratos peptídicos. Os produtos da fosforilação destes resíduos são monoésteres fosfato de álcool primário (fosfoserina), secundário (fosfotreonina) ou fenol (fosfotirosina). De acordo com a especificidade dos resíduos que fosforilam nos seus substratos, as proteínas quinases eucarióticas são divididas em dois grupos principais: as serina/treonina-quinases, que fosforilam hidroxilas alcoólicas de Ser e Thr, e as tirosina-quinases, que fosforilam hidroxilas fenólicas de Tyr. No sistema de classificação de enzimas da União Internacional de Bioquímica e Biologia Molecular (IUBMB), as CDKs recebem a numeração EC 2.7.11.22, correspondente à Classe 2 (transferases), Subclasse 7 (fosfotransferases), Sub-subclasse 11 (serina/treonina-quinases) e Número Serial 22 (CDKs) (ALBERTS et al., 2010; NELSON et al., 2011).

As proteínas quinases eucarióticas compartilham uma sequência conservada de 250 a 300 resíduos de aminoácidos, que se enovela em uma estrutura terciária característica da superfamília. Este domínio conservado é composto por 12 subdomínios próprios que, juntos, reúnem os elementos estruturais responsáveis pela atividade enzimática das quinases. Em cada subdomínio os motivos de estrutura secundária (hélices α e fitas β) são particularmente conservados, enquanto as variações principais ocorrem nos segmentos de alças que conectam estes motivos. Estas variações produzem a diversidade estrutural das numerosas PKs em uma célula eucariótica e lhes conferem sua especificidade funcional, como a afinidade enzimática por determinados substratos peptídicos (HANKS et al., 1995; ALBERTS et al., 2010). A determinação da estrutura cristalográfica da CDK9 foi alcançada somente em 2008, 14 anos após a sua descoberta (BAUMLI et al., 2008). Na Figura 2 está representada a sequência de aminoácidos da CDK9₄₂ humana, incluindo seus elementos de estrutura secundária. Alguns pequenos trechos da sequência que possuem estrutura secundária não estão nomeados porque não correspondem a motivos comumente conservados entre outras PKs.

Figura 2 – Estrutura primária e secundária da CDK9. O diagrama apresenta a sequência de 372 resíduos da isoforma CDK9₄₂ humana, delimitando os motivos de estrutura secundária. Devido à preponderância de resíduos básicos no polipeptídeo, o pI da CDK9 possui um valor atipicamente elevado, ao redor de 9,0. A sequência PITALRE está destacada em fundo amarelo. Os asteriscos marcam os resíduos que fazem contato com a ciclina T1 e o inibidor AMPPNP no complexo ternário (PDB 3BLQ).

Fonte: Adaptado de Baumli et al. (2008, p. 1910).

O domínio quinase se enovela em uma estrutura terciária bilobada, formada por um lobo N-terminal menor com predominância de fitas β , e um lobo C-terminal maior com predominância de hélices α (BOSSEMEYER, 1995; HANKS et al., 1995). O lobo N-terminal da

CDK9 é compreende os resíduos 16 a 108, contendo cinco trechos de fitas β ($\beta 1$ - $\beta 5$) e uma hélice α (αC). O lobo C-terminal inclui os resíduos de 109 a 330, que engloba quatro pequenas fitas β ($\beta 6$ - $\beta 9$) e sete hélices α principais (αD - αJ) (BAUMLI et al., 2008). Observando a Figura 2, fica evidente que a interação entre a CDK e a ciclina ocorre através do lobo N-terminal da CDK, envolvendo principalmente resíduos de aminoácidos na região da hélice αC e suas proximidades.

Na Figura 2 podem ser identificados os resíduos que interagem diretamente com o inibidor AMPPNP – também chamado de ADPNP ou 5'-(β,γ -imido)trifosfato de adenosina – um análogo estrutural do ATP caracterizado por uma substituição do átomo de oxigênio em ponte entre os fosfatos β e γ do ATP por um grupo -NH-. Devido à grande semelhança estrutural, é esperado que a interação da CDK9 com o ATP envolva precisamente os mesmos resíduos e o mesmo modo de ligação observado para o AMPPNP. Como o sítio de ligação ao ATP encontra-se na fenda entre os dois lobos da quinase, há interações com resíduos de ambos os lobos voltados para o interior da fenda, e também alguns resíduos da alça que liga os dois lobos (entre $\beta 5$ e αC), designada “dobradiça”. No sítio de ligação ao ATP em uma PK, a adenina ocupa a região mais profunda da fenda, de forma que o grupo trifosfato fica orientado para o exterior. O fosfato γ (terminal) da cadeia trifosfato situa-se próximo à superfície da enzima, bem ao lado do sítio de ligação ao substrato proteico acceptor de fosforila (BOSSEMEYER, 1995; HANKS et al., 1995).

O modo de ligação do ATP à CDK9 está mostrado na Figura 3. Todas as proteínas quinases utilizam o cátion Mg^{2+} como co-fator para a ligação ao ATP (mais raramente podem utilizar Mn^{2+}). Como o Mg^{2+} já se encontra associado ao ATP em solução, podemos considerar que o complexo $MgATP^{2-}$ é o verdadeiro substrato doador de fosforila na reação catalisada pelas PKs (NELSON et al., 2011). Na estrutura do complexo, o Mg^{2+} apresenta-se coordenado pelos fosfatos β e γ do ATP, pelo grupo carboxilato de um resíduo conservado de aspartato (Asp167 na CDK9) e duas moléculas de água. Este aspartato pertence ao motivo DFG, situado na alça entre $\beta 8$ e $\beta 9$, que é universalmente conservado entre as PKs, mas não está destacado na Figura 3(b) porque se encontra oculto atrás do cátion Mg^{2+} nesta perspectiva (BOSSEMEYER, 1995; HANKS et al., 1995).

Figura 3 – Estrutura cristalográfica da CDK9 complexada ao inibidor AMPPNP (PDB 3BLQ – resolução 2,90 Å). A estrutura da CDK9 em conformação ativa é derivada de um complexo CDK9/ciclina T1, do qual a ciclina foi omitida por clareza. Para facilitar a cristalização, foi utilizada uma forma encurtada da CDK9 (resíduos 1-330). Em **(a)** estão evidenciados os motivos de estrutura secundária com as mesmas cores da Figura 2. Em **(b)** são destacados os principais subdomínios funcionais de uma PK na estrutura da CDK9.

Fonte: Autoria própria, a partir da estrutura depositada no *RCBS Protein Data Bank* (PDB 3BLQ), utilizando o programa PyMOL versão 1.1 (DeLano Software).

No lobo N-terminal encontram-se a alça G (rica em resíduos de glicina) e a hélice α_C . Já discutimos o papel da hélice α_C como a principal região de interação das CDKs com as ciclinas que as ativam. Quando o ATP se liga ao sítio ativo da enzima, a alça G se posiciona cobrindo a cadeia de trifosfato e seus resíduos formam ligações de hidrogênio específicas com os fosfatos não transferíveis do ATP, utilizando os grupos amida na cadeia principal da enzima (HANKS et al., 1995). Devido à sua propriedade de interagir com a cadeia trifosfato, a alça G é também denominada alça de ligação ao fosfato. O sítio ativo se estende desde a alça de ligação ao fosfato, no lobo N-terminal, até a alça catalítica, no domínio C-terminal. Esta região circunda o sítio de ligação do ATP, cuja estrutura é especialmente conservada por todo o grupo das PKs. (BOSSEMEYER, 1995; HANKS et al., 1995). Conectando os lobos N-

e C-terminal de uma PK existe uma alça designada dobradiça, que está indicada por uma seta na Figura 3(a) e destacada em azul claro na Figura 3(b). A dobradiça é uma região importante de interação entre a quinase e o ATP, por formar ligações de hidrogênio específicas com o anel purínico da adenina e grupos hidroxila da ribose (HANKS et al., 1995). De especial interesse são as ligações de hidrogênio formadas entre a dobradiça e a adenina, devido ao padrão universalmente conservado pelas proteínas quinases e sua importância para a interação com inibidores enzimáticos que se ligam ao sítio do ATP (FISCHER, 2004). O padrão formado pelas ligações de hidrogênio entre a adenina do ATP e o segmento da dobradiça está ilustrado na Figura 4.

Figura 4 – Ligações de H no reconhecimento da adenina. A Figura 4(a) representa o modo de interação do núcleo purínico do ATP com grupos amida na cadeia principal do segmento da dobradiça. Uma ligação de hidrogênio (I) é estabelecida entre o grupo NH_2 exocíclico da adenina (doador) e a carbonila de um resíduo de ordem i . Outra ligação de hidrogênio (II) envolve o nitrogênio N1 da adenina (aceptor) e o grupo NH de um resíduo de ordem $i+2$. Embora a eletronegatividade do carbono C2 da adenina não seja grande o suficiente para formar uma ligação de hidrogênio verdadeira (III) com a carbonila do resíduo de ordem $i+2$, a ligação possui a orientação correta para atuar como doadora de H. Este potencial para formação de ligações de H na posição (III) é efetivamente aproveitado por muitos inibidores de quinases, que interagem com a dobradiça utilizando dois ou três dos pontos de ligação mostrados. (A representação das ligações de H aqui não indica sua geometria e orientação real no espaço). Curiosamente, este modo de ligação da adenina apresenta o mesmo padrão doador-aceptor das ligações de H envolvidas no pareamento de bases A = T no DNA, como mostrado na Figura 4(b).

Fonte: Autoria própria.

Considerando agora o lobo C-terminal, os subdomínios funcionais indicados na Figura 3(b) são a alça T e a alça catalítica. A alça T - ou segmento de ativação - é assim chamada devido à ocorrência de um resíduo de treonina cuja fosforilação é essencial para a atividade enzimática. No caso da CDK9 a fosforilação do segmento de ativação ocorre em Thr186, que corresponde à Thr160 da CDK2. Entretanto, a CDK2 é ativada pela quinase ativadora de CDKs (CAK), enquanto a CDK9 se auto-fosforila nesta posição (BAUMLI et al., 2008). Para a CDK2 o processo de ativação já foi elucidado, e envolve a mudança de conformação da alça T de modo a desobstruir o sítio ativo ao mesmo tempo em que produz a organização local dos resíduos para o reconhecimento específico do substrato peptídico. A alça catalítica, por sua vez, contém os resíduos diretamente envolvidos na reação de transfosforilação, que na CDK9 correspondem a Asp149, Lys151 e Thr191. A lisina e a treonina desempenham um papel secundário, enquanto o aspartato é a chave do mecanismo de catálise. No mecanismo utilizado pelas proteínas quinases, a transferência do grupo fosforila ocorre diretamente do fosfato γ do ATP para a hidroxila do resíduo-alvo no substrato peptídico. Desta maneira, a reação ocorre sem produzir intermediários contendo resíduos fosforilados da própria enzima, como fazem as proteínas histidina-quinases (HANKS et al., 1995) e algumas proteínas fosfatases (ADAMS, 2001).

A tríade catalítica Asp/Lys/Thr forma uma rede de ligações de hidrogênio entre si e com a hidroxila do resíduo a ser fosforilado no substrato peptídico, posicionando-a para o ataque nucleofílico sobre o fosfato γ do ATP. O carboxilato do resíduo de Asp catalítico atua como uma base sobre a hidroxila do resíduo-alvo, ativando-a como nucleófilo. O alcóxido formado então ataca o grupo fosfato γ adjacente no sítio ativo, deslocando o ADP e gerando o monoéster de fosfato do resíduo-alvo (SMITH et al., 2011). Muitas linhas de evidências indicam que as reações de transfosforilação catalisadas por proteínas quinases seguem um mecanismo concertado de substituição nucleofílica, como a inversão universal na configuração estereoquímica ao redor do átomo de fósforo (ADAMS, 2001). Este mecanismo concertado é equivalente à substituição nucleofílica de ordem 2 (S_N2) no carbono, que também ocorre através de um mecanismo sem intermediários e com um único estado de transição separando reagentes e produtos. A Figura 5 ilustra simplificada o mecanismo da reação catalisada. A conformação do segmento de ativação que determina a especificidade para o motivo Ser/Thr-Pro do substrato é muito similar nas estruturas dos

complexos CDK2/ciclina A e CDK9/ciclina T1 (BAUMLI et al., 2008). Os motivos Ser/Thr-Pro são observados nos principais substratos do P-TEFb, como o domínio C-terminal (CTD) da subunidade maior da RNAP II. Este domínio contém dezenas de repetições em série do heptâmero Tyr-Ser-Pro-Thr-Ser-Pro-Ser (YSPTSPS), contabilizando um total de 52 repetições (21 exatas) para a enzima humana (NELSON et al., 2011).

Figura 5 – Esquema simplificado do mecanismo da reação de transfosforilação catalisada por Ser/Thr-quinases. O resíduo de aspartato da alça catalítica – Asp149 na CDK9 – atua como base sobre a hidroxila do resíduo-alvo no substrato peptídico. A retirada do próton produz um ânion alcóxido que ataca nucleofilicamente o fosfato γ do ATP, formando uma nova ligação com o fósforo e deslocando o ADP como grupo abandonador. A representação do mecanismo é simplificada porque omite a participação de outros resíduos do sítio ativo que atuam como coadjuvantes, no alinhamento dos grupos reativos, polarização das ligações dos reagentes e estabilização das cargas formadas nos produtos. O único resíduo coadjuvante mostrado é o aspartato do motivo DFG na alça de ligação ao metal, que coordena o cátion Mg^{2+} .

Fonte: Adaptado de Adams (2001, p. 2275) e Smith et al. (2011, p. 14).

A princípio, foi atribuído à CDK9 o papel de fosforilar os motivos repetidos do CTD na posição Ser2, mas estudos recentes contradizem esta função de Ser2 CTD quinase proposta originalmente. Foi primeiro relatado que a perda total da atividade da CDK9 *in vivo* resulta na desfosforilação de ambas Ser2 e Ser5 (GARRIGA et al., 2010). Em seguida, foi descoberto que as CDKs 12 e 13 possuem atividade de Ser2 CTD quinases e devem ser as principais responsáveis pela fosforilação do CTD nesta posição em células de metazoários (BARTKOWIAK et al., 2010). Por fim, a especificidade da CDK9 na fosforilação do CTD foi testada diretamente sobre o substrato em ensaios *in vitro*. Os resultados de uma série de experimentos independentes convergem para a conclusão de que a CDK9 é capaz de

promover apenas a fosforilação de Ser5, mas não de Ser2 como se pensava anteriormente (CZUDNOCHOWSKI et al., 2012).

Além de atuar regulando a função dos seus substratos por meio da fosforilação, a própria CDK9 é regulada por múltiplas fosforilações e outras formas de modificação covalente. Conforme mencionado, a fosforilação no resíduo Thr186 do segmento de ativação é a modificação mais importante que a CDK9 sofre, por ser essencial à sua atividade enzimática. Além deste, a CDK9 ainda possui cerca de uma dezena de sítios de fosforilação e 2 sítios de acetilação. Os demais sítios de fosforilação da CDK9 só foram descobertos mais recentemente e correspondem a Thr29, Ser90 e Ser175. A Thr29 também é um sítio de auto-fosforilação, embora esta dependa da associação do P-TEFb com Brd4, uma das principais proteínas com que interage nas células (ZHOU et al., 2009). Já a fosforilação em Ser90 é promovida pela CDK2 *in vivo* e corresponde ao único sítio da CDK9 que é reconhecidamente fosforilado por outra quinase (BREUER et al., 2012). A quinase responsável pela fosforilação em Ser175 permanece indeterminada, embora seja provável que se trate de mais um sítio de auto-fosforilação, pois o P-TEFb é capaz de fosforilar na posição equivalente peptídeos derivados da alça T da CDK9 (AMMOSOVA et al., 2011). Os resíduos que sofrem estas modificações estão indicados na Figura 6, assim como os domínios funcionais da CDK9 que foram discutidos no texto.

Uma vez que a forma enzimaticamente ativa da CDK9 ocorre obrigatoriamente associada a uma das ciclinas com as quais interage, a estrutura destas ciclinas também é importante para a função biológica do P-TEFb. A principal característica estrutural das ciclinas é a presença de um domínio particular - formado pelos *boxes* ciclina - que apresenta um modo de enovelamento classicamente associado a esta família. No caso da ciclina T1, observamos a ocorrência de dois motivos que não são encontrados em nenhuma das demais ciclinas ativadoras da CDK9: um motivo TRM de interação com a proteína Tat do HIV, parcialmente sobreposto ao final do segundo *box* ciclina, e uma sequência *PEST* de reconhecimento por ubiquitina-ligases, localizada na extremidade C-terminal da ciclina. As ciclinas T2a e T2b são isoformas produzidas por *splicing* alternativo a partir de um mesmo gene. A Figura 7(a) ilustra a organização dos domínios funcionais nestas ciclinas.

Figura 6 – Correlação entre elementos estruturais importantes e a sequência de aminoácidos da CDK9. A figura apresenta o alinhamento das sequências da CDK9 e CDK2 humanas, as quais compartilham 40% de identidade. Os resíduos conservados são indicados pelo sinal “#”. Na sequência da CDK9 estão destacados os domínios funcionais e sítios de fosforilação e acetilação. Podemos observar que os resíduos conservados se concentram justamente nas regiões dos domínios funcionais, os quais compõem o sítio ativo da quinase. As cores utilizadas para representar cada domínio funcional seguem a mesma correspondência da Figura 3.

Fonte: Autoria própria.

Figura 7 – Estrutura das ciclinas que se associam à CDK9. (a): Representação em bloco dos principais domínios funcionais das ciclinas que se ligam à CDK9 para ativá-la. (b) e (c): As estruturas cristalográficas dos boxes das ciclinas T1 (PDB 3BLQ – resolução 2,90 Å) e K (PDB 2I53 – resolução 1,50 Å) estão alinhadas para evidenciar a notável semelhança estrutural. (d): Diagrama topológico dos *boxes* ciclina de T1, ilustrando a estrutura típica deste domínio. A coloração das hélices segue o mesmo padrão utilizado nas estruturas (b) e (c), permitindo identificá-las.

(a) Organização dos domínios funcionais nas ciclinas ativadoras da CDK9

(b) Ciclina K (boxes)

(c) Ciclina T1 (boxes)

(d) Diagrama topológico dos boxes ciclina de T1

Fonte: (a) Adaptado de Wei et al. (1998, p.453) e Lin et al. (2002, p. 16874). (b) e (c) Autoria própria, a partir das estruturas depositadas no *RCBS Protein Data Bank* (2I53 e 3BLQ). (d) Adaptado de Baek et al. (2007, p. 12).

1.4 Função molecular do P-TEFb

A transcrição é o processo que apresenta as maiores diferenças na Biologia Molecular entre procariotos e eucariotos. Enquanto os procariotos possuem uma única RNA polimerase, os eucariotos possuem três - designadas RNAP I, RNAP II e RNAP III. Cada RNA polimerase eucariótica é responsável pela transcrição de certos tipos de genes, mas a RNAP II se destaca por transcrever todos os genes que codificam proteínas, sintetizando assim a totalidade do RNAm celular. Embora as três RNA polimerases eucarióticas sejam muito semelhantes em sua estrutura e composição de subunidades, a subunidade maior da RNAP II apresenta um domínio C-terminal (CTD) exclusivo, o qual é dispensável para a atividade enzimática da RNAP II, mas essencial à sua função *in vivo*. Notavelmente, o CTD contém dezenas de repetições do heptâmero YSPTSPS, cuja fosforilação nos resíduos de serina assume um papel muito importante na regulação da transcrição. Completamente estendido, ele seria cerca de dez vezes mais longo que o restante da polimerase. Hoje sabemos que o CTD funciona como uma plataforma de ligação para proteínas que participam do processo de transcrição pela RNAP II, especialmente os fatores que realizam o processamento co-transcricional do RNAm eucariótico (ALBERTS et al., 2010; NELSON et al., 2011).

A transcrição - assim como a síntese de biomoléculas poliméricas em geral - pode ser dividida em três processos básicos: (1°) Iniciação, (2°) Elongação e (3°) Terminação. Mais que uma divisão conceitual, os processos diferem essencialmente na sua natureza, demandando que a polimerase e fatores associados atuem de modo distinto em cada fase. No caso da RNAP II a iniciação é particularmente complexa, por ser subdividida em uma série de etapas intermediárias que regulam a passagem para o modo de elongação. A descoberta do P-TEFb ocorreu no contexto das pesquisas que elucidaram este mecanismo de iniciação da RNAP II. Mesmo antes que a CDK9 fosse identificada como a subunidade catalítica do P-TEFb, sua atividade de proteína quinase já havia sido reconhecida e associada à fosforilação do CTD durante a iniciação transcricional. Embora os detalhes do mecanismo ainda representem um tema corrente de investigação e debate, o modelo proposto por Nicholas Marshall e David Price há quase 20 anos - o qual atribui ao P-TEFb o papel central na passagem para a elongação - permanece como um paradigma na regulação da RNAP II (MARSHALL et al., 1992; MARSHALL et al., 1995; PETERLIN et al., 2006; PRICE, 2008).

Muito do que sabemos hoje sobre a iniciação da RNAP II é devido aos experimentos que investigaram o bloqueio da transcrição em células eucarióticas pelo DRB (5,6-dicloro-1- β -D-ribofuranosilbenzimidazol), um análogo nucleosídico halogenado cuja estrutura é apresentada na Figura 8. Há muito tempo foi percebido que a exposição de células de mamíferos ao DRB é letal, por provocar uma redução dramática dos níveis de RNAm (SEHGAL et al., 1976). Os mesmos efeitos da exposição celular ao DRB foram observados em ensaios de transcrição *in vitro* utilizando extratos nucleares brutos, o que facilitou muito o trabalho subsequente dos pesquisadores (PRICE, 2000). Talvez a maior surpresa revelada pelos primeiros estudos sobre o DRB seja a sua incapacidade de inibir diretamente a RNAP II purificada, como seria de se esperar para um análogo nucleosídico que apresenta a capacidade de suprimir a transcrição.

Figura 8 – Comparação entre as estruturas do DRB e da adenosina. O DRB (5,6-dicloro-1- β -D-ribofuranosilbenzimidazol) é um análogo nucleosídico que contém o grupo 5,6-diclorobenzimidazol em lugar da base nitrogenada. É reconhecido por sua capacidade de suprimir a síntese de RNAm eucariótico, bloqueando a passagem da RNAP II à fase de alongação através da inibição da CDK9. A estrutura cristalográfica do heterodímero CDK9/ciclina T em complexo com o DRB está disponível em uma resolução de 2,8 Å (PDB 3MY1).

Fonte: Autoria própria.

Utilizando extratos nucleares de *Drosophila* para estabelecer um sistema de transcrição *in vitro*, Marshall e Price realizaram vários experimentos com o DRB para elucidar seu mecanismo de bloqueio da síntese de RNAm. Assim, foi determinado que a RNAP II forma dois tipos distintos de complexos de alongação após iniciar a transcrição em um promotor. O tipo mais abundante é incapaz de produzir longos transcritos, interrompendo

sua síntese depois de adicionar algumas dezenas de nucleotídeos. O segundo tipo de complexo de alongação não apresenta esta limitação, e prossegue estendendo a cadeia até transcrever o molde de DNA completo. Estes dois processos foram denominados alongação abortiva e produtiva, respectivamente. A adição do DRB ao sistema mostrou impedir a formação de complexos do segundo tipo, resultando em um modo exclusivamente abortivo de alongação (MARSHALL et al., 1992).

Estas observações da influência do DRB sobre os complexos de alongação levaram à proposição de um modelo consistente para explicar o comportamento da RNAP II nas fases iniciais da transcrição. Ele estabelece que a RNAP II tende naturalmente a assumir um modo abortivo de alongação logo após iniciar a síntese do transcrito, assim a passagem à alongação produtiva depende de uma etapa posterior que é impedida pela adição de DRB ao meio. Para explicar a transição entre os dois modos de alongação da RNAP II, os autores teorizaram a existência de dois tipos de fatores envolvidos no processo: Fatores positivos (genericamente designados P-TEF) deveriam favorecer a passagem à alongação produtiva, enquanto fatores negativos (genericamente designados N-TEF) dificultariam esta passagem. Logo após a iniciação, a RNAP II passa a sofrer a influência dos fatores negativos de alongação e não pode sintetizar uma cadeia de RNA que ultrapasse algumas dezenas de nucleotídeos de comprimento. Para evitar um ciclo fútil de alongação abortiva *in vivo*, a polimerase deve realizar uma pausa pouco depois de abandonar o promotor até que a atuação dos fatores positivos supere o bloqueio, promovendo sua passagem à alongação produtiva. O efeito do DRB consiste justamente na inibição destes fatores positivos (MARSHALL et al., 1992).

A continuação das pesquisas permitiu identificar, nos extratos nucleares de *Drosophila*, alguns fatores que exerciam a função atribuída ao P-TEF. Dentre eles, o único que se provou essencial à alongação produtiva foi um heterodímero de natureza desconhecida, então designado P-TEFb. Em concordância com o modelo proposto, o P-TEFb mostrou ser o verdadeiro alvo de ligação do DRB (MARSHALL et al., 1995). A seguir, a caracterização do P-TEFb revelou uma atividade enzimática de proteína quinase capaz de fosforilar o CTD da RNAP II (MARSHALL et al., 1996). É importante mencionar que alguns anos antes havia sido descoberto o papel do fator TFIIF na fosforilação do CTD durante a

iniciação, promovendo a passagem da polimerase à fase de alongação (LU et al., 1992). Assim, a ação do P-TEFb corresponde a uma segunda etapa de fosforilação do CTD, responsável por desencadear o modo de alongação produtiva. Embora muitos estudos tenham acrescentado detalhes ao processo de regulação da RNAP II, a validade deste modelo permanece aceita até hoje, e o P-TEFb continua sendo o único fator conhecido capaz de reativar as polimerases estagnadas após a iniciação (PRICE, 2008).

O reconhecimento de fatores com as propriedades previstas de N-TEF só foi possível mais tarde, com a descoberta do DSIF - *DRB-sensitivity inducing factor* (WADA et al., 1998) e NELF - *negative elongation factor* (YAMAGUCHI et al. 1999). Nos ensaios *in vitro*, embora nenhum dos fatores produzisse qualquer efeito sozinho, a combinação de DSIF e NELF mostrou bloquear a alongação produtiva da RNAP II. A atividade do P-TEFb reverte o efeito negativo destes dois fatores, o que acabou levando à descoberta de que as subunidades Spt5 do DSIF e RD do NELF também são seus alvos de fosforilação (KIM et al., 2001; FUJINAGA et al., 2004). A fosforilação de Spt5 converte o DSIF de um fator negativo para um fator positivo de alongação, e a fosforilação de RD provoca a dissociação do NELF da RNAP II (PETERLIN et al., 2006). Assim, o P-TEFb desempenha um papel múltiplo na passagem da RNAP II à alongação produtiva, removendo o bloqueio imposto por DSIF e NELF ao mesmo tempo em que prepara o CTD para recrutar os fatores de alongação e processamento do RNAm. Neste processo também ocorre a liberação do restante dos fatores que se associam à polimerase inicialmente para formar o complexo de pré-iniciação (*PIC*) no promotor. A Figura 9 ilustra a função do P-TEFb na superação do bloqueio e passagem da RNAP II à alongação produtiva.

O mecanismo de pausa e retomada da transcrição pela RNAP II fornece uma oportunidade de regulação da expressão gênica, através do controle da passagem à alongação produtiva. De fato, hoje sabemos que um conjunto importante de genes dos eucariotos superiores têm sua expressão regulada neste ponto da síntese de RNAm (PRICE, 2008). O avanço da pesquisa nas últimas décadas tornou claro que o modo de controle da expressão destes genes é bem mais comum do que se imaginava a princípio: Nas células de *Drosophila* estima-se que cerca de 10% do total de genes apresentem a RNAP II pausada pouco além do promotor, depois de iniciar a síntese do transcrito (PRICE, 2008).

Figura 9 – Função do P-TEFb na fosforilação do CTD, DSIF e NELF. Durante a iniciação, o CTD da RNAP II é fosforilado pelas CDKs 7 e 8 (componentes do TFIIF e Mediador, respectivamente). Esta fosforilação inicial faz com que a polimerase comece a síntese do transcrito e recruta os fatores de capeamento (processamento 5'), que adicionam o *cap* ao RNA nascente. Porém, devido à influência do DSIF e NELF, a RNAP II sofre uma pausa após adicionar algumas poucas dezenas de nucleotídeos ao transcrito. O P-TEFb atua sobre a RNAP II pausada **(a)**, fosforilando o seu CTD e os fatores negativos DSIF e NELF **(b)**. A fosforilação provoca a dissociação do NELF, converte o DSIF em um fator positivo de elongação e promove o recrutamento dos fatores de elongação e processamento do RNAm **(c)**. Depois desta segunda etapa de fosforilação, a RNAP II entra no modo de elongação produtiva e retoma a transcrição até a síntese do RNAm completo.

Fonte: Autoria própria.

Agora está bastante claro que o princípio de funcionamento do CTD consiste na mudança em seus padrões de fosforilação ao longo das sucessivas fases da transcrição, e na capacidade destes diferentes padrões em recrutar os fatores específicos necessários a cada estágio. O CTD encontra-se originalmente desfosforilado quando a RNAP II inicia a montagem do complexo de pré-iniciação (*PIC*) nos promotores. Além dos fatores gerais de transcrição, que incluem o TFIIF (CDK7), também são recrutadas proteínas ativadoras específicas de cada gene e proteínas co-ativadoras, responsáveis por mediar a interação entre as proteínas ativadoras e o *PIC*. O principal co-ativador eucariótico é um grande

complexo denominado Mediador, que contém a CDK8 entre suas dezenas de componentes, e possui uma afinidade específica pela forma desfosforilada do CTD. O complexo Mediador é um forte estimulador da atividade quinásica do TFIIH, além de atuar diretamente na fosforilação do CTD por meio da CDK8. Esta etapa inicial envolve a fosforilação de sítios Ser5 e Ser7, sendo o TFIIH responsável pela maior parte da fosforilação em Ser5 e a totalidade da fosforilação em Ser7 (BURATOWSKI, 2009; CZUDNOCHOWSKI et al., 2012; BARTKOWIAK et al., 2010; BAUMLI et al., 2012).

Vários eventos são funcionalmente relacionados à mudança no estado de fosforilação do CTD. Por exemplo, é sabido que as enzimas participantes do processamento 5' (capeamento) do transcrito se ligam especificamente ao CTD fosforilado em Ser5 (ALBERTS et al., 2010). Outros fatores importantes para a iniciação também devem fazer uso da fosforilação de Ser5 para determinar sua associação ao CTD, como enzimas modificadoras de histonas. Além do recrutamento dos fatores necessários ao início da transcrição, a fosforilação do CTD leva à dissociação do Mediador, liberando a polimerase da região promotora. Enquanto a RNAP II se desloca no sentido 5' da fita molde, o complexo Mediador permanece ligado ao promotor de modo a facilitar o recrutamento de mais moléculas de polimerase para realizar novas rodadas de iniciação (BURATOWSKI, 2009).

Uma das consequências mais importantes da atividade do TFIIH e fosforilação do CTD durante a iniciação é o recrutamento dos fatores DSIF e NELF, que funcionam em conjunto para ocasionar a pausa da RNAP II pouco após ter começado a transcrição (GLOVER-CUTTER et al., 2009; LAROCHELLE et al., 2012). Embora a CDK9 promova a fosforilação da posição Ser5, assim como as CDKs 7 e 8 precedentes, deve haver uma diferença fundamental que justifique seu papel exclusivo na liberação das moléculas de RNAP II pausadas após a iniciação. Um dos motivos é o fato de que o CTD não representa o único alvo importante do P-TEFb na passagem à elongação produtiva, pois seu efeito sobre a RNAP II depende também da fosforilação de DSIF e NELF. Todas estas modificações que o P-TEFb realiza são fundamentais para os eventos seguintes da transcrição, representando um ponto de verificação (*checkpoint*) que regula a capacidade da polimerase de proceder à elongação completa do transcrito (BURATOWSKI, 2009).

A atividade da CDK9 é o fator determinante na passagem da polimerase à elongação produtiva, entretanto ela não pode ser a única quinase envolvida no processo, pois uma característica central das moléculas de RNAP II em fase de elongação é a intensa fosforilação dos sítios Ser2 no CTD (CZUDNOCHOWSKI et al., 2012; BARTKOWIAK et al., 2010). A má interpretação da relação entre elongação produtiva e fosforilação em Ser2 levou à conclusão errônea de que a CDK9 seria uma Ser2 CTD quinase. Agora que as CDKs 12 e 13 foram reconhecidas como as legítimas responsáveis pela fosforilação de Ser2, torna-se necessário rever o mecanismo de passagem à elongação produtiva, que deve envolver seu recrutamento ou ativação em consequência dos eventos desencadeados pelo P-TEFb. Uma vez ultrapassado este ponto crítico, a presença de sítios Ser2 fosforilados promove várias interações novas do CTD com fatores de transcrição.

Já se sabe que a fosforilação de Ser2 recruta fatores de poliadenilação (AHN et al., 2004) e remodeladores de cromatina (YOH et al., 2007), que permitem à polimerase elongar através dos nucleossomos. Ao contrário do aumento observado nos níveis de sítios Ser2 fosforilados, os níveis de fosforilação em Ser5 diminuem enquanto a RNAP II avança na elongação do transcrito. Assim, os níveis relativos de Ser2 e Ser5 fornecem a informação posicional que é transmitida aos fatores que interagem com o CTD para regular as diferentes fases do processo (BURATOWSKI, 2009). Além das variações ao longo do percurso na transcrição de um gene, os padrões de fosforilação do CTD também variam entre diferentes classes de genes, o que deve refletir uma especialização na sua regulação individual (TIETJEN et al., 2010). De modo geral, o resultado da atuação conjunta das CTD quinases durante a transcrição é o estabelecimento deste duplo gradiente de fosforilação envolvendo as posições Ser2 e Ser5.

A descoberta da participação essencial do P-TEFb na transcrição do genoma do HIV teve grande importância para a elucidação dos mecanismos moleculares envolvidos na infecção de células humanas pelo vírus. Além disso, a transcrição do HIV tornou-se o principal modelo para estudo da função do P-TEFb na regulação da RNAP II. A primeira indicação de que a transcrição do HIV apresentava um mecanismo peculiar de regulação veio com o reconhecimento de que a proteína viral Tat (trans-ativadora da transcrição) não influencia a iniciação da síntese do RNA do vírus, mas é indispensável para a produção de

transcritos com mais que algumas dezenas de nucleotídeos. A extremidade 5' do genoma do HIV apresenta uma sequência designada *LTR* (*long terminal repeat*) que é responsável pela incorporação do provírus ao DNA do hospedeiro e controle da sua expressão. O *LTR* do HIV contém um típico promotor eucariótico para a RNAP II - incluindo um *TATA box* - como ocorre geralmente entre os retrovírus. Após a RNAP II iniciar a transcrição no promotor do HIV, ocorre um bloqueio na elongação que impede a polimerase de avançar além do ponto em que a cadeia de RNA atinge 59 nucleotídeos de extensão (KAO et al., 1987).

O pequeno segmento inicial transcrito pela RNAP II antes de sofrer o bloqueio contém uma sequência designada TAR (*transactivation response element*), com a qual a Tat interage para produzir seu efeito estimulador sobre a transcrição do HIV. É muito comum em células eucarióticas que a transcrição de um gene seja estimulada pela ligação de proteínas ativadoras a sequências regulatórias de DNA (*enhancers*) que complementam a função do promotor. No entanto, para surpresa geral dos pesquisadores, foi demonstrado que o alvo de ligação da Tat não era a sequência TAR no DNA, mas sim o produto da sua transcrição em RNA. O RNA TAR assume uma estrutura secundária em grampo que se liga com elevada afinidade ao complexo formado pela Tat e o P-TEFb contendo a ciclina T1. Como a função do TAR é equivalente àquela das sequências reforçadoras encontradas no DNA, ele tem sido denominado *RNA enhancer* (PETERLIN et al., 2006).

A Tat interage exclusivamente com a ciclina T1 humana, devido à existência de um motivo próprio de reconhecimento da Tat (TRM). Este motivo contém um resíduo-chave de cisteína (Cys261) que participa da coordenação de um cátion Zn^{2+} essencial para a associação entre a ciclina T1 e a Tat (GARBER et al., 1998). O resultado desta interação é o recrutamento do P-TEFb para o RNA TAR, favorecendo o contato físico com seus alvos de fosforilação localizados na polimerase pausada (o processo está ilustrado na Figura 10). A presença do P-TEFb desencadeia a liberação da RNAP II por promover as fosforilações do CTD, DSIF e NELF, assim como discutido no caso de genes celulares. Então ocorrerá a síntese de transcritos completos do HIV, capazes de originar todos os produtos virais por *splicing* diferencial. Além da expressão das proteínas do HIV, a transcrição também é essencial à replicação do vírus, pois o transcrito completo será incorporado como RNA genômico por novas partículas virais (HOFFMANN et al., 2012).

Figura 10 – Recrutamento do P-TEFb pela proteína Tat do HIV. Ao iniciar a transcrição do HIV, a RNAP II é incapaz de proceder à elongação produtiva, e sofre um bloqueio após adicionar algumas dezenas de nucleotídeos ao transcrito. O curto segmento de RNA sintetizado antes da pausa da RNAP II contém uma sequência conhecida como TAR, que assume uma estrutura secundária em *hairpin*. O transativador viral Tat reconhece o RNA TAR, mas apenas sua ligação não é capaz de superar o bloqueio da RNAP II e induzir a elongação produtiva. A Tat interage também com a ciclina T1 humana, o que aumenta muito a sua afinidade pelo RNA TAR. O efeito desta associação é o recrutamento do P-TEFb para o provírus do HIV, onde a RNAP II se encontra pausada após transcrever o RNA TAR. Assim como atua nos genes celulares, o P-TEFb fosforila o CTD, DSIF e NELF, levando a RNAP II a retomar a síntese do transcrito do HIV no modo de elongação produtiva. O RNA transcrito funciona simultaneamente como RNAm (para codificar as proteínas virais) e como RNA genômico (para ser incorporado como material genético a novas partículas do vírus).

Fonte: Adaptado de Wei et al. (1998, p. 459) e Garriga et al. (2004, p. 17).

Devido à presença do resíduo essencial de cisteína, apenas a ciclina T1 de primatas é capaz de interagir com a Tat. A ciclina T1 murina possui um resíduo de tirosina na posição equivalente à Cys261 humana, o que impossibilita sua associação com a Tat quando é realizada a transfecção do HIV em células de camundongo. Contudo, a transativação do HIV mediada pela Tat pode ser viabilizada nas células murinas através da co-transfecção da ciclina T1 humana, ou substituição do resíduo de tirosina por cisteína no motivo TRM de sua própria ciclina T1 (GARBER et al., 1998). A estrutura cristalográfica do complexo Tat/P-TEFb só pôde ser elucidada recentemente (TAHIROV et al., 2010). Ela mostra que a Tat possui dois sítios atípicos de ligação ao zinco, localizados em um pequeno trecho notavelmente rico em resíduos de cisteína. Um dos cátions Zn^{2+} é coordenado apenas por resíduos da própria Tat,

enquanto a coordenação do segundo envolve também a cisteína localizada no motivo TRM da ciclina T1.

O modelo de interação com a Tat ilustra bem o princípio geral da função do P-TEFb na célula. Como no caso da transativação do HIV, o recrutamento do P-TEFb estimula a transcrição de uma variedade de genes celulares que são naturalmente reprimidos pelo bloqueio na passagem da RNAP II à elongação produtiva. A expressão destes genes é modulada por fatores de transcrição específicos, capazes de interagir simultaneamente com o P-TEFb e elementos locais de regulação, que podem ser sequências de DNA (*enhancers*), RNA (*RNA enhancers*), ou ainda proteínas recrutadas previamente pelos genes (PETERLIN et al., 2006). Muitos fatores que interagem com o P-TEFb para mediar seu recrutamento por genes-alvo já foram identificados, e alguns exemplos serão apresentados mais adiante nesta seção. A lista completa de fatores que recrutam o P-TEFb conta atualmente com dezenas de membros e continua crescendo em ritmo acelerado.

Alguns estudos investigaram o efeito das diferentes modalidades de recrutamento do P-TEFb sobre a transcrição de genes-alvo. Afinal, a descoberta do sistema de transativação do HIV havia levantado a suspeita de que um mecanismo análogo pudesse ocorrer na regulação normal de genes eucarióticos. Era preciso determinar se o P-TEFb seria capaz de estimular de modo geral a atividade da RNAP II quando recrutado diretamente (*via* RNA ou DNA) para interagir com o complexo de transcrição. Com este objetivo, foram realizados ensaios de prendimento heterólogo (*tethering*) do P-TEFb a elementos regulatórios de DNA ou RNA introduzidos em regiões apropriadas de um gene repórter. A estratégia consiste em fusionar à proteína de interesse um domínio específico de ligação ao DNA ou RNA, o que promove seu recrutamento direto para sítios que contenham as sequências nucleotídicas correspondentes. Experimentos utilizaram os sistemas Gal4/UAS e Rev/RRE para obter o prendimento heterólogo do P-TEFb a sítios específicos de DNA e RNA, respectivamente (FUJINAGA et al., 1998; TAUBE et al., 2002; LIN et al., 2002).

O recrutamento artificial do P-TEFb *via* DNA demonstrou produzir o mesmo efeito atribuído ao seu papel postulado de co-ativador, no qual atua recrutado indiretamente ao DNA por proteínas ativadoras ligadas a *enhancers* (TAUBE et al., 2002). Sendo assim, o prendimento heterólogo do P-TEFb ao DNA faz com que seus sítios de ligação passem a

satisfazer os critérios clássicos para a definição de *enhancers* (PETERLIN et al., 2006). O prendimento do P-TEFb ao RNA também induziu um aumento significativo na expressão do gene repórter. O aumento observado desapareceu quando a CDK9 foi substituída nos ensaios pelo mutante D167N, sem atividade quinásica (LIN et al., 2002). Apenas proteínas que interagem com o P-TEFb têm mostrado a capacidade de induzir a transcrição em sistemas de recrutamento *via* RNA. Por este motivo, o prendimento ao RNA já foi sugerido como um ensaio genético de identificação dos fatores que recrutam o P-TEFb para promover a ativação transcricional (PETERLIN et al., 2006).

Outro importante aspecto do P-TEFb desvendado pelos ensaios de prendimento heterólogo refere-se a uma diferença essencial nas propriedades de recrutamento da ciclina K em relação às ciclinas do tipo T. Os experimentos mostraram que ambos os tipos de ciclinas (T e K) podem funcionar ativando a transcrição quando recrutadas *via* RNA, por meio do sistema Rev/RRE. Entretanto, apenas as ciclinas do tipo T se provaram capazes de induzir a transcrição na modalidade de recrutamento *via* DNA, pelo sistema Gal4/UAS (LIN et al., 2002). Este resultado foi associado à observação de que a ciclina T1 perde a habilidade de ativar a transcrição no mesmo sistema quando são realizadas deleções envolvendo seu motivo rico em resíduos de histidina, situado entre as posições 481 e 551 da cadeia polipeptídica. A caracterização subsequente do trecho indicado revelou que ele funciona *in vivo* como um domínio de ligação ao CTD da RNAP II (TAUBE et al., 2002). Quando esta sequência rica em histidina foi fundida à extremidade C-terminal da ciclina K, o híbrido tornou-se capaz de ativar a transcrição pelo recrutamento *via* DNA (LIN et al., 2002).

Hoje são conhecidos diversos exemplos de especialização entre os diferentes heterodímeros CDK9/ciclina, aos quais temos nos referido indistintamente como P-TEFb. Considerando-se que um complexo P-TEFb contém um dos quatro diferentes tipos de ciclina (T1, T2a, T2b e K) associado a uma das duas isoformas da CDK9 (42 ou 55 kDa), a análise combinatória simples fornece 8 potenciais heterodímeros distintos (KOHOUTEK, 2009). A rigor, cada um destes complexos deve funcionar de modo independente, atuando sobre um subconjunto particular dos genes que são regulados pelo P-TEFb. Obviamente poderá ocorrer algum grau de sobreposição entre os subconjuntos de genes regulados por duas ou mais formas heterodiméricas, inclusive o recrutamento de formas distintas do P-TEFb para

um mesmo gene sob diferentes contextos de regulação. As diferentes formas do P-TEFb ainda admitem o desenvolvimento de mecanismos de regulação dirigidos separadamente a cada uma delas, como de fato constatamos em alguns casos de modificações covalentes na CDK ou ciclina. Obviamente, a especificidade do programa de transcrição ativado pelo P-TEFb em cada célula dependerá do seu repertório particular de fatores induzíveis capazes de recrutá-lo (GARRIGA et al., 2004).

Por fim, é preciso considerar a variedade dos processos de regulação que coordenam o funcionamento do P-TEFb em todas as células. A Figura 11 apresenta os principais mecanismos responsáveis por modular a atividade do complexo P-TEFb em diferentes contextos da sua regulação. O primeiro nível de controle sobre a atividade celular do P-TEFb é exercido pela regulação da expressão de suas subunidades (PETERLIN et al., 2006). Os heterodímeros CDK9/ciclina são bastante estáveis, ao contrário da CDK9 monomérica livre que é rapidamente degradada quando excede a quantidade disponível de ciclinas. A formação dos complexos P-TEFb é mediada por uma via de enovelamento assistido da qual participam as chaperonas Hsp70 e Hsp90/Cdc37. A CDK9 recém-sintetizada é transferida sequencialmente da Hsp70 para o complexo Hsp90/Cdc37, e deste para uma ciclina formando o P-TEFb (O'KEEFFE et al., 2000). As etapas enumeradas de I a VI na Figura 11 representam esta fase inicial de montagem dos complexos P-TEFb na célula.

Uma vez produzida, a forma enzimaticamente ativa da CDK9 deve sofrer auto-fosforilação espontânea na Thr186 do segmento de ativação (correspondente à etapa VII no diagrama). Mesmo contendo a CDK9 ativa, o P-TEFb livre não é capaz de fosforilar seus principais alvos – CTD da RNAPII, DSIF e NELF – com os quais só poderá interagir após ser recrutado ao complexo de transcrição. No estado heterodimérico livre, o P-TEFb pode atuar na fosforilação de alvos nucleares secundários (etapa XI na Figura 11), como pRb (GRAÑA et al., 1994) e p53 (RADHAKRISHNAN et al., 2006). Embora a CDK9 deva se auto-fosforilar com a formação do P-TEFb, seu recrutamento ao complexo de transcrição requer a desfosforilação prévia da Thr186 (representada pela etapa XXII). A presença do fator TFIIH no complexo de pré-iniciação impede a auto-fosforilação da Thr186 e consequente ativação enzimática da CDK9 de ocorrer antes da sua dissociação (ZHOU et al., 2001).

Figura 11 – Panorama da função celular e regulação do P-TEFb. O diagrama representa um esboço da função global do P-TEFb na célula eucariótica, reunindo os principais mecanismos regulatórios descritos na literatura. A via principal - que envolve as etapas de XII a XXVIII no diagrama - corresponde ao recrutamento do P-TEFb para o complexo de transcrição, onde ele realiza sua função como fator positivo de elongação da RNAP II. Embora a ordem das etapas relacionadas ao recrutamento do P-TEFb não seja precisamente conhecida, ela envolve uma série de modificações que inativam enzimaticamente a CDK9 ao mesmo tempo em que aproximam o P-TEFb do complexo de pré-iniciação (CPI). Até o momento da pausa transcricional da RNAP II, todas as modificações inibitórias terão sido desfeitas, promovendo a ativação *de novo* da quinase no interior do complexo de elongação.

Fonte: Autoria própria.

A desfosforilação da Thr186 pelas proteínas fosfatases PP1 ou PPM1A – etapa VIII (AMMOSOVA et al., 2005; WANG et al., 2008) segue a mesma tendência apresentada por várias outras modificações covalentes que dirigem o recrutamento do P-TEFb para os complexos de transcrição da RNAP II. Tais modificações, como a acetilação da CDK9 em Lys44/48 – etapa XII (SABÒ et al., 2008), a fosforilação em Ser175 – etapa XIV (AMMOSOVA et al., 2011; MBOYNE et al., 2013), e a associação a Brd4 – etapa XIV – que provoca a fosforilação em Thr29 – etapa XXV (ZHOU et al., 2009) possuem em comum o efeito paradoxal de inibir a atividade enzimática da CDK9 ao mesmo tempo em que aumentam a transcrição de genes-alvo do P-TEFb. Portanto, o P-TEFb é recrutado em uma forma enzimaticamente inativa para o complexo de pré-iniciação e assim permanece inibido até que uma série de eventos promova a ativação *de novo* da CDK9 no princípio da fase de alongação da RNAP II – etapa XXVII (KOHOUTEK, 2009). A inativação do P-TEFb antes do seu recrutamento ao complexo de transcrição provavelmente desempenha um papel organizador na montagem do complexo, evitando a fosforilação prematura das proteínas-alvo até o momento certo de liberação (ZHOU et al., 2009).

Com a descoberta de que a CDK9 sofre ubiquitinação pelo complexo SCF^{skp2} ubiquitina-ligase – mostrada pelas etapas XVIII e XIX – foi sugerido que a atividade celular do P-TEFb pudesse ser regulada pela degradação proteolítica da quinase em sincronia com o ritmo do ciclo celular (KIERNAN et al., 2001). De fato, a degradação controlada de ciclinas pelo proteossomo é um mecanismo importante de regulação da atividade de complexos CDK/ciclina durante o ciclo celular, e muitas ciclinas contêm motivos *destruction box* que marcam proteínas para serem alvos de ubiquitina-ligases (ALBERTS et al., 2010). O caso da CDK9 foi o primeiro exemplo conhecido no qual um complexo ubiquitina-ligase é recrutado por uma ciclina (motivo *PEST* da ciclina T1), mas realiza a ubiquitinação da CDK associada (KIERNAN et al., 2001). Estudos posteriores, entretanto, indicaram que a expressão endógena da CDK9 não é regulada por esta ubiquitina-ligase de modo dependente do ciclo celular (GARRIGA et al., 2003). Embora ainda haja a possibilidade de que outra ubiquitina-ligase semelhante atue mediando a degradação proteolítica da CDK9 ou ciclina T1 (GARRIGA et al., 2004), a única função atribuída à poliubiquitinação da CDK9 pelo SCF^{skp2} consiste na estabilização do complexo ternário P-TEFb/Tat/TAR e correspondente aumento na

transcrição do HIV (BARBORIC et al., 2005). Por este motivo, a ubiquitinação foi incluída na via principal (de recrutamento do P-TEFb para o complexo de pré-iniciação) na Figura 11.

O principal desvio que ocorre no recrutamento do P-TEFb ativo para os complexos de transcrição é causado pela sua ligação inibitória à ribonucleoproteína snRNP 7SK – etapa IX (NGUYEN et al., 2001; YANG et al., 2001; MICHELS et al., 2003; YIK et al., 2003). O cerne do complexo snRNP 7SK é formado pelo snRNA 7SK associado a um dímero de HEXIM1 e/ou HEXIM2 e pode se ligar a duas unidades do P-TEFb (necessariamente fosforilado na Thr186) (LI et al., 2005). Embora HEXIM1 e HEXIM2 possam formar homo ou heterodímeros livres, é necessário que estes se associem primeiro ao snRNA 7SK para que possam interagir com o P-TEFb. A ligação a proteínas inibitórias (designadas CKIs) é um mecanismo comum de regulação das CDKs, mas este é o primeiro caso conhecido de uma CKI que possui RNA associado (KOHOUTEK, 2009). Cerca de metade do P-TEFb celular encontra-se associado à snRNP 7SK na forma de um complexo inativo, que funciona como um grande estoque do fator no núcleo. A liberação do P-TEFb deste complexo ocorre sempre que a transcrição é dramaticamente inibida, como quando as células são tratadas com luz UV, actinomicina D, ou inibidores da CDK9 – etapa X (PETERLIN et al., 2006).

Praticamente todo o P-TEFb celular que não se encontra ligado ao complexo snRNP 7SK está associado à proteína de duplo bromodomínio Brd4 (YANG et al., 2005). A associação do P-TEFb com estes complexos é mutuamente exclusiva, provavelmente porque a fosforilação da CDK9 em Ser175 impede a ligação à snRNP 7SK ao mesmo tempo em que favorece a interação com Brd4 (MBOYNE et al., 2013). Portanto, se considera que o P-TEFb existe no núcleo na forma de dois complexos distintos: Um complexo maior (inativo) que envolve a snRNP 7SK e um complexo menor (ativo) que envolve Brd4 (WU et al., 2007). Integrante do complexo Mediador, Brd4 tem a capacidade de se associar a histonas acetiladas, como as demais proteínas que possuem bromodomínio (JANG et al., 2005). Desta maneira, além das modalidades alternativas de recrutamento *via* DNA e RNA já consideradas, o fator Brd4 atua recrutando o P-TEFb *via* cromatina para diversos genes-alvo marcados pela acetilação em H3 e H4 (PETERLIN et al., 2006; JANG et al., 2005).

Como Brd4 apresenta duplo bromodomínio, foi sugerido que este interagisse simultaneamente com as histonas e a forma acetilada da CDK9 (FU et al., 2007). Entretanto,

a interação de Brd4 com o P-TEFb parece ocorrer através da ciclina T1, mesmo que dependa indiretamente da fosforilação da CDK9 em Ser175 (YANG et al., 2005; JANG et al., 2005). É importante mencionar ainda que o efeito ativador sobre a transcrição da RNAP II promovido por Brd4 também pode ser constatado em experimentos *in vitro*, nos quais o DNA não se encontra organizado na forma de cromatina. Então, deve haver um mecanismo alternativo para o recrutamento do P-TEFb mediado por Brd4 que seja independente de cromatina, o qual provavelmente está relacionado ao complexo Mediador (YANG et al., 2005).

Desde a formação do complexo de pré-iniciação, o P-TEFb permanece associado a Brd4 até o início da fase de alongação da RNAP II. A auto-fosforilação inibitória da CDK9 em Thr29 é induzida pela ligação de Brd4, mas só ocorre após a incorporação de ambos os fatores ao PIC. Uma mudança crítica na composição dos complexos de transcrição é observada no início da fase de alongação – equivalente ao intervalo entre as posições +14 e +36 para o HIV-1 – que inclui a liberação de TFIIH e Brd4, os quais mantêm a CDK9 inativa por impedir a fosforilação da Thr186 e desfosforilação da Thr29, respectivamente. Para a desfosforilação da Thr29 é necessária ainda a associação da fosfatase PP2A ao complexo de transcrição, que ocorre neste mesmo intervalo (ZHOU et al., 2001; ZHOU et al., 2009). Com a dispersão dos fatores inibitórios, ocorre a ativação *de novo* da CDK9 no interior do complexo de alongação, evitando desta maneira que o P-TEFb fosforile seus alvos prematuramente durante a iniciação.

A interação do P-TEFb com Brd4 finalmente estabeleceu uma conexão entre a CDK9 e o ciclo celular, a qual havia sido desacreditada por muito tempo devido ao reconhecimento de que os níveis de expressão e atividade quinásica da CDK9 não variam ao longo do ciclo (GARRIGA et al., 2003). De fato, foi demonstrado um aumento dramático na interação entre o P-TEFb e Brd4 desde o final da mitose até o início da fase G₁ (YANG et al., 2008). Este processo se revelou essencial para o recrutamento do P-TEFb e expressão de genes-chave essenciais à transição de fase G₁/S (MOCHIZUKI et al., 2008). Outra importante consequência da interação com Brd4 é a sugestão de um papel para o P-TEFb na transmissão de memória epigenética para as células-filhas na divisão celular. Afinal, Brd4 apresenta a propriedade de permanecer associada aos cromossomos durante a mitose, e poderia assim representar uma marca de memória para os genes transcricionalmente ativos (YANG et al., 2008).

1.5 Apresentação e genética da CDK11

A CDK11 humana foi originalmente identificada em 1990 como uma proteína de 58 kDa cuja superexpressão em células murinas resulta na suspensão do progresso do ciclo celular através da fronteira telófase/G₁, ocasionando anomalias estruturais na mitose e citocinese (BUNNELL et al., 1990). A sequência predita de aminoácidos revelou se tratar de uma nova proteína relacionada à família CDC de quinases reguladoras do ciclo celular, evidenciando a presença de um motivo PITSLRE na região N-terminal correspondente ao *box* PSTAIRE das CDKs 1 e 2. Esta proteína quinase ficou então conhecida como PITSLRE por mais de uma década até a descoberta da sua associação às ciclinas do tipo L (DICKINSON et al., 2002), quando passou a ser designada CDK11.

A isoforma de 58 kDa original representa apenas uma dentre mais de 20 diferentes variedades da CDK11 (com massa molecular entre 50 e 110 kDa) que são expressas a partir de dois genes altamente similares, originados por um evento recente de duplicação gênica. Os genes CDC2L1 e CDC2L2 – antes denominados PITSLRE B e A, respectivamente – estendem-se por um trecho com tamanho aproximado de 100 kb no locus 1p36.2, situado próximo à extremidade do braço curto do cromossomo 1 humano (XIANG et al., 1994). Em células de camundongo existe um único gene ortólogo da CDK11 (CDC2L), encontrado em uma região sintênica do cromossomo 4 murino (LI et al., 2004). Os genes da CDK11 mostraram ser constitutivamente expressos em uma ampla variedade de tecidos humanos e murinos.

É preciso esclarecer que a designação “CDK11” também foi empregada no passado para se referir ao produto do gene humano CDC2L6, localizado no cromossomo 6. Modernamente, a proteína codificada por CDC2L6 passou a ser referida como CDK19 na literatura (MALUMBRES et al., 2009). A CDK19 apresenta maior identidade de sequência com a CDK8, podendo inclusive substituí-la como subunidade quinase do complexo Mediador. A CDK11, por outro lado, apresenta-se filogeneticamente mais próxima da CDK10 (PISSLRE). Os genes CDC2L1 e CDC2L2 apresentam um número total de 21 éxons cada, sendo que os 11 éxons finais (3') da sequência estão invariavelmente presentes em todos os RNAs transcritos. A tradução destes 11 éxons produz as isoformas com massa aparente de 58 kDa designadas CDK11^{p58}. Os 10 éxons restantes são indicados pelas letras A-I e estão presentes

em diferentes combinações na porção 5' dos RNAs devido à ocorrência de *splicing* alternativo do transcrito primário (XIANG et al., 1994). As isoformas da CDK11 que contêm estes éxons adicionais possuem massa molecular aparente entre 65 e 110 kDa, sendo aquelas de maior massa – referidas coletivamente como CDK11^{p110} – as variedades mais bem caracterizadas no seu grupo.

O domínio quinase da CDK11 encontra-se na região C-terminal da sequência, enquanto a extensa porção N-terminal apresenta diversos elementos regulatórios identificados, que incluem três motivos PEST, um sinal de localização nuclear e uma região acídica caracterizada pela elevada prevalência de resíduos de ácido glutâmico (BUNNELL et al., 1990; XIANG et al., 1994). Existe também uma variedade celular designada CDK11^{p46} que corresponde ao fragmento C-terminal oriundo da clivagem proteolítica das isoformas CDK11^{p110} e CDK11^{p58} por caspases, consistindo basicamente no domínio quinase livre da CDK11 (SHI et al., 2003; CHEN et al., 2003; MIKOLAJCZYK et al., 2004). A espécie CDK11^{p60}, por outro lado, se refere ao fragmento N-terminal (complementar à CDK11^{p46}) destituído do domínio quinase, produzido quando as isoformas CDK11^{p110} são clivadas por caspases (FENG et al., 2005).

Como a diversidade de isoformas da CDK11 permanece subexplorada, praticamente todos os trabalhos publicados na área são dedicados às variedades CDK11^{p110}, CDK11^{p58} e CDK11^{p46}. É curioso notar que o contexto funcional relacionado a cada uma destas três isoformas parece ser completamente independente das demais (LOYER et al., 2005). A CDK11^{p110} funciona na regulação da transcrição pela RNAP II (DROGAT et al., 2012) e *splicing* do pré-RNA (HU et al., 2003). Por sua vez, a CDK11^{p58} somente é expressa durante as fases G₂ e M do ciclo celular (CORNELIS et al., 2000), e atua na regulação de eventos mitóticos (PETRETTI et al., 2006). Finalmente, a CDK11^{p46} é produzida pela ação de caspases sobre as isoformas maiores e participa como quinase mediadora na sinalização pró-apoptótica (BEYAERT et al., 1997; SHI et al., 2003).

As ciclinas do tipo L também são produtos da expressão de dois genes humanos distintos mas muito similares, denominados Ania-6a e Ania-6b pela sua relação próxima com a ciclina Ania-6 (ortóloga murina da ciclina L). As proteínas codificadas por estes dois genes são designadas ciclinas L1 ou L2, respectivamente. A ciclina L1 existe como um conjunto de

três isoformas diferentes – L1 α , L1 β e L1 γ – enquanto a ciclina L2 é representada pelas isoformas L2 α , L2 β 1 e L2 β 2 (LOYER et al., 2008). As isoformas L1 α e L2 α são maiores e contêm um extenso domínio C-terminal com repetições do dipeptídeo Ser-Arg, uma característica de fatores de *splicing* da família de proteínas SR. As ciclinas L β e L γ são formas truncadas que consistem basicamente no *box* ciclina N-terminal, o qual apresenta maior identidade de sequência com aquele da ciclina K (DICKINSON et al., 2002).

Considerando-se o número de diferentes isoformas conhecidas da CDK11 e ciclina L, é provável que existam *in vivo* dezenas de complexos heterodiméricos CDK11/ciclina L com combinações distintas destas subunidades. Já foi determinado, por exemplo, que as ciclinas L1 α e L2 α formam complexos endógenos com todas as três isoformas principais da CDK11, enquanto L1 β se associa preferencialmente à CDK11^{p58} e CDK11^{p46} (LOYER et al., 2008). Ainda, em relação às ciclinas que interagem com a CDK11, é preciso acrescentar que já foi demonstrada a interação da ciclina D3 exclusivamente com as isoformas CDK11^{p58}. As ciclinas do tipo D (D1, D2 e D3) interagem normalmente com as CDKs 4 e 6 na fase G₁, e apenas D3 mostrou interagir com a CDK11^{p58} no estágio G₂/M do ciclo celular (ZHANG et al., 2002).

Ao contrário das diversas isoformas CDK11^{p110}, a CDK11^{p58} não representa um produto de *splicing* diferencial do RNAm da CDK11. Como mencionado, todos os transcritos de RNAm originados dos genes da CDK11 possuem os mesmos 11 éxons na sua terminação 3', os quais codificam a proteína CDK11^{p58}. A tradução da CDK11^{p58} é iniciada a partir de um códon AUG interno no RNAm, que encontra-se na mesma pauta de leitura em relação ao ORF da CDK11^{p110}. Portanto, a CDK11^{p58} corresponde a um fragmento C-terminal das isoformas CDK11^{p110} cuja tradução é iniciada através de um sítio interno de entrada para o ribossomo (*IRES*). O elemento *IRES* identificado no RNAm da CDK11 apresenta uma sequência de Kozak discernível flanqueando o códon de iniciação da CDK11^{p58} (CORNELIS et al., 2000).

A síntese da isoforma CDK11^{p58} apenas é realizada durante o estágio G₂/M porque depende da presença de um fator trans-ativador específico, denominado Unr (*upstream of N-ras*), que experimenta um aumento na própria expressão nesta mesma fase do ciclo

celular. Também devem contribuir para a tradução da CDK11^{p58} a fosforilação da subunidade α do fator de iniciação eucariótico eIF2 que se observa em G₂/M, e o fato de ser a tradução convencional – dependente de *cap* 5' – marcadamente inibida na célula durante a mitose (TINTON et al., 2005). O *IRES* da CDK11 foi o primeiro a ser descoberto no interior da sequência codificante e fora da região 5' não traduzida de um RNAm. Ainda, a CDK11^{p58} forneceu o primeiro exemplo de controle traducional na expressão de proteínas reguladoras do ciclo celular, uma vez que todos os mecanismos classicamente associados à regulação do ciclo operam no nível transcricional ou pós-traducional (CORNELIS et al., 2000).

Assim como acontece em relação a outras CDKs, a estabilidade da CDK11 na célula depende de uma etapa de enovelamento assistido da qual participam a chaperona eucariótica Hsp90 e sua co-chaperona Cdc37. A Cdc37 atua direcionando especificamente as proteínas quinases que reconhece para o complexo da Hsp90, o qual auxilia o polipeptídeo a alcançar sua conformação nativa num processo que envolve a hidrólise de ATP. Na ausência desta etapa de enovelamento assistido, a proteína sofre ubiquitinação e subsequente degradação proteolítica por uma via dependente de proteossomo (MIKOLAJCZYK et al., 2004). Foi mencionado anteriormente que a CDK9 também se utiliza do sistema Hsp90/Cdc37 para mediar o enovelamento correto da proteína recém-sintetizada e sua associação à subunidade ciclina (O'KEEFE et al., 2000). De modo semelhante, é reconhecido que a chaperona Hsp90/Cdc37 assiste no estabelecimento da conformação nativa das CDKs 4 e 6, mas parece não interagir com as CDKs 2, 3 e 5 (MIKOLAJCZYK et al., 2004).

Um obstáculo que compromete bastante o progresso do conhecimento científico sobre a CDK11 é a ausência de estudos dedicados à caracterização estrutural da proteína. Neste sentido, com a recente publicação das estruturas de outras CDKs do grupo – CDK7 (LOLLI et al., 2004), CDK9 (BAUMLI et al., 2008), CDK8 (SCHNEIDER et al., 2011) e CDK12 (BÖSKEN et al., 2014) – a CDK11 representa uma das poucas CDKs que ainda permanecem sem estrutura cristalográfica determinada. Igualmente, não se dispõe de nenhuma estrutura depositada no *Protein Data Bank* (PDB) para as ciclinas L até o presente momento.

A Tabela 2 é um sumário das mais importantes descobertas resultantes das pesquisas que vêm desvendando ao longo dos últimos 25 anos a natureza e função da CDK11.

Tabela 2 – Principais eventos no desenvolvimento da pesquisa sobre a CDK11. A Tabela relaciona as principais contribuições para o conhecimento científico a respeito da CDK11 desde sua descoberta.

Ano	Acontecimento
1990	*A descoberta de uma proteína humana de 58 kDa, cuja superexpressão afeta negativamente a progressão do ciclo celular em células CHO, representa o início das pesquisas sobre a quinase CDK11/PITSLRE (BUNNEL et al., 1990).
1994	*É reconhecido que a PITSLRE se origina de dois genes adjacentes e altamente relacionados no cromossomo 1 humano, capazes de produzir numerosas isoformas da proteína por <i>splicing</i> alternativo. (XIANG et al., 1994).
1995	*A continuação dos estudos de expressão ectópica em células CHO revela a associação entre a CDK11 e a apoptose (LAHTI et al., 1995).
2000	*Identificado o <i>IRES</i> no RNAm da CDK11, responsável pela expressão da isoforma de 58 kDa durante o estágio G ₂ /M do ciclo celular (CORNELIS et al., 2000).
2002	*A descoberta da ciclina L humana e sua interação com a PITSLRE leva a proteína a ser renomeada CDK11 (DICKINSON et al., 2002). *Paralelamente à identificação das ciclinas L, é reconhecida a interação da ciclina D3 com a isoforma CDK11 ^{p58} (ZHANG et al., 2002).
2003	*Demonstrada a fosforilação do fator eucariótico de iniciação eIF3f pela isoforma CDK11 ^{p46} , ativada por caspases durante a apoptose (SHI et al., 2003).
2004	*É reconhecida a interação da CDK11 com a chaperona Hsp90/Cdc37, e sua degradação por uma via dependente de proteossomo na ausência desta etapa de enovelamento assistido (MIKOLAJCZYK et al., 2004).
2005	*Identificado o papel do fragmento N-terminal da CDK11 processada por caspases (CDK11 ^{p60}) na apoptose, o qual depende da sua translocação para a mitocôndria (FENG et al., 2005).
2006	*Descoberta a função essencial da isoforma CDK11 ^{p58} na maturação do centrôssomo e nucleação dos microtúbulos para montagem do fuso durante a mitose (PETRETTI et al., 2006).
2011	*Um estudo revela a importância da autofosforilação no resíduo Thr370 e subsequente dimerização para a atividade quinásica e função da isoforma CDK11 ^{p58} (CHI et al., 2011).
2012	*O Mediador é reconhecido como um substrato da CDK11 ^{p110} , que promove a associação do módulo quinase ao cerne do complexo através da fosforilação das subunidades Med4 e Med27 (DROGAT et al., 2012).

1.6 Função molecular da CDK11

Um bom indicativo do papel exercido por uma proteína na célula pode ser obtido pelo estudo do fenótipo associado à perda da sua função. A inativação simultânea de ambos os alelos CDC2L em camundongos provoca a letalidade embrionária do indivíduo *knockout* no estágio de blastocisto (LI et al., 2004). Deleções da banda p36 do cromossomo 1 humano são um achado frequente em diversos tipos de câncer, sugerindo que o gene da CDK11 deve corresponder ao supressor tumoral há muito tempo postulado nesta região (XIANG et al., 1994). Por exemplo, foi constatado que um alelo do locus CDC2L encontrava-se deletado ou translocado em oito de quatorze diferentes linhagens de melanoma analisadas. A correlação entre a perda de função da CDK11 e o fenótipo maligno das células cancerosas provavelmente é resultado da grande resistência à apoptose que estas linhagens apresentam em relação às células normais, que contêm alelos CDC2L intactos (ARIZA et al., 1999).

A princípio foi sugerido que a CDK11 poderia ser mais uma CDK envolvida na fosforilação do CTD da RNAP II, devido à observação de que a CDK11^{p110} encontra-se majoritariamente associada aos complexos de transcrição da RNAP II no núcleo, e a introdução de anticorpos contra a CDK11 suprime de modo específico a síntese de transcritos *in vitro* (TREMBLEY et al., 2002). A hipótese de que a CDK11 poderia ser uma nova CTD quinase foi momentaneamente reforçada pela identificação de uma atividade quinásica capaz de fosforilar o CTD em amostras de CDK11 purificadas por imunoafinidade. Entretanto, logo foi esclarecido que a fosforilação do CTD era devida à presença da caseína quinase 2 (CK2), que se associa e é co-purificada com a CDK11^{p110}. De fato, este estudo acabou revelando que a interação entre as duas quinases ocorre porque a CK2 fosforila o domínio N-terminal da CDK11^{p110} (TREMBLEY et al., 2003). Embora a CK2 seja capaz de fosforilar o CTD da RNAP II *in vitro*, como muitas outras Ser/Thr-quinases conhecidas, não foram precisamente determinadas as circunstâncias e a extensão em que ela contribui para o estado de fosforilação da polimerase *in vivo*.

A função celular da CDK11 vem sendo elucidada gradualmente desde a sua descoberta, à medida que são identificados os substratos da quinase e os fatores que regulam sua atividade. Todavia, uma descrição geral do papel da CDK11 permanece difícil de

ser elaborada, pois as diferentes variedades da quinase têm sido implicadas em contextos funcionais bastante distintos que parecem não apresentar inter-relação direta (LOYER et al., 2005). Por este motivo as isoformas da CDK11 são tratadas de forma independente, cada uma delas tendo como alvo um conjunto particular de substratos envolvidos em processos celulares específicos.

CDK11^{p46} – A CDK11^{p46} corresponde ao fragmento C-terminal produzido pela clivagem das isoformas CDK11^{p58} e CDK11^{p110} por caspases, e é capaz de induzir a apoptose quando ectopicamente expressa em células humanas (MIKOLAJCZYK et al., 2003). A atividade pró-apoptótica foi a primeira função da CDK11 a ser reconhecida (LAHTI et al., 1995), tornando-a o primeiro substrato de caspases para o qual um papel direto na apoptose foi descrito (BEYAERT et al., 1997). A clivagem da CDK11 foi observada em diversas linhagens celulares durante a apoptose induzida pelo receptor Fas, fator de necrose tumoral (TNF), e estaurosporina; sendo a caspase-3 a principal responsável pela sua ativação proteolítica (TANG et al., 1998; BEYAERT et al., 1997; ARIZA et al., 1999). Diversos alvos já foram identificados para a CDK11^{p46}, incluindo a subunidade p47 (f) do fator de iniciação eucariótico eIF3 (SHI et al., 2003), a proteína RanBPM ligadora de Ran GTPase (MIKOLAJCZYK et al., 2004) e a proteína de função indeterminada NOT2 (SHI et al., 2005). O substrato da CDK11^{p46} cujo papel como efector na apoptose está mais bem caracterizado é o fator eIF3, que participa da iniciação da tradução pelo ribossomo eucariótico. A fosforilação dos resíduos Ser46 e Thr119 da subunidade eIF3f fortalece a sua interação com as subunidades no cerne do complexo eIF3, ocasionando um efeito inibitório sobre a síntese de proteínas e presumivelmente estimulando a degradação do RNA ribossomal (SHI et al., 2009). Igualmente importante para a função pró-apoptótica da CDK11 parece ser a geração do fragmento CDK11^{p60} pela atividade proteolítica das caspases. A CDK11^{p60} corresponde ao fragmento N-terminal – complementar à CDK11^{p46} – que é produzido quando as isoformas CDK11^{p110} são clivadas durante a apoptose. Uma vez gerada, a CDK11^{p60} é translocada para a mitocôndria através do motor de importação mitocondrial Hsp70, em um processo que induz o desprendimento do citocromo c (FENG et al., 2005). A liberação de citocromo c pela mitocôndria leva indiretamente à ativação da caspase-3, promovendo a retroalimentação da cascata apoptótica.

CDK11^{p58} – Ao contrário das isoformas maiores CDK11^{p110}, cuja expressão não é influenciada pelo ciclo celular, as isoformas CDK11^{p58} estão presentes exclusivamente durante as fases G₂ e M, e sua tradução é iniciada por um sítio interno de entrada do ribossomo (*IRES*) no RNAm (CORNELIS et al., 2000). Como seria de se esperar, todas as funções atribuídas à CDK11^{p58} são relacionadas a eventos celulares que fazem parte da mitose e citocinese. Em primeiro lugar, foi demonstrado que a atividade da CDK11^{p58} é essencial para a maturação dos centrossomos e formação do fuso mitótico nas células em divisão (PETRETTI et al., 2006). Além disso, a CDK11^{p58} funciona como reguladora da coesão de cromátides-irmãs, evitando a sua separação prematura antes do estágio de anáfase (HU et al., 2007). Devido a estes papéis importantes na regulação de eventos mitóticos, a depleção parcial da CDK11^{p58} resulta em incapacidade de duplicação dos centríolos, formação de fuso encurtado ou monopolar, aneuploidias cromossômicas e surgimento de células multinucleadas. O acúmulo de defeitos mitóticos nas células-filhas acaba por inviabilizar a ocorrência de divisões subsequentes, desencadeando em muitas delas a morte celular programada (HU et al., 2007; FRANCK et al., 2011). Em concordância com a função a ela atribuída, a CDK11^{p58} se concentra nos polos do fuso durante a mitose (PETRETTI et al., 2006), embora não tenha sido identificada como um componente dos centrossomos mitóticos purificados (FRANCK et al., 2011). A localização subcelular da CDK11^{p58} é resultado da sua associação direta aos microtúbulos, exercendo um importante papel como reguladora da nucleação de MTs no processo de formação do fuso (YOKOYAMA et al., 2008). Os estudos que desvendaram a função mitótica da CDK11^{p58} levantaram a suspeita de que as quinases do polo Plk1 e Plk4 poderiam representar seus alvos de fosforilação envolvidos na montagem do fuso, mas esta hipótese permanece sem confirmação (PETRETTI et al., 2006; FRANCK et al., 2011). De fato, um dos poucos substratos conhecidos da CDK11^{p58} é a quinase PAK1, cuja fosforilação em Ser174 promove seu recrutamento para o complexo motor da miosina V, onde deve atuar como efetora na progressão da mitose (KONG et al., 2009). Na verdade, foi primeiro identificada a fosforilação de PAK1 pela CDK11^{p46} (CHEN et al., 2003), contudo ficou esclarecido posteriormente que apenas a isoforma CDK11^{p58} pode ser responsável pela fosforilação de PAK1 dependente da CDK11 *in vivo* (KONG et al., 2009). Por último, foi descoberto que o resíduo Thr370 representa um sítio cuja auto-fosforilação é essencial à atividade quinásica da CDK11^{p58} e provoca a sua dimerização (CHI et al., 2011).

CDK11^{p110} – Enquanto existem apenas 2 variedades conhecidas da CDK11^{p58} – cada uma referente a um dos genes da CDK11 – a ocorrência de *splicing* alternativo e a existência de dois promotores distintos para cada gene resulta na expressão de uma multiplicidade de isoformas CDK11^{p110}. Embora referidas coletivamente como CDK11^{p110} em razão da sua massa aparente, estas proteínas possuem massa molecular predita entre 62 e 90 kDa, apresentando uma migração marcadamente anômala em géis SDS-PAGE (XIANG et al., 1994). O extenso domínio N-terminal exclusivo das isoformas CDK11^{p110} contém diversos elementos regulatórios, incluindo uma sequência de localização nuclear que condiz com sua função associada à transcrição pela RNAP II e *splicing* do pré-RNAm. Este papel na regulação da transcrição e *splicing* foi originalmente sugerido quando os primeiros estudos demonstraram a co-imunoprecipitação e co-purificação da CDK11^{p110} com o fator de *splicing* RNPS1 (LOYER et al., 1998) e os fatores de elongação ELL2, TFIIIF, TFIIIS e FACT a partir de extratos nucleares (TREMBLEY et al., 2002). Apesar da interação física com a quinase em complexos de transcrição, nenhuma destas proteínas parece funcionar como substrato para a CDK11. Entretanto, logo em seguida foi descoberta a associação da CDK11^{p110} a um outro fator geral de *splicing*, a proteína SR 9G8, a qual representa um dos seus principais alvos de fosforilação conhecidos (HU et al., 2003). Evidências diretas para a função transcricional da CDK11^{p110} começaram a ser obtidas quando foi demonstrado que a introdução de anticorpos contra o domínio quinase da CDK11 suprime especificamente a síntese de transcritos *in vitro* (TREMBLEY et al., 2002), e a imunodepleção da CDK11^{p110} em extratos nucleares reduz drasticamente a formação de produtos de *splicing* (HU et al., 2003). De maneira semelhante, a depleção das ciclinas L também se mostrou capaz de afetar negativamente o *splicing* do pré-RNAm em ensaios *in vitro* e *in vivo* (LOYER et al., 2008). Mais recentemente, foi atribuída à CDK11^{p110} a importante função de regular a associação do módulo quinase – que contém o heterodímero CDK8/ciclina C – ao cerne do complexo Mediador (S-Mediador). A interação do S-Mediador com o módulo quinase se revelou dependente da fosforilação das subunidades Med4 e Med27 pela CDK11^{p110} (DROGAT et al., 2012). Apesar da evidente relevância funcional, o papel exato deste processo na dinâmica de iniciação da transcrição pela RNAP II não é conhecido.

A Figura 12 ilustra a função celular da CDK11, resumindo as principais informações que são apresentadas nesta seção.

2 JUSTIFICATIVA

O grupo de CDKs reguladoras da transcrição tem atraído o interesse crescente de pesquisadores das mais diversas especialidades afins à Biologia Molecular. Estas CDKs desempenham papéis de controle na intrincada atividade de transcrição e processamento do RNA mensageiro pela RNA Polimerase II. Pesquisas envolvendo as CDKs da subfamília conduziram à elucidação de muitos detalhes dos mecanismos de funcionamento e regulação da RNAP II. Por exercerem uma série de funções relacionadas à transcrição, estas quinases assumem naturalmente grande importância como elementos reguladores da expressão gênica nas células eucarióticas. De fato, muitos estudos têm desvendado conexões entre o grupo de CDKs transcrpcionais e processos celulares característicos dos eucariotos superiores, como a diferenciação, ciclo celular, apoptose e desenvolvimento.

Talvez, porém, o grande interesse atual nas pesquisas sobre CDKs reguladoras da transcrição seja devido à associação constante de suas disfunções a uma variedade de patologias graves. A desregulação da atividade destas quinases tem sido frequentemente observada em células de diversos tipos de câncer humano. As próprias CDKs do grupo e as ciclinas que as ativam são codificadas por oncogenes e genes supressores de tumor, bem como regulam a expressão de numerosos genes nestas categorias. Além do câncer, as CDKs também provaram ter relações com uma série de outras doenças: Um exemplo notável é a AIDS, em razão do papel essencial da CDK9 para a replicação do HIV. Portanto, as CDKs reguladoras da transcrição estão emergindo como importantes alvos terapêuticos a serem explorados pela indústria no desenvolvimento de fármacos antineoplásicos e antivirais.

A obtenção de CDKs humanas a partir de sua expressão heteróloga por um sistema bacteriano oferece muitas vantagens práticas ao trabalho experimental. Os principais atrativos são a simplicidade da técnica, baixo custo, altos níveis de expressão e facilidade no isolamento do produto. Assim, as quinases poderiam ser obtidas em grandes quantidades para aplicações que demandam amostras com elevada pureza.

3 OBJETIVOS

O objetivo deste trabalho é fornecer precedentes experimentais para a obtenção das CDKs humanas 9 e 11 enzimaticamente ativas, através da expressão heteróloga em um sistema bacteriano (*E. coli*) seguida do seu isolamento e purificação por cromatografia de afinidade. Nosso projeto busca o aperfeiçoamento de métodos que permitam obter estas enzimas em quantidade e pureza adequadas para utilização em estudos estruturais. Assim, esperamos contribuir para a pesquisa sobre as relações entre estrutura e função biológica neste grupo de CDKs.

4 MATERIAIS E MÉTODOS

Para obtenção das proteínas quinases CDK9 e CDK11 humanas em *E. coli*, o material de partida foram plasmídeos recombinantes contendo as sequências de cDNA que codificam as CDKs já inseridas nos vetores de expressão pET-23a(+) e pET-28a(+) (Novagen). A primeira parte desta seção (item 4.1) trata dos métodos relacionados à tecnologia do DNA recombinante e outros empregados na manipulação do clone original da CDK11. A segunda parte (item 4.2) trata dos métodos para obtenção das CDKs 9 e 11, através da sua expressão no sistema bacteriano e subsequente purificação cromatográfica.

4.1 Modificação do plasmídeo pET28a::cdk11

No caso da CDK11, as dificuldades de expressão observadas a princípio exigiram que o clone original fosse modificado para gerar uma versão encurtada do polipeptídeo.

4.1.1 Amplificação do pET28a::cdk11(Δ) por PCR

A amplificação de DNA foi utilizada como estratégia para simplificar a deleção de um extenso segmento (1.284 pb) da sequência codificante da CDK11 originalmente inserida no plasmídeo pET28a::cdk11. Para produzir este efeito, foram desenhados oligonucleotídeos iniciadores (*primers*) complementares a duas regiões no interior da sequência codificante da CDK11, porém voltados para a amplificação em direção às suas extremidades. O segmento contido entre os sítios complementares aos *primers* utilizados é assim excluído da amplificação, enquanto as extremidades remanescentes da sequência codificante permanecem ligadas ao vetor pET28a(+). Portanto, o fragmento a ser amplificado corresponde à maior parte do plasmídeo pET28a::cdk11 original, excetuando-se apenas um trecho interno na sequência codificante da CDK. As sequências de DNA da CDK11 e dos *primers* utilizados são apresentadas na Figura 15, na seção de Resultados e Discussão.

As reações de amplificação empregaram uma mistura comercial das DNA polimerases *Taq* e *Pfu* – *High Fidelity PCR Enzyme Mix* (Thermo Scientific). Um volume de 50 μ L foi preparado para cada reação, contendo ao final 1,0 μ mol/L de cada *primer* (*forward* e *reverse*), 0,2 mmol/L de cada desoxirribonucleosídeo 5'-trifosfato (dNTP), 2,5 mmol/L de

MgCl₂ e 0,84 ng do molde de DNA – o plasmídeo pET28a::cdk11. Para prevenir a degradação dos oligonucleotídeos, as amostras foram preparadas em uma cabine de fluxo laminar utilizando água *nuclease-free*. A amplificação foi realizada por um termociclador *MyCycler Thermal Cycler* (BioRad) com o volume total (50 µL) dividido em 5 alíquotas de 10 µL. Inicialmente, cada uma das 5 alíquotas idênticas foi submetida a uma temperatura diferente durante a etapa de hibridização (no intervalo de 51 a 58 °C), obtendo-se o melhor resultado na temperatura de 53,5 °C. O processo foi então repetido colocando-se o volume total de 50 µL sob esta temperatura para a hibridização dos *primers*. Como o tamanho do fragmento de DNA a ser amplificado (> 6,3 kb) é relativamente grande em relação à capacidade das polimerases, a programação do termociclador teve de ser adaptada para incrementar o rendimento do produto de amplificação. Assim, foram programados 35 ciclos térmicos com a temperatura de alongação (68 °C) diminuída em relação ao seu valor usual (72 °C), e tempo extra foi gasto nesta etapa – 6 minutos a mais para alongação em cada ciclo (além do 1 min/kb empregado normalmente). Os resultados das amplificações foram visualizados por eletroforese em gel de agarose, para revelar o rendimento e a pureza do produto obtido (amplicon).

4.1.2 Eletroforese em gel de agarose

A eletroforese em gel de agarose foi realizada em diversas ocasiões para analisar o conteúdo das amostras de DNA. Todas as aplicações utilizaram géis 0,8% de agarose em tampão TAE (Tris-acetato 40 mmol/L; EDTA 1,0 mmol/L; pH 8,0). As amostras de DNA foram preparadas para aplicação pela mistura na razão 5:1 com o tampão de carregamento 6x (*loading buffer*) (Tris-HCl 10 mmol/L; azul de bromofenol 0,03%; xileno cianol 0,03%; glicerol 60%; EDTA 60 mM; pH 7,6). Como referência para o tamanho dos fragmentos de DNA, cada gel teve uma raia destinada à aplicação do marcador *GeneRuler™ 1 kb DNA Ladder* (Fermentas). Uma diferença de potencial entre 90 e 110 V foi aplicada a cada gel na cuba de eletroforese, preenchida com tampão TAE, por um período de 15 a 25 minutos. Após a corrida, os géis foram imersos em uma solução 0,5 µg/mL de brometo de etídio por 15 a 20 minutos para permitir a visualização das bandas por fluorescência. Imagens dos géis foram capturadas utilizando um fotodocumentador *MiniBIS pro* (DNR Bio-Imaging Systems) com lâmpada de ultravioleta.

Além da sua extensa aplicação na análise das amostras de DNA, a eletroforese em gel de agarose também foi empregada em uma etapa do trabalho como método de separação, para purificar o fragmento amplificado por *PCR*.

4.1.3 Ciclização do DNA plasmidial amplificado

Após a reação de amplificação, todo o volume resultante do *PCR* foi aplicado em um único gel de agarose, para visualização e separação do fragmento de interesse. O gel foi tratado com a solução de brometo de etídio para permitir a visualização das bandas de DNA sob luz UV. A região do gel correspondente à banda de interesse foi recortada com um estilete e sua amostra de DNA recuperada por extração e purificação utilizando o *QIAquick Gel Extraction kit* (QIAGEN) conforme o protocolo do fabricante. Ao final obtivemos uma amostra de 50 µL contendo o fragmento purificado em uma concentração de 80,5 ng/µL, quantificada pela absorbância a 260 nm em um espectrofotômetro *NanoDrop™ 1000* (Thermo Scientific). Assim, o rendimento total foi cerca de 4,00 µg de DNA para o processo completo de recuperação a partir do gel.

O fragmento isolado consiste em uma cadeia linear de DNA em dupla fita com 6.391 pb, o que representa uma fração dos 7.657 pb do plasmídeo pET28a::cdk11 original. No processo de amplificação, o par de *primers* utilizado introduziu um sítio de clivagem pela endonuclease de restrição *NheI* em cada extremidade do fragmento de interesse. Esta sequência de reconhecimento corresponde ao palíndromo de 6 pares de bases (5')GCTAGC, e permite que sejam geradas extremidades complementares nas duas pontas do fragmento através de sua digestão pela enzima *NheI*. Assim, o fragmento linear pode ser ciclizado por uma recombinação intramolecular ligando as suas extremidades. O produto – que designamos pET28a::cdk11(Δ) – possui o sítio de reconhecimento da *NheI* substituindo um extenso trecho (1284 pb) deletado no interior da sequência codificante da CDK11.

O produto de amplificação purificado foi submetido a reações consecutivas de clivagem e religação em um mesmo meio, sem o isolamento do intermediário linear contendo as extremidades digeridas. Esta estratégia diminui o número de etapas e evita as perdas de amostra inerentes a uma etapa extra de purificação. As reações consecutivas foram realizadas em um tampão compatível com ambas as enzimas empregadas (a

endonuclease de restrição *NheI* e a DNA ligase). Após a reação de clivagem, efetuamos a inativação térmica da endonuclease de restrição antes de prosseguir à etapa de ligação do DNA. O produto de clivagem/ligação foi então utilizado diretamente para transformar uma cepa bacteriana de clonagem (DH5 α) e assim obter o plasmídeo recombinante pET28a::cdk11(Δ) purificado.

Para realizar a síntese do novo plasmídeo, empregamos 45 dos 50 μ L do produto de amplificação do *PCR*, portanto cerca de 3,60 μ g de DNA. Os 45 μ L de amostra foram primeiramente misturados a 5,0 μ L do tampão *NEBuffer 4* (New England Biolabs) concentrado 10x (concentrações finais: acetato de potássio 50 mmol/L; tris-acetato 20 mmol/L; acetato de magnésio 10 mmol/L; DTT 1,0 mmol/L; pH 7,9). Esta mistura recebeu 1,0 μ L da enzima *NheI* (New England Biolabs) e foi incubada na estufa a 37 °C durante 1 hora e meia para digestão do DNA. Após este tempo de reação, a mistura foi retirada da estufa e submetida a um banho térmico de 20 minutos a 80 °C, o suficiente para eliminar toda a atividade residual da endonuclease de restrição. Em seguida, a amostra foi resfriada no gelo por alguns minutos antes de receber 2,5 μ L da T4 DNA ligase (New England Biolabs) e ATP até uma concentração final de 2,0 mmol/L. A reação de ligação foi então realizada com a amostra imersa em um banho térmico a 25 °C durante 20 minutos.

Estimando-se a massa molecular do novo plasmídeo em cerca de 3.940 kDa (calculado com o software disponível online em *bioinformatics.org*), a concentração de extremidades complementares geradas no meio após a reação de clivagem deve ser inferior a 0,020 μ mol/L. Este valor é cerca de 5 vezes menor que a concentração mínima recomendada para a reação de ligação pelo fabricante da T4 DNA ligase. Entretanto, devemos considerar que pretendemos unir as duas extremidades de um mesmo fragmento, e não de dois fragmentos distintos como é usual. Sabemos que as reações intramoleculares são cineticamente favorecidas, em relação às reações intermoleculares paralelas, pela diluição dos reagentes. Assim, esperamos que a concentração mais baixa do fragmento linear aumente especificamente o rendimento da recombinação intramolecular, na qual estamos interessados.

O produto final da recombinação foi utilizado para transformar a cepa de clonagem DH5 α . Os procedimentos de preparação e transformação de células competentes são idênticos àqueles descritos a seguir para as cepas de expressão. Seis colônias crescidas na placa foram escolhidas para inocular culturas em meio líquido, as quais utilizamos no isolamento do pET28a::cdk11(Δ) empregando o kit de extração de DNA plasmidial *GeneJet™ Plasmid Miniprep* (Fermentas). O sucesso da recombinação foi confirmado por meio do sequenciamento de três destas amostras do plasmídeo purificado.

4.1.4 Sequenciamento do DNA

As amostras do novo plasmídeo foram enviadas para sequenciamento do DNA no Laboratório de Biologia Molecular do Grupo de Cristalografia do Instituto de Física de São Carlos – IFSC/USP. O sequenciamento pelo *3130 Genetic Analyzer* (Applied Biosystems) emprega o método de Sanger automatizado, no qual os terminadores de cadeia (ddNTPs) possuem marcadores fluorescentes que permitem a sua identificação em uma única raia. Os resultados do sequenciamento foram alinhados com a sequência original da CDK11 e aquela esperada após a deleção. Em todos os casos foi possível observar a integridade do sítio de recombinação e ausência de mutações em toda a sequência codificante da CDK11(Δ).

4.2 Expressão das proteínas e purificação da CDK9

Uma vez obtido o plasmídeo pET28a::cdk11(Δ), seguimos para os testes de expressão e finalmente alcançamos um resultado positivo para a CDK11. Por outro lado, a expressão da CDK9 a partir dos plasmídeos pET28a::cdk9 e pET23a::cdk9 foi prontamente realizada pela bactéria utilizando a cepa BL21(DE3).

4.2.1 Transformação das cepas de expressão

As amostras de células competentes para transformação foram preparadas pelo método do cloreto de cálcio, no qual são congeladas em uma suspensão contendo CaCl₂ 0,1 mol/L e glicerol 10% e mantidas a -80 °C até a sua utilização. Cada experimento consumiu uma quantidade de 10 a 100 ng de DNA plasmidial para transformar alíquotas de 100 a 200 μ L da suspensão de células competentes recém-descongeladas. A mistura homogeneizada

foi inicialmente mantida no gelo por 30 minutos, sofrendo na sequência um choque térmico de 42 °C durante 2 minutos. O choque térmico é o agente responsável pela transformação bacteriana, facilitando a passagem do DNA exógeno através da parede celular e membrana plasmática da célula. Após o choque, as amostras receberam 500 µL de meio LB líquido e foram incubadas em um *shaker* a 37 °C/200 rpm por 1 hora. As culturas crescidas no *shaker* foram aplicadas em placas de Petri com 20 mL de meio sólido LB-ágar contendo o antibiótico apropriado para selecionar as células transformadas (canamicina 30 µg/mL para o pET28a ou ampicilina 25 µg/mL para o pET23a). As placas foram então incubadas em estufa a 37°C *overnight* e examinadas na manhã seguinte para observar o crescimento de colônias. A Tabela 3, na página seguinte, reúne as cepas de *E. coli* que foram empregadas neste trabalho para clonagem e expressão das CDKs.

4.2.2 Experimentos de indução

A expressão das CDKs foi induzida com IPTG, uma vez que as cepas bacterianas derivadas da BL21(DE3) possuem o gene da T7 RNA polimerase sob controle do operador *lac*. As cepas transformadas são utilizadas para crescer uma cultura em meio *Lysogen Broth* (LB) líquido, composto por 10 g/L de triptona, 5 g/L de extrato de levedura e 10 g/L de NaCl em água destilada. Antes do experimento de indução, as cepas transformadas são inoculadas em um pequeno volume (pré-inóculo) do meio LB líquido e crescidas a 30 °C *overnight*. A cultura obtida no pré-inóculo é então transferida para os frascos com maior capacidade – 100 a 500 mL – nos quais a indução é realizada. Os frascos são mantidos no agitador orbital (*shaker*) a 37 °C enquanto o crescimento da cultura é monitorado constantemente pela densidade óptica da amostra a 600 nm. Quando a densidade óptica atinge um valor entre 0,4 e 0,6 unidades os frascos recebem o IPTG e a temperatura do *shaker* é ajustada para o valor escolhido para a indução.

Inicialmente, os experimentos de indução foram realizados no próprio meio LB. Devido à necessidade de se aprimorar a indução, ele passou a ser substituído por um meio de composição semelhante ao SOB (*Super Optimal Broth*), mais conveniente para a superexpressão de proteínas: 20 g/L de triptona; 10 g/L de extrato de levedura; 5,0 g/L de NaCl; 4,0 g/L de KCl; 3,0 g/L de MgCl₂; 1,0% de glicerol. Além de possuir o dobro do conteúdo de triptona do meio LB, o meio de indução alternativo também inclui 15 mmol/L de MgCl₂,

Tabela 3 – Cepas bacterianas de clonagem e expressão utilizadas. A tabela lista as cepas de *Escherichia coli* empregadas neste trabalho para clonagem dos plasmídeos e expressão das CDKs no sistema bacteriano. Para cada cepa é apresentado o respectivo genótipo resumido e as características principais que determinam a sua aplicação. A única cepa de clonagem utilizada foi DH5 α , que também representa a única derivada da linhagem K-12, todas as demais correspondem a cepas de expressão derivadas da BL21.

Cepa	Genótipo	Características Principais
DH5α	F ⁻ <i>deoR endA1 gyrA96 hsdR17</i> $\Delta(lac)U169$ <i>recA1 relA1 supE44 thi-1</i> ($\phi 80 lacZ\Delta M15$)	Apresenta grande eficiência na transformação e mutações que favorecem a estabilidade de plasmídeos incorporados pela célula.
BL21(DE3)	F ⁻ <i>dcm ompT lon hsdS_B gal</i> λ (DE3)	Possui o gene da T7 RNA polimerase (sob controle do operador <i>lac</i>), o que permite a expressão de proteínas a partir de vetores que utilizam o promotor T7.
BL21(DE3) CodonPlus-RIL	F ⁻ <i>dcm+ ompT lon hsdS_B gal endA Hte Tet^r λ(DE3)</i> [pACYC-RIL <i>argU ileY leuW Cam^r</i>]	Contém o plasmídeo pACYC-RIL, que possui cópias extras dos genes de RNAs que traduzem os códons AGG/AGA (Arg), AUA (Ile) e CUA (Leu).
BL21(DE3) CodonPlus-RP	F ⁻ <i>dcm+ ompT lon hsdS_B gal endA Hte Tet^r λ(DE3)</i> [pACYC-RP <i>argU proL Cam^r</i>]	Contém o plasmídeo pACYC-RP, que possui cópias extras dos genes de RNAs que traduzem os códons AGG/AGA (Arg) e CCC (Pro).
Rosetta (DE3)	F ⁻ <i>dcm ompT hsdS_B gal lacY1</i> λ (DE3) [pRARE <i>araW argU glyT ileX leuW proL metT thrT tyrU thrU Cam^r</i>]	Contém o plasmídeo pRARE, que possui cópias extras dos genes de RNAs que traduzem os códons AGG/AGA (Arg), CGG (Arg), AUA (Ile), CUA (Leu), CCC (Pro) e GGA (Gly).
BL21(DE3) pLysS	F ⁻ <i>dcm ompT lon hsdS_B gal</i> λ (DE3) [pLysS Cam ^r]	Contém o plasmídeo pLysS com o gene da T7 lisozima, um inibidor natural da T7 RNA polimerase que diminui a expressão basal de RNAm a partir de promotores T7.
BL21(DE3) pLysE	F ⁻ <i>dcm ompT lon hsdS_B gal</i> λ (DE3) [pLysE Cam ^r]	Contém o plasmídeo pLysE com o gene da T7 lisozima, alcançando níveis mais altos de expressão da enzima viral que a pLysS.
BL21(DE3) pT-GroE	F ⁻ <i>dcm ompT lon hsdS_B gal</i> λ (DE3) [pT-GroE Cam ^r]	Contém o plasmídeo pT-GroE com o operon <i>groE</i> , que codifica as chaperonas bacterianas GroEL e GroES.

Fonte: Adaptado de Casali et al. (2003, p. 42).

visto que o cátion magnésio é um co-fator essencial para a atividade enzimática da CDK9. O meio também passou a ser tamponado com 50 mmol/L de tris, para evitar a queda no pH inerente ao processo de indução. As condições de indução variaram bastante nos experimentos, empregando-se concentrações de IPTG entre 0 e 1,0 mmol/L e temperaturas entre 20 e 37 °C. Após um período de 3 a 5 horas de indução, os frascos foram retirados do *shaker* e centrifugados a 8.000 x g por 30 min a 4 °C para recuperar o sedimento bacteriano. As amostras da cultura induzida foram então ressuspensas no tampão de lise e armazenadas a -20 °C até a sua utilização.

4.2.3 Lise das células

Para cada grama de sedimento bacteriano da cultura induzida foram utilizados 5,0 mL do tampão de lise. Amostras destinadas à purificação pela coluna de afinidade ao níquel foram lisadas em diversos tampões para determinar as condições mais favoráveis de ligação à resina. O melhor resultado foi obtido com um tampão fosfato desnaturante em pH 7,5 (100 mmol/L de Na₂HPO₄; 600 mmol/L de NaCl; 10 mmol/L de imidazol; 10% glicerol; 8,0 mol/L de ureia). Paralelamente, as tentativas de purificação em condições nativas empregaram tampão Hepes em vez de fosfato, devido à incompatibilidade com o cátion Mg²⁺, que é essencial à atividade da CDK9. Neste caso, após serem ressuspensas no tampão apropriado, as amostras receberam as enzimas lisozima (0,1 mg/mL) e DNase (0,5 u), além de 10 µL/ml do coquetel de inibidores de proteases do *kit CellLytic*TM B Plus (Sigma), permanecendo em repouso no gelo por 30 minutos antes da sonicação.

As amostras foram sonicadas no equipamento *Branson Sonifier S250A* (Soni-Tech Ultrasonic Systems) em 10 ciclos de 10 pulsos com intervalos de 50 segundos para dissipação do calor. Ao final, as amostras foram centrifugadas a 15.000 x g por 30 min a 4 °C para separar a fração insolúvel (*pellet*) do sobrenadante. O sobrenadante foi reservado para purificação da proteína, enquanto o *pellet* foi ressuspendido no próprio tampão de lise, em um volume equivalente ao do sobrenadante, para comparação das amostras em SDS-PAGE.

4.2.4 Eletroforese em gel de poliacrilamida (SDS-PAGE)

A eletroforese em gel de poliacrilamida foi realizada para avaliar o conteúdo de proteínas das amostras provenientes dos testes de indução. Os géis foram preparados com grau de entrecruzamento de 15%, empregando-se tampão de amostra desnaturante (Tris-HCl 0,188 mol/L; pH 6,8 contendo SDS 6%; EDTA 6 mmol/L; glicerol 30%, azul de bromofenol 0,01% e β -mercaptoetanol 10%). As amostras foram misturadas com o tampão na proporção volumétrica 1 (amostra) : 1 (tampão) e aquecidas por 5 minutos a 95 °C antes da aplicação de 10 μ l em cada raia do gel. A eletroforese foi realizada segundo o método de Laemmli (1970), com voltagem inicial de 120 V até entrada no gel de resolução e 200 V a partir de então até o final da corrida.

Como referência, os géis tiveram uma raia reservada para aplicação do marcador de massa molecular *Unstained Protein Molecular Weight Marker* (Fermentas), contendo bandas entre 14,4 e 116,0 kDa. Após a eletroforese, os géis foram impregnados com o corante *Coomassie Brilliant Blue R* (BioRad) e então descorados por aquecimento em água destilada para visualização das bandas protéicas. Imagens de cada gel foram obtidas em um *scanner* comum (Hewlett-Packard) e tratadas com um software de edição (Microsoft Office Picture Manager) para realçar a nitidez das bandas.

4.2.5 Western Blot

Para confirmar a presença da CDK9 em algumas amostras, foi realizado um *Western blot* utilizando um anticorpo primário que reconhece a cauda de histidina. As amostras de interesse foram aplicadas em SDS-PAGE, e então transferidas do gel para uma membrana de nitrocelulose. A transferência, por um método semi-seco, aplicou uma diferença de potencial de 25 V por 30 minutos em um equipamento Trans-Blot Turbo (BioRad). Após a transferência, a membrana foi tingida com Ponceau para demarcação das posições das bandas de interesse com uma agulha. A membrana foi então descorada com solução de ácido acético a 1,0% e incubada *overnight* a 4 °C com a solução de bloqueio, contendo 5% de leite em pó desnatado. No dia seguinte, a membrana foi lavada com tampão TBS e incubada por aproximadamente 2 horas com uma mistura dos anticorpos primário e secundário. Por fim, a membrana foi tratada com a solução reveladora contendo BCIP[®]/NBT (Sigma).

4.2.6 Purificação da CDK9 por cromatografia *IMAC*

O plasmídeo pET28a::cdk9 codifica a proteína de interesse fusionada a uma cauda de poli-histidina (His₆) N-terminal, ao contrário do pET23a::cdk9. A His-tag permite que seja empregada uma estratégia de purificação da proteína por afinidade a uma resina *IMAC* (*immobilized metal affinity chromatography*). A fração sobrenadante do lisado bacteriano foi incubada com 2,0 mL da resina *Profinity™* (BioRad) carregada com os cátions Ni²⁺ e pré-equilibrada no próprio tampão de lise. Após um período de incubação de até 4 horas sob agitação, a resina foi transferida para uma coluna afixada em suporte para a eluição das amostras. Inicialmente, foram aplicadas três alíquotas de 2,0 mL do próprio tampão de lise para lavar a coluna, removendo os componentes que não se ligam à resina. Em seguida, foi realizada a eluição do material retido pela resina, através de quatro etapas sucessivas de lavagem (2,0 mL cada) utilizando concentrações crescentes de imidazol no tampão (50, 100, 250 e 500 mmol/L). As frações eluídas em cada etapa foram coletadas separadamente para análise do seu conteúdo proteico.

4.2.7 Reenovelamento *in vitro* da CDK9

A purificação da His-tag-CDK9 só pôde ser realizada em condições desnaturantes (8,0 mol/L de ureia), devido à ausência da proteína na fração sobrenadante. Assim, era muito importante examinar a possibilidade de a proteína desnaturada recuperar parcialmente a sua conformação nativa e atividade enzimática após a purificação. Para tanto, primeiro foi realizada a lise em condições nativas e a separação dos corpos de inclusão pela centrifugação do lisado a 10.000 × g por 30 min a 4 °C. Em seguida, o material sedimentado foi ressuscitado em tampão desnaturante contendo 30 mmol/L de Hepes; 600 mmol/L de NaCl; 15 mmol/L de imidazol; 8,0 mol/L de ureia; 10% glicerol; pH 7,4. Após a amostra ser novamente sonicada e centrifugada a 15.000 × g por 30 min a 4 °C, a fração sobrenadante foi separada e recebeu SDS para uma concentração final de 0,5% antes de ser incubada com a resina *IMAC* previamente equilibrada com o mesmo tampão.

Ao final da incubação (que durou 4 horas sob agitação à temperatura ambiente), a resina foi transferida para uma coluna afixada em suporte. Após a eluição do *flowthrough*, foi aplicado um gradiente negativo de 8,0 a 0 mol/L de ureia em um volume total de 10 mL

(4,0; 2,0; 1,0; 0,5 e 0 mol/L em alíquotas de 2,0 mL cada) para remoção do desnaturante. Em seguida, foi aplicado um gradiente positivo de 15 a 500 mmol/L de imidazol em um volume total de 10 mL (25, 50, 100, 250 e 500 mmol/L em alíquotas de 2,0 mL cada) para promover a eluição da proteína ligada à resina. Apesar da remoção da ureia, a proteína continua desnaturada e mantém a sua solubilidade devido à presença do SDS 0,5%. Finalmente, as frações eluídas com imidazol da resina de níquel foram misturadas e receberam $MgCl_2$ para uma concentração final de 10 mmol/L e MPD (2-metil-2,4-pentanodiol) para uma concentração final de 10%. A amostra foi deixada em repouso *overnight* e diafiltrada no dia seguinte em um concentrador *Amicon Ultra-4* (Millipore), que contém um filtro molecular com limite de 30 kDa, até a composição do tampão atingir 15 mmol/L de Hepes; 100 mmol/L de NaCl; 10 mmol/L de $MgCl_2$; 10% glicerol; pH 7,4. A amostra resultante deste processo foi então submetida ao teste de atividade.

4.2.8 Ensaio de atividade enzimática

A atividade enzimática da proteína purificada é testada pela capacidade de ligação da quinase a uma resina cromatográfica de afinidade ao ATP. Esta resina de agarose entrecruzada contém unidades de ATP ligadas covalentemente à matriz, através do C8 no núcleo purínico. Proteínas que se ligam ao ATP são especificamente adsorvidas na resina e podem ser eluídas por um tampão contendo ATP ou elevada força iônica. A amostra purificada de CDK9 no tampão final de diálise foi incubada com 2,0 mL da resina A2767 (Sigma) pré-equilibrada com o mesmo tampão (15 mmol/L de Hepes; 100 mmol/L de NaCl; 10 mmol/L de $MgCl_2$; 10% glicerol; pH 7,4). Após 24 horas de incubação na geladeira (4 °C), a resina foi transferida para uma coluna e lavada com três volumes de 2,0 mL do tampão. Por fim, foi aplicado 1,0 mL de uma solução 50 mmol/L de ATP, que ficou incubada por 2 horas à temperatura ambiente com a resina antes de ser eluída.

Além de testar a atividade enzimática pela ligação da quinase à resina, este experimento cromatográfico permite ainda a separação da fração ativa da enzima em relação à fração incapaz de interagir com o ATP. O resultado é visualizado pela aplicação das amostras coletadas em gel SDS-PAGE e a quantificação direta do produto permite determinar o rendimento final.

4.2.9 Quantificação de proteínas totais

Os ensaios para quantificação de proteínas totais foram realizados para avaliar o conteúdo proteico dos produtos amostrados em cada etapa e determinar o rendimento na purificação da proteína. Diferentes métodos de quantificação foram utilizados em razão do amplo intervalo de concentrações de proteínas totais, desde o extrato celular bruto de *E. coli* até as amostras diluídas coletadas ao final das separações cromatográficas. As amostras do extrato celular bruto (fração sobrenadante após a lise bacteriana) foram quantificadas pelo método do biureto, devido à elevada concentração de proteínas. As amostras referentes ao produto da separação cromatográfica (IMAC) foram quantificadas pelo método de Bradford, utilizando o reagente *Protein Assay* (BioRad), devido à menor concentração de proteínas. Estes ensaios colorimétricos foram realizados em um espectrofotômetro BioMate™ (Thermo Scientific), que constrói curvas de calibração e calcula a regressão linear de maneira automatizada para determinar a concentração de amostras desconhecidas.

Por fim, as amostras da CDK9 purificada foram quantificadas pelo método de Edelhoch (leitura direta da absorbância a 280 nm), dada a elevada pureza da proteína e diluição, em um espectrofotômetro de micro-volume DS-11 (DeNovix). Entretanto, nos referimos às quantificações pela absorção das amostras a 280 nm como sendo de proteínas totais, uma vez que a determinação automática da concentração pelo instrumento não utilizou o coeficiente de absorvidade específico da CDK9.

5 RESULTADOS E DISCUSSÃO

5.1 Construção do plasmídeo pET28a::cdk11(Δ)

As primeiras tentativas de obtenção da CDK11 humana por meio do sistema bacteriano (*E. coli*) empregaram o plasmídeo pET28a::cdk11, produzido sob encomenda pela companhia *Epoch Life Science*. O plasmídeo contém a sequência codificante de uma isoforma da CDK11 humana - GI: 148763347; 2.343 pb - inserida entre *Nde*I e *Sal*I no sítio múltiplo de clonagem do pET-28a(+) (Novagen). O mapa do vetor pET-28a(+) está representado na Figura 13.

Figura 13 – Mapa do vetor bacteriano de expressão pET-28a(+) (Novagen). O mapa ilustra a organização do pET-28a(+), destacando os genes próprios do vetor, sua origem de replicação e o sítio múltiplo de clonagem (*polylinker*). O plasmídeo pET28a::cdk11 (com 7.657 pb totais), contém uma sequência de 2.343 pb inserida entre os sítios *Nde*I e *Sal*I do vetor, a qual codifica uma isoforma da CDK11 humana com 780 resíduos de aminoácidos e massa molecular aproximada de 91,0 kDa.

Fonte: Catálogo Novagen pET-28a-c(+) vectors.

O plasmídeo pET28a::cdk11 foi utilizado para transformar células de *E. coli* pertencentes à cepa de expressão BL21(DE3) para proceder aos experimentos de indução. Entretanto, após uma série de tentativas de indução em condições variadas, ficou evidente a dificuldade de se obter a expressão da proteína no sistema bacteriano. O resultado negativo contrasta com a relativa facilidade observada para a expressão da CDK9 humana no mesmo sistema. A sequência codificante da CDK11 originalmente introduzida no pET-28a(+) corresponde a uma isoforma com 780 resíduos de aminoácidos e cerca de 91 kDa. Esta isoforma contém uma extensa região N-terminal que não inclui o domínio responsável pela sua atividade de proteína quinase. Ao contrário, a isoforma de 42 kDa da CDK9 possui apenas 372 resíduos de aminoácidos, o que representa pouco mais que a estrutura mínima de uma proteína quinase (aproximadamente 300 resíduos). Assim, devido à similaridade entre as sequências das duas CDKs, raciocinamos que a expressão da CDK11 em *E. coli* poderia ser facilitada por uma extensa deleção N-terminal, com o objetivo de reduzi-la basicamente ao seu domínio de proteína quinase.

5.1.1 Estratégia para a deleção interna na sequência da CDK11

Existem várias abordagens experimentais possíveis para se gerar a deleção pretendida. Inicialmente, analisamos os sítios de restrição presentes na sequência codificante da CDK11 e no próprio vetor pET-28a(+). Caso fossem identificados dois sítios de restrição (equivalentes e únicos) delimitando o segmento N-terminal da CDK11, a alternativa mais simples seria digerir o plasmídeo recombinante com a enzima apropriada para então religar as extremidades do fragmento de interesse, excluindo assim o trecho a ser deletado. Infelizmente, porém, não foi identificada nenhuma enzima comercialmente disponível que satisfizesse as condições necessárias para esta operação. Decidimos então utilizar uma abordagem mais elaborada - que envolve a amplificação de DNA por PCR - porém simplificando-a pela minimização do número de etapas do processo.

A Figura 14 ilustra a estratégia que empregamos para produzir a deleção na sequência de DNA. O diagrama mostra que o resultado final do procedimento é a substituição de um trecho contendo 1.284 pb na sequência codificante da CDK11 por um sítio de restrição *NheI*, com apenas 6 pb. A proteína que esta nova sequência codifica será designada CDK11(Δ) por corresponder basicamente a uma versão truncada da CDK11.

Figura 14 – Etapas para obtenção do pET28a::cdk11(Δ). O esquema apresenta as etapas do processo de modificação da sequência codificante da CDK11 inserida no vetor pET-28a(+). Primeiramente, foram desenhados oligonucleotídeos iniciadores (*primers*) complementares a duas regiões no interior da sequência codificante da CDK11, porém voltados para a amplificação em direção às suas extremidades. Os *primers* delimitam o segmento a ser excluído e também introduzem um sítio de restrição para *NheI* em cada extremidade do fragmento amplificado por PCR. O produto da reação de amplificação foi purificado e digerido com a enzima *NheI* para gerar extremidades coesivas nos sítios de restrição. Por fim, a recombinação destas extremidades pela T4 DNA ligase produziu o novo plasmídeo pET28a::cdk11(Δ) com 6.379 pb.

A enzima *NheI* foi escolhida para o sítio de restrição introduzido pelos *primers* porque sua sequência de reconhecimento – (5')GCTAGC – codifica o dipeptídeo Ala-Ser. Por conter dois aminoácidos com cadeias laterais pequenas e eletricamente neutras, o dipeptídeo acrescentado à CDK11 deve minimizar a interferência na estrutura original da proteína. A deleção líquida de 1.278 pb na sequência codificante faz com que a cadeia polipeptídica seja encurtada de 780 para 354 resíduos de aminoácidos, o que corresponde a uma diminuição de 91,0 para 40,1 kDa na sua massa molecular. Considerando ainda a cauda de histidina N-terminal adicionada pelo vetor (com um total de 20 resíduos incluindo His₆), a massa

molecular da proteína His-tag-CDK11(Δ) é estimada em 42,3 kDa pela ferramenta *ProtParam* (Expasy). As Figuras 15 e 16 ilustram o resultado pretendido para a modificação das sequências de nucleotídeos e aminoácidos da CDK11.

Figura 15 – Sequência nucleotídica da CDK11 e primers utilizados. (a) Sequência codificante da CDK11 humana originalmente introduzida no vetor pET-28a(+). A sequência está dividida em seus 781 códons, com os códons de iniciação (ATG) e terminação (TGA) destacados em amarelo. Os trechos em azul representam as regiões de hibridização dos primers e o trecho cinza corresponde ao segmento deletado. (b) Sequências dos primers utilizados na amplificação (*Integrated DNA Technologies - IDT*). Os primers possuem uma sequência de segurança (em verde) que evita o posicionamento dos sítios de restrição (em vermelho) nas extremidades do amplicon, onde sua clivagem seria menos eficaz. As sequências de hibridização (em azul) são complementares aos trechos do DNA molde.

(a)

ATG	GGT	GAT	GAA	AAG	GAC	TCT	TGG	AAA	GTG	AAA	ACT	TTA	GAT	GAA	ATT	CTT	CAG	GAA	AAG	AAA	CGA
AGG	AAG	GAA	CAA	GAG	GAG	AAA	GCA	GAG	ATA	AAA	CGC	TTA	AAA	AAT	TCT	GAT	GAC	CGG	GAT	TCC	AAG
CGG	GAT	TCC	CTT	GAG	GAG	GGG	GAG	CTG	AGA	GAT	CAC	TGC	ATG	GAG	ATC	ACA	ATA	AGG	AAC	TCC	CCG
TAT	AGA	AGA	GAA	GAC	TCA	ATG	GAA	GAC	AGA	GGA	GAA	GAA	GAT	GAT	TCT	TTG	GCC	ATC	AAA	CCA	CCC
CAG	CAA	ATG	TCT	CGG	AAA	GAA	AAA	GTT	CAT	CAC	AGA	AAA	GAT	GAA	AAG	AGA	AAA	GAA	AAA	TGT	AGG
CAT	CAT	AGC	CAT	TCA	GCA	GAA	GGG	GGG	AAG	CAT	GCT	AGA	GTG	AAA	GAA	AGA	GAG	CAC	GAA	CGT	CGG
AAA	CGA	CAT	CGA	GAA	GAA	CAG	GAT	AAA	GCT	CGC	CGG	GAA	TGG	GAA	AGA	CAG	AAG	AGA	AGG	GAA	ATG
GCA	AGG	GAG	CAT	TCC	AGG	AGA	GAA	AGG	GAC	CGC	TTG	GAG	CAG	TTA	GAA	AGG	AAG	CGG	GAG	CGG	GAG
CGC	AAG	ATG	CGG	GAG	CAG	CAG	AAG	GAG	CAG	CGG	GAG	CAG	AAG	GAG	CGC	GAG	CGG	CGG	CGG	GAG	GAG
CGG	CGC	AAG	GAG	CGG	GAG	GCC	CGC	AGG	GAA	GTG	TCT	GCA	CAT	CAC	CGA	ACG	ATG	AGA	GAG	GAC	TAC
AGC	GAC	AAA	GTG	AAA	GCC	AGC	CAC	TGG	AGT	CGC	AGC	CCG	CCT	CGG	CCG	CCG	CGG	GAG	CGG	TTC	GAG
TTG	GGA	GAC	GGC	CGG	AAG	CCA	GTA	AAA	GAA	GAG	AAA	ATG	GAA	GAA	AGG	GAC	CTG	CTG	TCC	GAC	TTA
CAG	GAC	ATC	AGC	GAC	AGC	GAG	AGG	AAG	ACC	AGC	TCG	GCC	GAG	TCC	TCG	TCA	GCG	GAA	TCA	GGC	TCA
GGT	TCT	GAG	GAA	GAA	GAG	GAG	GAG	GAG	GAG	GAG	GAG	GAG	GAG	GAA	GGG	AGC	ACC	AGT	GAA	GAA	TCA
GAG	GAG	GAG	GAG	GAG	GAA	GAG	GAA	GAG	GAG	GAG	GAG	GAG	GAG	ACC	GGC	AGC	AAC	TCT	GAG	GAG	GCA
GAG	CAG	TCT	GCT	GAA	GAA	GTA	AGT	GAG	GAA	GAA	GAT	AGT	GAA	GAT	GAA	GAA	CGA	AAT	GAA	AAC	
CAC	CTC	TTG	GTT	GTT	CCA	GAG	TCA	CGG	TTC	GAC	CGA	GAT	TCC	GGG	GAG	AGT	GAA	GAA	GCA	GAG	GAA
GAA	GTG	GGT	GAG	GGA	ACG	CCG	CAG	AGC	AGC	GCC	CTG	ACA	GAG	GGC	GAC	TAT	GTG	CCC	GAC	TCC	CCT
GCC	CTG	TTG	CCC	ATC	GAG	CTC	AAG	CAG	GAG	CTG	CCC	AAG	TAC	CTG	CCG	GCC	CTG	CAG	GGC	TGC	CGG
AGC	GTC	GAG	GAG	TTC	CAG	TGC	CTG	AAC	AGG	ATC	GAG	GAG	GGC	ACC	TAT	GGA	GTG	GTC	TAC	AGA	GCA
AAA	GAC	AAG	AAA	ACA	GAT	GAA	ATT	GTG	GCT	CTA	AAG	CGG	CTG	AAG	ATG	GAG	AAG	GAG	AAG	GAG	GGC
TTC	CCG	ATC	ACG	TCC	CTG	AGG	GAG	ATC	AAC	ACC	ATC	CTC	AAG	GCC	CAG	CAT	CCC	AAC	ATT	GTC	ACC
GTT	AGA	GAG	ATT	GTG	GTG	GGC	AGC	AAC	ATG	GAC	AAG	ATC	TAC	ATC	GTG	ATG	AAC	TAC	GTG	GAG	CAC
GAC	CTC	AAG	AGC	CTG	ATG	GAG	ACC	ATG	AAA	CAG	CCC	TTC	CCA	GGG	GAG	GTG	AAG	ACC	CTG	ATG	
ATC	CAG	CTG	CTG	CGG	GGG	GTG	AAA	CAC	CTG	CAC	AAC	TGG	ATC	CTG	CAC	CGT	GAC	CTC	AAG	ACG	
TCC	AAC	CTG	CTG	CTG	AGC	CAC	GCC	GGC	ATC	CTC	AAG	GTG	GGT	GAT	TTT	GGG	CTG	GCG	CGG	GAG	TAC
GGA	TCC	CCT	CTG	AAG	GCC	TAC	ACC	CCG	GTC	GTG	GTG	ACC	CAG	TGG	TAC	CGC	GCC	CCA	GAG	CTG	CTG
CTT	GGT	GCC	AAG	GAA	TAC	TCC	ACG	GCC	GTG	GAC	ATG	TGG	TCA	GTG	GGC	TGC	ATC	TTC	GGG	GAG	CTG
CTG	ACT	CAG	AAG	CCT	CTG	TTC	CCC	GGG	AAT	TCG	GAA	ATC	GAT	CAG	ATC	AAC	AAA	GTG	TTC	AAG	GAG
CTG	GGG	ACC	CCC	AGT	GAG	AAA	ATC	TGG	CCC	GGC	TAC	AGT	GAG	CTC	CCA	GTA	GTC	AAA	AAG	ATG	ACC
TTC	AGC	GAG	CAC	CCC	TAC	AAC	AAC	CTC	CGC	AAG	CGC	TTC	GGG	GCT	CTG	CTC	TCA	GAC	CAG	GGC	TTC
GAC	CTC	ATG	AAC	AAG	TTC	CTG	ACC	TAC	TTC	CCC	GGG	AGG	ATC	AGC	GCT	GAG	GAG	GGC	CTC	AAG	
CAT	GAG	TAT	TTC	CGC	GAG	ACC	CCC	CTC	CCC	ATC	GAC	CCC	TCC	ATG	TTC	CCC	ACG	TGG	CCC	GCC	AAG
AGC	GAG	CAG	CAG	CGT	GTG	AAG	CGG	GGC	ACC	AGC	CCG	AGG	CCC	CCT	GAG	GGA	GGC	CTG	GGC	TAC	AGC
CAG	CTG	GGT	GAC	GAC	GAC	CTG	AAG	GAG	ACG	GGC	TTC	CAC	CTT	ACC	ACC	ACG	AAC	CAG	GGG	GCC	TCT
GCC	GCG	GGC	CCC	GGC	TTC	AGC	CTC	AAG	TTC	TGA											

cdk11 Forward: (5')–AGA.GCT.AGC.GTC.TAC.AGA.GCA.AAA.GAC

(b) cdk11 Reverse: (5')–AGA.GCT.AGC.CCA.AGA.GTC.CTT.TTC.ATC

Seq. de segurança – Sítio da *NheI* – Seq. de hibridização

Figura 16 – Sequência de aminoácidos da proteína truncada CDK11(Δ). O polipeptídeo completo possui 374 resíduos de aminoácidos, e uma massa estimada em 42,3 kDa. Os 20 resíduos iniciais, destacados em fundo amarelo, correspondem à cauda de histidina (His-tag) introduzida pelo pET-28a(+). A seguir, os resíduos em cinza escuro correspondem aos 8 primeiros aminoácidos da CDK11, que foram preservados na sua região N-terminal. Os resíduos em fundo azul claro pertencem ao dipeptídeo Ala-Ser, codificado pelo sítio de restrição da enzima *NheI* introduzido na sequência de DNA. Por fim, os resíduos em cinza claro correspondem à região C-terminal da CDK11, que inclui seu domínio de proteína quinase. Nesta região está destacada a sequência PITSLRE (em rosa), que representa o principal motivo conservado na família das CDKs, por mediar as interações funcionais com ciclinas.

CDK11(Δ)

M	G	S	S	H	H	H	H	H	S	S	G	L	V	P	R	G	S	H	M	G	D	E	K	
D	S	W	A	S	V	Y	R	A	K	D	K	K	T	D	E	I	V	A	L	K	R	L	K	M
E	K	E	K	E	G	F	P	I	T	S	L	R	E	I	N	T	I	L	K	A	Q	H	P	N
I	V	T	V	R	E	I	V	V	G	S	N	M	D	K	I	Y	I	V	M	N	Y	V	E	H
D	L	K	S	L	M	E	T	M	K	Q	P	F	L	P	G	E	V	K	T	L	M	I	Q	L
L	R	G	V	K	H	L	H	D	N	W	I	L	H	R	D	L	K	T	S	N	L	L	L	S
H	A	G	I	L	K	V	G	D	F	G	L	A	R	E	Y	G	S	P	L	K	A	Y	T	P
V	V	V	T	Q	W	Y	R	A	P	E	L	L	L	G	A	K	E	Y	S	T	A	V	D	M
W	S	V	G	C	I	F	G	E	L	L	T	Q	K	P	L	F	P	G	N	S	E	I	D	Q
I	N	K	V	F	K	E	L	G	T	P	S	E	K	I	W	P	G	Y	S	E	L	P	V	V
K	K	M	T	F	S	E	H	P	Y	N	N	L	R	K	R	F	G	A	L	L	S	D	Q	G
F	D	L	M	N	K	F	L	T	Y	F	P	G	R	R	I	S	A	E	D	G	L	K	H	E
Y	F	R	E	T	P	L	P	I	D	P	S	M	F	P	T	W	P	A	K	S	E	Q	Q	R
V	K	R	G	T	S	P	R	P	P	E	G	G	L	G	Y	S	Q	L	G	D	D	D	L	K
E	T	G	F	H	L	T	T	T	N	O	G	A	S	A	A	G	P	G	F	S	L	K	F	

374 resíduos de aminoácidos

MM $\approx$ 42,3 kDa

Como veremos adiante nesta seção (item 5.2), a modificação realizada no DNA plasmidial tornou finalmente possível a expressão da versão encurtada da CDK11 humana em *E. coli*. Portanto, nosso resultado confirma uma influência negativa exercida pela região deletada sobre a expressão da CDK11 no sistema bacteriano. As causas prováveis deste efeito serão discutidas em detalhes no item 5.2.1. A seguir, apresentamos os resultados experimentais das etapas de modificação do DNA que permitiram obter o plasmídeo contendo a sequência codificante da CDK11(Δ). A precisão desta operação foi avaliada pelo sequenciamento do DNA, discutido no item 5.1.3, que demonstrou exatamente o resultado planejado para a sequência nucleotídica do sítio de recombinação.

5.1.2 Obtenção do pET28a::cdk11(Δ) a partir do pET28a::cdk11

De posse dos *primers* desenhados para a etapa de amplificação – Figura 15 (b) – as reações foram preparadas em um volume total de 50 μ L utilizando o *High Fidelity PCR Enzyme Mix* (Thermo Scientific). O volume total foi a princípio dividido em 5 alíquotas de 10 μ L, para que pudéssemos avaliar os produtos da reação experimentando diferentes temperaturas durante a etapa de hibridização. O resultado do teste de amplificação das 5 amostras idênticas, distribuídas em um gradiente de temperatura de 51 a 58 $^{\circ}$ C, está apresentado na Figura 17. Em cada raia podemos observar uma banda principal, entre 6.000 e 8.000 pb, que corresponde ao fragmento de interesse (6.391 pb). Na última raia aplicamos uma amostra do pET28a::cdk11 linearizado, para indicar o tamanho do molde de DNA. Todas as amostras apresentaram um elevado rendimento para a amplificação do fragmento de interesse, entretanto todas elas também geraram produtos secundários na reação. Neste sentido, o melhor resultado foi obtido para a temperatura de 53,5 $^{\circ}$ C.

Figura 17 – Teste de amplificação em diferentes temperaturas de hibridização. As amostras do PCR foram aplicadas em gel de agarose para podermos visualizar os produtos da amplificação. Os valores sobre as raias indicam a temperatura escolhida para a hibridização de cada amostra. O marcador *GeneRuler™ 1 kb DNA Ladder* (Fermentas) foi aplicado na primeira raia, e o tamanho dos fragmentos na região de interesse está indicado pelas setas. Na última raia foi aplicada uma amostra do plasmídeo pET28a::cdk11, previamente linearizado pela digestão com a enzima *Sall*. O tamanho do amplicon de interesse é de 6.391 pb, enquanto o pET28a::cdk11 (DNA molde) possui 7.657 pb. Observamos que a temperatura de 53,5 $^{\circ}$ C revela a menor fração de subprodutos amplificados.

Como a temperatura de 53,5 °C parece ter fornecido a menor quantidade de subprodutos na amplificação, repetimos a reação utilizando somente esta temperatura para o volume total de 50 µL. A amostra foi aplicada em gel de agarose para visualização do seu conteúdo e separação do DNA de interesse. A Figura 18 mostra o resultado da aplicação de todo o produto do PCR em um único gel. Nesta aplicação, os poços do gel foram fundidos para que a fenda resultante pudesse comportar todo o volume da reação de amplificação, facilitando assim o isolamento da banda.

Figura 18 – Separação do fragmento amplificado a partir do pET28a::cdk11. Após o PCR empregando a temperatura ótima de 53,5 °C na hibridização dos *primers*, o volume total da reação (50 µL) foi aplicado no gel para separação do fragmento de interesse. Como sugerido pelo teste anterior, o gel revela um bom rendimento e especificidade da reação de amplificação para o produto desejado. O fragmento de DNA foi extraído deste gel e purificado com o *QIAquick Gel Extraction kit* (QIAGEN).

A quantificação espectrofotométrica (pela absorbância a 260 nm) do DNA purificado forneceu uma concentração de 80,5 ng/µL para a amostra de 50 µL. Assim, o rendimento total foi cerca de 4,00 µg para o processo completo de recuperação do DNA a partir do gel. O amplicon purificado foi então submetido às reações consecutivas de clivagem e religação das extremidades, conforme esquematizado na Figura 14 e descrito na seção de Materiais e Métodos. Por fim, o produto de clivagem/religação foi utilizado para transformar a cepa DH5α, e obtivemos o plasmídeo recombinante pET28a::cdk11(Δ) purificado empregando o kit de extração de DNA plasmidial *GeneJet™ Plasmid Miniprep* (Fermentas).

5.1.3 Ensaios confirmatórios para o produto de recombinação

Após as reações de clivagem e religação com o fragmento de DNA amplificado, era necessário confirmar o sucesso da recombinação antes de partir para a transformação das cepas de expressão. Desta maneira, o produto da recombinação foi inicialmente utilizado para transformar a cepa de clonagem DH5 α . Na placa de transformação, seis colônias individuais (designadas pelas letras de A a F) foram escolhidas para crescer culturas em meio LB líquido. Os plasmídeos clonados em cada uma destas culturas foram extraídos e purificados separadamente, empregando o kit de extração de DNA plasmidial *GeneJet™ Plasmid Miniprep* (Fermentas). As seis amostras resultantes foram submetidas a um teste de clivagem com as endonucleases de restrição *Sall* e *NdeI*, cujo resultado está apresentado na Figura 19.

Figura 19 – Teste de clivagem dos plasmídeos clonados a partir do produto de recombinação. As amostras de DNA plasmidial foram obtidas de seis colônias DH5 α individuais transformadas com o produto bruto das reações de clivagem e ligação. Cada uma das amostras purificadas - A, B, C, D, E e F - foi aplicada no gel em três condições distintas: (I) sem sofrer digestão, (II) digerida com *Sall*, e (III) digerida com *Sall* e *NdeI*. Considerando-se que os sítios *Sall* e *NdeI* delimitam a sequência codificante da CDK11 no pET-28a(+), a clivagem com *Sall* apenas lineariza o plasmídeo, enquanto a clivagem com ambas *Sall* e *NdeI* separa o inserto do vetor, gerando dois fragmentos distintos. É possível observar a liberação de um fragmento menor (entre 1.000 e 1.500 pb) na raia (III) de todas as amostras, que deve corresponder à sequência codificante da CDK11(Δ), com 1.065 pb.

Podemos constatar na Figura 19 que todas as amostras revelaram um padrão semelhante para a altura das bandas em raias correspondentes. As raias (I) devem conter o plasmídeo pET28a::cdk11(Δ) íntegro, naturalmente na forma de DNA circular. Devido à existência de diferentes topoisômeros do plasmídeo – ou seja, diferentes formas supertorcidas do mesmo DNA circular – observamos sempre mais de uma banda nestas canaletas. Por outro lado, as raias (II) devem conter o plasmídeo pET28a::cdk11(Δ) linearizado. Como não existem diferentes topoisômeros para fragmentos de DNA linear, observamos uma única banda situada na mesma altura em todas elas. Finalmente, nas raias (III) devemos observar a presença de duas bandas distintas, referentes à separação do inserto e do vetor. Assim, a banda mais alta corresponde ao vetor clivado – pET28a-[*Sall/NdeI*] – enquanto a banda mais baixa corresponde à sequência codificante da CDK11(Δ).

Embora os perfis de todas as seis amostras sigam o mesmo padrão no teste de clivagem, duas variações importantes precisam ser esclarecidas. Em primeiro lugar, chamamos atenção para a inversão acidental na ordem das raias (I) e (II) para a amostra B. A segunda discrepância refere-se à presença na raia (III), em alguns casos, de uma banda mais alta originária da raia (II), a qual corresponde ao plasmídeo linearizado (6.379 pb). Como ocorre apenas nas amostras A e B, a banda extra deve indicar problemas com a qualidade da preparação destas amostras. Normalmente, a presença de etanol residual, a remoção ineficiente do meio de cultura e a eluição da amostra utilizando concentrações salinas muito elevadas são as principais causas de resistência à clivagem do DNA plasmidial purificado pelas enzimas de restrição. Portanto, embora o teste de clivagem tenha indicado êxito na obtenção do pET28a::cdk11(Δ) em todos os casos, optamos por descartar as amostras A e B antes de avançar na caracterização dos produtos clonados em DH5 α .

Para complementar o resultado do teste de clivagem em paralelo das amostras isoladas nas mini-preparações (*minipreps*), também realizamos um teste de clivagem comparativo destas com o plasmídeo original, pET28a::cdk11. A Figura 20 apresenta o resultado do teste com a amostra C, que revela um perfil idêntico ao produzido pelas demais amostras restantes (D, E, e F). Do mesmo modo que no teste anterior, as amostras foram clivadas com *Sall* e *NdeI* para liberar o inserto correspondente à sequência codificante da

CDK11. Como podemos observar na Figura 20, os plasmídeos testados diferem apenas pelo tamanho do inserto recombinado entre os sítios *Sall* e *NdeI* no pET28a. A comparação entre os perfis de clivagem do DNA antes e depois da modificação realizada indica um resultado consistente com o produto almejado.

Figura 20 – Teste de clivagem comparativo dos plasmídeos contendo a sequência codificante da CDK11 antes e depois da modificação. As amostras foram clivadas pelas mesmas enzimas de restrição - *Sall* e *NdeI* - e divididas entre as raia (I), (II) e (III), como descrito para o teste anterior. As setas à direita do gel indicam os produtos de clivagem representados pelas bandas presentes na raia (III) de cada amostra. Em comum, ambas apresentam uma banda superior que corresponde ao vetor pET-28a(+) com o trecho entre *Sall* e *NdeI* removido. Por outro lado, observamos uma diferença de tamanho para os insertos liberados pelo vetor nas duas amostras. A diferença de mais de 1.000 pb entre estas bandas está de acordo com o resultado pretendido para a deleção interna na sequência codificante da CDK11.

As amostras que forneceram o resultado esperado nos testes de clivagem (C, D, E, e F) foram submetidas ao próximo ensaio confirmatório, baseado em uma reação de amplificação utilizando os *primers* T7 do vetor. Os *primers* T7 são complementares a regiões localizadas nas extremidades do sítio múltiplo de clonagem dos vetores pET (Novagen), sobrepondo-se parcialmente aos elementos promotor e terminador reconhecidos pela T7 RNA polimerase. Desta maneira, a amplificação com o par de *primers* T7 abrange toda a sequência a ser transcrita pela T7 RNA polimerase durante o processo de indução. A partir do tamanho do fragmento amplificado é fácil deduzir o tamanho de um inserto recombinado

ao *polylinker*, e então analisamos se este é compatível com o produto de deleção planejado. O resultado do teste de amplificação das amostras de DNA plasmidial com os *primers* T7 está mostrado na Figura 21. Todas as quatro amostras exibiram um perfil semelhante no gel de agarose, o qual é consistente com o padrão esperado para uma recombinação bem sucedida.

Figura 21 – Teste de amplificação com os plasmídeos pET28a::cdk11(Δ) utilizando os primers T7 do vetor. A reação de amplificação foi realizada utilizando como molde as amostras de DNA plasmidial que apresentaram resultados consistentes com o produto desejado nos testes de clivagem (C, D, E, e F). Os *primers* T7 promovem a amplificação de toda a região que contém o sítio múltiplo de clonagem e as sequências regulatórias para expressão no vetor, cobrindo um total de 318 pb no vetor pET-28a(+) intacto. Subtraindo-se o trecho de 55 pb removido entre *NdeI* e *Sall* e adicionando os 1.065 pb da sequência codificante da CDK11(Δ), resulta um tamanho de 1.328 pb esperado para o amplicon. O gel revela a presença de grandes bandas pouco abaixo da marca de 1.500 pb para todas as amostras testadas, confirmando que o inserto possui o tamanho pretendido pela deleção.

Os ensaios confirmatórios de clivagem e amplificação, portanto, indicam êxito no processo de modificação do DNA plasmidial para produzir a recombinação desejada na sequência codificante da CDK11. Contudo, estes experimentos servem apenas para descartar a presença dos subprodutos mais comuns da reação de recombinação e confirmar o padrão esperado para o produto de interesse. Afinal, os géis de agarose não fornecem sequer uma estimativa precisa do tamanho das bandas observadas, muito menos informações sobre a sequência nucleotídica dos fragmentos. Mesmo assim, a obtenção dos

resultados previstos nestes ensaios confirmatórios é um forte indicativo de que as amostras de DNA plasmidial realmente correspondem ao produto de recombinação desejado. Para confirmar definitivamente o sucesso da nossa estratégia de modificação na sequência codificante da CDK11, as amostras foram enviadas para sequenciamento do DNA. O sequenciamento foi realizado pelo método de Sanger (dos terminadores de cadeia) automatizado, empregando os mesmos *primers* T7 que utilizamos no teste de amplificação. A Figura 22 ilustra o resultado do sequenciamento de uma amostra, destacando o sítio *NheI* produzido pela recombinação das extremidades coesivas durante o fechamento (ciclização) do pET28a::cdk11(Δ).

Figura 22 – Resultado do sequenciamento de DNA evidenciando o sítio de recombinação na sequência codificante da CDK11(Δ). A Figura apresenta o cromatograma gerado pelo software *Chromas* 2.4.1 (Technelysium) com o resultado do sequenciamento para o trecho correspondente aos primeiros 18 códons da CDK11(Δ). O códon de iniciação original da CDK11 – que neste caso está localizado no interior da sequência codificante devido à cauda de histidina N-terminal – encontra-se destacado em amarelo. No centro, os códons destacados em cinza constituem o sítio *NheI* introduzido no lugar do segmento de 1.284 pb deletado. A sequência que observamos aqui deve ser comparada à sequência codificante original da CDK11, mostrada na Figura 15(a). O sítio *NheI* substituiu todo o trecho marcado em cinza na sequência original. Embora só esteja apresentado o resultado para o trecho de interesse, o sequenciamento completo forneceu exatamente a sequência planejada para a CDK11(Δ).

O sequenciamento confirmou o resultado esperado para todas as três amostras de DNA plasmidial enviadas (C, D, e E), com ausência de quaisquer mutações na sequência codificante da CDK11(Δ). Entretanto, focamos nossa análise nas vizinhanças do sítio de recombinação porque este trecho envolve toda a sequência modificada (inserida/deletada) e representa um intervalo crítico para o surgimento de mutações. É importante mencionar que, como observamos na Figura 22, os cromatogramas resultantes do sequenciamento exibiram em geral bandas bem definidas e regularmente espaçadas na região de interesse, com intensidade muito superior à linha de base do sinal. Estas características funcionam como indicadores de qualidade para o sequenciamento e demonstram a precisão do resultado obtido. Portanto, os ensaios realizados com as amostras de DNA plasmidial clonadas em DH5 α comprovam definitivamente que o produto de recombinação isolado corresponde ao pET28a::cdk11(Δ). Uma vez verificado que nossa estratégia de deleção foi bem sucedida, podemos finalmente testar seu efeito sobre a expressão da CDK11 em *E. coli*. Para concluir, a Tabela 4 sumariza todas as diferentes sequências de DNA relacionadas ao trabalho de obtenção do pET28a::cdk11(Δ), descrito neste tópico.

Tabela 4 – Sequências de DNA relacionadas à CDK11. A Tabela resume as sequências de DNA que são mencionadas no texto, indicando seu tamanho em pares de bases e a massa molecular do polipeptídeo codificado (quando for o caso) estimada pelo *ProtParam* (ExPASy). A sequência nucleotídica que pretendemos expressar no sistema bacteriano - a partir do novo plasmídeo pET28a::cdk11(Δ) - corresponde na realidade à His-tag-CDK11(Δ). Contudo, passaremos a omitir a His-tag por conveniência, nos referindo à proteína resultante simplesmente como CDK11(Δ).

Sequência	Observação	DNA	Proteína
cdk11		2.343 pb	91,02 kDa
	- Deleção	- 1.284 pb	- 51,09 kDa
	+ Inserção	+ 6 pb	+ 0,16 kDa
cdk11(Δ)		1.065 pb	40,09 kDa
	+ Cauda de poli-histidina	+ 60 pb	+ 2,18 kDa
His-tag-cdk11(Δ)		1.125 pb	42,27 kDa
pET28a		5.369 pb	-
	- Trecho entre <i>Sall</i> e <i>NdeI</i>	- 52 pb	-
pET28a-[<i>NdeI/Sall</i>]		5.317 pb	-
pET28a::cdk11	= pET28a - [<i>NdeI/Sall</i>] + cdk11 – 3 pb (sobreposição entre <i>NdeI</i> e ATG iniciação)	7.657 pb	-
pET28a::cdk11(Δ)	= pET28a - [<i>NdeI/Sall</i>] + cdk11(Δ) – 3 pb (sobreposição entre <i>NdeI</i> e ATG iniciação)	6.379 pb	-

5.2 Obtenção das proteínas CDK9 e CDK11 humanas em *E. coli*

O trabalho que desenvolvemos para obtenção das CDKs 9 e 11 através do sistema bacteriano foi precedido por um estudo pioneiro, do qual a Prof.^a Dr.^a Fernanda Canduri participou como autora (LEOPOLDINO et al., 2006). Neste estudo foi conduzida pela primeira vez a expressão e purificação da CDK9 humana biologicamente ativa a partir de um sistema heterólogo (*E. coli*). Embora a pesquisa sobre CDKs reguladoras da transcrição já tenha se desenvolvido por mais de duas décadas, sua expressão ectópica tem sido invariavelmente realizada em células eucarióticas. A expressão de CDKs humanas em *E. coli* representaria uma alternativa muito vantajosa em termos do rendimento obtido, custos operacionais, simplicidade do método e facilidade de purificação. Por outro lado, a expressão de proteínas eucarióticas em células bacterianas é frequentemente prejudicada pela sua incompatibilidade natural com o sistema heterólogo. Ainda que tenhamos observado uma série de dificuldades neste sentido, os resultados publicados (LEOPOLDINO et al., 2006) haviam demonstrado ser viável a obtenção de CDKs humanas no sistema bacteriano em quantidade e pureza suficientes para a realização de estudos estruturais.

A expressão da CDK9 humana em *E. coli* já tinha sido alcançada no nosso grupo a partir do mesmo plasmídeo pET23a::cdk9 originalmente produzido e utilizado pelo estudo de Leopoldino et al., (2006). Entretanto, pela necessidade de se aperfeiçoar o procedimento de purificação e adequá-lo às disponibilidades do laboratório, foram comercialmente adquiridos os plasmídeos pET28a::cdk9 e pET28a::cdk11 (*Epoch Life Science*) que codificam uma cauda de histidina N-terminal fusionada às CDKs. No caso da CDK11, sua expressão nunca havia sido realizada no sistema bacteriano, e o plasmídeo foi empregado com o intuito de se adaptar o protocolo estabelecido para obtenção da CDK9. Uma vez em posse dos novos plasmídeos, realizamos alguns ensaios preliminares de expressão em células BL21(DE3). Os resultados mostraram que a CDK9 é expressa com relativa facilidade em *E. coli* a partir do pET28a::cdk9, de modo semelhante ao observado previamente para o plasmídeo pET23a::cdk9. Ao contrário, a expressão da CDK11 não foi obtida mesmo após várias tentativas de indução em diferentes condições experimentais.

A Figura 23 apresenta o resultado inicial dos ensaios de expressão com os plasmídeos pET28a::cdk9 e pET28a::cdk11 em BL21(DE3).

Figura 23 – Resultado preliminar para a expressão das CDKs 9 e 11 em *E. coli*. Gel de poliacrilamida a 15% (SDS-PAGE) mostrando o perfil de diferentes amostras coletadas em um ensaio preliminar de expressão das CDKs 9 e 11 no sistema bacteriano. A expressão foi induzida em culturas de células BL21(DE3) - transformadas com os plasmídeos pET28a::cdk9 ou pET28a::cdk11 - empregando 0,2 mmol/L de IPTG a 25 °C durante 15:00 horas (*overnight*). Para cada CDK, na raia (I) foi aplicada uma amostra da cultura bacteriana imediatamente antes da adição do IPTG, enquanto a raia (II) corresponde à amostra coletada ao final da indução. As raias (III) e (IV) consistem, respectivamente, na fração sobrenadante (solúvel) e sedimentada (insolúvel) coletadas após a centrifugação do lisado bruto das células induzidas. Constata-se no gel a expressão da CDK9, bem como sua presença exclusivamente na fração insolúvel do lisado. Contudo, para a CDK11 esperávamos a ocorrência de uma banda de expressão por volta de 110 kDa, a qual não foi possível observar.

Examinando a Figura 23 fica evidente a superexpressão da CDK9 pelo sistema bacteriano, embora ocorram algumas desconformidades - quanto à posição da banda e sua ausência na raia (III) - que serão discutidas mais adiante. Porém, o experimento de indução parece não ter produzido qualquer mudança apreciável no perfil de expressão das células transformadas com o plasmídeo pET28a::cdk11. Este resultado negativo se repetiu em todas as tentativas posteriores de obter a expressão da CDK11 na cepa BL21(DE3) utilizando condições mais drásticas na indução (temperaturas mais altas e maiores concentrações de IPTG). A alternativa proposta foi a deleção do segmento de DNA que codifica a região N-terminal da CDK11 no pET28a::cdk11, de modo a reduzir a proteína basicamente ao seu domínio quinase. O processo de obtenção do novo plasmídeo contendo esta modificação - pET28a::cdk11(Δ) - é o tema central do tópico anterior (5.1).

5.2.1 Expressão da CDK11(Δ) no sistema bacteriano

A expressão da CDK11 em *E. coli* assume especial interesse no nosso trabalho pelo fato de não existirem referências que descrevam sua obtenção a partir de um sistema bacteriano. Ainda, a CDK11 permanece sendo um dos membros mais deficientes em termos de estudos estruturais na família das CDKs. Neste sentido não há prejuízo na expressão de uma forma truncada da proteína, pois seu domínio quinase é o responsável pela atividade enzimática da CDK11 e nele se situam todos os elementos determinantes da especificidade de ligação aos substratos e catálise. Assim, a expressão heteróloga em *E. coli* pode representar o primeiro passo na obtenção da CDK11 em grandes quantidades para aplicações que demandem amostras proteicas com elevada pureza (como ensaios de cristalização, dicroísmo circular, espectrometria de massas, etc.). Uma vez alcançada a expressão heteróloga, o próximo desafio naturalmente consiste em adaptar as condições experimentais para obter a proteína solúvel e em sua forma biologicamente ativa.

A proteína truncada CDK11(Δ) possui apenas 40,1 kDa (sem incluir a cauda de polihistidina) dos 91,0 kDa referentes à variedade da CDK11 humana cuja expressão fora inicialmente tentada a partir do pET28a::cdk11. Esta CDK11 original representa uma das múltiplas isoformas de maior massa – geradas por *splicing* alternativo do gene CDC2L2 – que são coletivamente designadas CDK11A^{p110}. É importante frisar que, embora sejam referidas pela sua massa molecular aparente de 110 kDa, tal valor resulta da migração eletroforética anômala das isoformas CDK11^{p110} em géis SDS-PAGE (XIANG et al., 1994). Por este motivo se afirma na legenda da Figura 23 que a banda correspondente à expressão da CDK11 deveria ser observada em torno de 110 kDa, mesmo sendo a massa real da His-tag-CDK11 estimada em 93,2 kDa. Como a CDK11(Δ) foi obtida pela deleção do N-terminal da CDK11 original, de modo que restasse apenas o domínio quinase da proteína, ela apresenta grande semelhança em relação às espécies CDK11^{p46} ativadas por caspases. A extensa região N-terminal removida, por sua vez, é semelhante ao fragmento CDK11^{p60}, o produto de clivagem por caspases complementar à CDK11^{p46}. Curiosamente, devido à proximidade entre a CDK11(Δ) e as isoformas CDK11^{p46}, é bastante provável que a CDK11(Δ) não pudesse ser obtida pela expressão em linhagens celulares de eucariotos superiores, por induzir as células à apoptose como observado na expressão ectópica da CDK11^{p46} (MIKOLAJCZYK et al., 2003).

A estratégia de deleção N-terminal foi bem sucedida e permitiu finalmente a expressão da CDK11(Δ) pelo sistema bacteriano, como mostra a Figura 24. Entretanto, constatamos no gel que a totalidade da CDK11(Δ) sintetizada pela bactéria se localiza na fração insolúvel do lisado, o que seria quase inevitável neste experimento considerando-se a elevada temperatura e concentração de IPTG aplicadas, além da longa duração para uma indução a 37 °C.

Figura 24 – Expressão da proteína CDK11(Δ) em *E. coli*. Este gel de poliacrilamida apresenta o primeiro resultado positivo que obtivemos para a expressão da CDK11 pelo sistema bacteriano. A expressão foi induzida em células BL21(DE3) transformadas com o plasmídeo pET28a::cdk11(Δ) empregando 1,0 mmol/L de IPTG e a temperatura de 37 °C durante 7:00 horas. A raia (I) corresponde à amostra retirada da cultura antes da adição do IPTG, e a raia (II) àquela coletada no final do processo de indução. A seguir, a raia (III) contém a fração insolúvel do lisado bruto, e a raia (IV) sua fração solúvel (as posições estão invertidas em relação ao gel da Figura anterior). Observamos a superexpressão da CDK11(Δ) no sistema heterólogo, embora a proteína esteja localizada exclusivamente na fração insolúvel do lisado.

Como ficou demonstrado, a deleção realizada na sequência codificante da CDK11 cumpriu seu propósito de favorecer a expressão da proteína no sistema bacteriano. O resultado indica, portanto, que a região deletada exerce uma influência negativa sobre a expressão da proteína pela bactéria. Tentamos determinar a causa deste efeito negativo analisando a sequência de DNA removida, que se encontra destacada em cinza na Figura 15. Logo ficou evidente que um fator com grande potencial para dificultar a expressão da CDK11

pela bactéria deveria ser a elevada quantidade de códons raros de Arg (AGG, AGA e CGA) presentes na sequência codificante. Originalmente, a sequência nucleotídica introduzida no vetor possuía 36 códons raros de arginina, sendo que 30 deles encontram-se no segmento que foi eliminado. O problema é ainda agravado pela ocorrência neste trecho de quatro repetições duplas dos códons raros de Arg. Os códons AGG e AGA são os mais raros em *E. coli* (ou seja, aqueles que ocorrem naturalmente com menor frequência no seu próprio genoma), e assim são também os que produzem o efeito negativo mais acentuado sobre a expressão. Portanto, a elevada prevalência de códons raros na sequência codificante parece fornecer motivo suficiente para um intenso efeito negativo da região deletada sobre a expressão da CDK11.

Para avaliar quantitativamente o impacto da presença de códons raros sobre a expressão de uma proteína, a variável mais utilizada é o índice de adaptabilidade dos códons (CAI). Conforme definido originalmente (SHARP et al., 1987), o valor de CAI corresponde à média geométrica das adaptabilidades relativas (w) de todos os códons da sequência, sendo a adaptabilidade relativa de cada códon dada pela razão entre a frequência relativa deste códon (f) e a frequência relativa do códon sinônimo mais comum. A fórmula que exprime matematicamente esta definição para o CAI está apresentada na Figura 25(b). Os valores numéricos das frequências relativas dependem da espécie considerada, e podem ser encontrados na forma de tabelas do uso relativo de códons sinônimos (*relative synonymous codon usage* – *RSCU*). Para a elaboração de uma tabela *RSCU*, no entanto, não são utilizadas as frequências relativas dos códons em todo o genoma do organismo: Apenas os genes com altos níveis de expressão são levados em conta, pois se presume nestes casos que a seleção natural tenha atuado no sentido de maximizar a eficiência da tradução.

De acordo com sua definição, o valor de CAI estimado para qualquer sequência codificante (assim como a frequência relativa de cada códon) estará sempre no intervalo entre 0 e 1. Nesta escala, quanto mais alto o valor resultante para o CAI, maior deve ser a eficiência no processo de tradução da sequência nucleotídica pela célula hospedeira (e maiores os níveis de expressão atingidos). Contudo, é preciso esclarecer que o valor de CAI não representa uma medida absoluta da eficiência na tradução – e muito menos dos níveis de expressão – devido à multiplicidade de fatores que normalmente influenciam a regulação

de qualquer gene. Os valores de CAI para as sequências codificantes da CDK11 (original) e CDK11(Δ) estão apresentados na Figura 25(a), juntamente com seus respectivos gráficos de frequência relativa para os códons.

Figura 25 – Avaliação do conteúdo de códons raros na sequência codificante da CDK11. (a) Comparação entre os valores de CAI para a sequência nucleotídica da CDK11 antes e depois da deleção interna. Nesta análise são consideradas as sequências completas a serem expressas em *E. coli*, incluindo os 60 pb referentes à His-tag codificada pelo vetor. Os gráficos e os valores de CAI foram gerados utilizando a ferramenta *Rare Codon Analysis Tool* (GenScript). (b) Equação que exprime o índice de adaptabilidade para uma sequência nucleotídica formada por L códons. Observamos que a variável CAI é representada pela média geométrica das adaptabilidades relativas dos códons na sequência, o que corresponde à raiz L -ésima do produto de todas as adaptabilidades relativas individuais (w_i). Os valores numéricos para as frequências relativas podem ser consultados em tabelas (*RSCU*) para cada espécie de interesse.

A análise dos valores de CAI fornecidos na Figura 25(a) para as sequências codificantes da CDK11 demonstra um resultado inesperado. Em primeiro lugar, a diferença de apenas 0,03 unidades observada entre os valores de CAI calculados é desprezível, e assim a variação no índice de adaptabilidade ocasionada pela deleção interna foi praticamente nula. Mais surpreendente, no entanto, é a constatação de que os valores de CAI obtidos não podem ser interpretados como desfavoráveis à expressão, pois um índice de adaptabilidade superior a 0,6 é considerado bastante elevado para uma proteína de origem heteróloga. De

fato, o valor de CAI estimado para a CDK9 em *E. coli* é de 0,65 (incluindo a cauda de histidina), e não observamos maiores dificuldades na sua expressão pelo sistema bacteriano. O cálculo dos índices de adaptabilidade, portanto, parece contrariar a hipótese anterior de que a expressão da CDK11 em *E. coli* seja comprometida pela abundância de códons raros na sua sequência nucleotídica.

Talvez, a maior limitação no emprego de CAI para avaliar o impacto de códons raros sobre a expressão de uma proteína seja sua indiferença em relação às posições que os códons ocupam na sequência codificante. Como o valor de CAI depende do produtório das adaptabilidades relativas individuais, importa apenas a quantidade de códons de cada tipo presentes, mas não a distribuição particular destes códons ao longo da sequência (o que equivale a afirmar matematicamente que a ordem dos fatores não altera o produto). Assim, o índice de adaptabilidade não leva em consideração o possível acúmulo de códons raros em determinadas regiões da sequência nucleotídica, embora os efeitos de uma distribuição heterogênea sejam reconhecidamente importantes (SHARP et al., 1987). Portanto, a grande concentração de códons raros de Arg na região deletada – e principalmente a ocorrência de quatro repetições duplas destes códons – poderia afinal justificar as dificuldades que relatamos para a expressão da CDK11 no sistema bacteriano. Então, o uso de cepas portadoras de cópias extras para os genes de RNAs decodificadores dos códons raros de arginina talvez fornecesse uma alternativa promissora para a expressão heteróloga da CDK11.

Um teste utilizando as cepas de expressão especializadas na decodificação de códons raros poderia esclarecer definitivamente se os códons raros da sequência são os verdadeiros responsáveis pelo fracasso das tentativas anteriores de obtenção da CDK11 humana em *E. coli*. Com o objetivo de confirmar a hipótese de que a presença de códons raros constitui o fator determinante para a incapacidade de expressão da CDK11 pela bactéria, as cepas Rosetta(DE3), BL21(DE3) CodonPlus-RP e BL21(DE3) CodonPlus-RIL foram transformadas com o plasmídeo pET28a::cdk11 e submetidas a condições muito favoráveis de indução. Surpreendentemente, porém, não foi possível detectar a expressão heteróloga da CDK11 em nenhum dos experimentos utilizando as cepas mencionadas acima. O resultado de um deles é apresentado na Figura 26.

Figura 26 – Teste de expressão das proteínas CDK11 e CDK11(Δ) em Rosetta (DE3). Este gel compara a expressão dos plasmídeos pET28a::cdk11 e pET28a::cdk11(Δ) em Rosetta(DE3). A cepa de expressão Rosetta(DE3) (Novagen) possui cópias extras de genes dos RNAs transportadores que decodificam 7 dos principais códons raros em *E. coli*, inclusive os códons AGG, AGA e CGG de arginina. A indução empregou 1,0 mmol/L de IPTG e a temperatura de 37 °C durante 4:00 horas. As raias (I) correspondem à amostra retirada da cultura antes da adição do IPTG, e as raias (II) àquelas coletadas no final do processo de indução. É evidente que foi obtida a expressão da CDK11(Δ) mas não da CDK11 original, que deveria aparecer em torno de 110 kDa no gel. O resultado do experimento sugere que a suplementação de códons raros não é suficiente para promover a expressão da CDK11A^{p110} no sistema bacteriano.

Assim como no teste de expressão em Rosetta(DE3) mostrado na Figura 26, as tentativas de obter a CDK11 a partir das cepas BL21(DE3) CodonPlus-RP e BL21(DE3) CodonPlus-RIL (Stratagene) foram igualmente infrutíferas. Este resultado contraria a expectativa inicial de que a CDK11 poderia ser expressa pela bactéria se fosse eliminado o obstáculo representado pela tradução dos códons raros. No entanto, se a ocorrência dos códons raros de arginina e suas duplas na porção removida não constitui o principal fator responsável pela dificuldade de obtenção da CDK11, qual seria afinal o motivo desta persistente inibição? De fato, ao analisar a sequência de aminoácidos do polipeptídeo codificado pela região deletada, desde o princípio fora constatado um desequilíbrio marcante na abundância relativa dos resíduos: mais de 25% da sua sequência corresponde

aos códons GAG e GAA, que codificam o aminoácido glutamato. A possibilidade de que esta proporção incomum de resíduos de glutamato causasse dificuldades na expressão heteróloga já havia sido considerada, mas a distribuição dos códons raros parecia fornecer um motivo mais importante para inibir a tradução da região N-terminal da CDK11.

A elevada prevalência de resíduos de ácido glutâmico na porção N-terminal das isoformas maiores da CDK11 foi reconhecida há bastante tempo (XIANG et al., 1994), embora praticamente nada tenha sido investigado em todos estes anos a respeito da sua função. Devido ao excesso de resíduos Glu, o segmento removido da proteína apresenta um ponto isoelétrico (pI) muito baixo, estimado em 4,96 pela ferramenta *ProtParam* (ExPASy). O caráter ácido da porção deletada da CDK11 só não é ainda mais acentuado porque os resíduos básicos de Arg e Lys representam, juntos, mais de 21% da sua sequência. Um valor de pI ao redor de 5,0 não é considerado atípico para proteínas de *E. coli* (SCHWARTZ et al., 2001), porém a distribuição do aminoácido Glu é muito heterogênea na sequência, incluindo um trecho crítico de 33 resíduos que contém 28 glutamatos. É bem provável que a densidade excepcionalmente alta de cargas negativas em uma pequena região prejudique de algum modo a síntese ou estabilidade da cadeia polipeptídica no sistema heterólogo. Em qualquer destes cenários, o mecanismo que impede a expressão da proteína deve atuar em um nível pós-transcricional, estando de acordo com a ausência de fatores identificados que poderiam influenciar negativamente a transcrição e estabilidade do RNAm da CDK11.

Para justificar a dificuldade observada na expressão da CDK11 pelo sistema bacteriano como consequência das repetições de glutamato na porção N-terminal da sequência polipeptídica, é necessário esclarecer que os princípios que regem a reiteração de aminoácidos em proteínas são essencialmente diferentes entre procariotos e eucariotos (GREEN et al., 1994; MARCOTTE et al., 1998; KATTI et al., 2000). Em primeiro lugar, repetições genéricas na sequência polipeptídica são três vezes mais comuns em proteínas eucarióticas que em procarióticas. Segundo, apenas 4% das repetições que ocorrem em eucariotos apresentam homólogos entre procariotos (MARCOTTE et al., 1998). As repetições de um único aminoácido em proteínas eucarióticas são, na grande maioria dos casos, relacionadas ao códon CAG (glutamina). Assim, repetições que possuam os dinucleotídeos CA, AG ou GC – mesmo ocorrendo em códons separados na pauta de leitura – são

privilegiadas, e ocorrem com alta frequência nos genes de eucariotos (GREEN et al., 1994). Os dois códons que correspondem ao glutamato (GAG e GAA) resultam num grande número de duplas AG quando repetidos em série, motivo pelo qual Glu é um dos aminoácidos mais comumente reiterados nas proteínas eucarióticas.

As proteínas mais bem caracterizadas que apresentam motivos de poliglutamato – ou poli(E) – parecem atuar justamente realizando funções típicas das células eucarióticas. Este é o caso da proteína do centrômero/cinetócoro CENP-B (XIANG et al., 1994), que se destaca por possuir uma região de 66 resíduos na porção C-terminal contendo 48 resíduos de glutamato, 17 dos quais formam uma série ininterrupta. Também pertencem ao grupo as proteínas HMGB1 e nucleoplasmina, que se utilizam da sua intensa carga eletrostática negativa para interagir com histonas e facilitar a montagem dos nucleossomos (DINGWALL et al., 1987). Outra proteína com motivos de poliglutamato é a Sec7p de *Saccharomyces cerevisiae*, que tem um papel regulatório na membrana do aparelho de Golgi (GREEN et al., 1994). Ainda, a sialoproteína óssea (BSP) é expressa em tecidos mineralizados de vertebrados, e atua como promotora da nucleação de hidroxiapatita na matriz extracelular (TYE et al., 2003).

Na realidade, embora motivos de poliglutamato não sejam comumente encontrados em proteínas bacterianas, é preciso reconhecer que a expressão da sialoproteína óssea murina já foi obtida com sucesso em *E. coli* (TYE et al., 2003). No entanto, a sialoproteína óssea apresenta sua maior proporção de glutamato em um trecho de 25 resíduos, dos quais 17 são Glu, incluindo uma série contínua de 10 resíduos de glutamato. Na CDK11, por outro lado, o trecho mais crítico ocorre em um segmento de 33 resíduos dos quais 28 são de glutamato, incluindo duas séries contínuas de 13 resíduos Glu. É provável que o sistema bacteriano não possua a mesma capacidade da célula eucariótica de traduzir uma sequência polipeptídica tão incomum, ou talvez devido à altíssima densidade de cargas negativas no trecho assinalado a proteína não seja estável no ambiente interno da célula procariótica. Esta hipótese certamente carece de confirmação, mas representa o motivo mais plausível para o intenso efeito negativo da região N-terminal da CDK11 sobre a expressão da proteína. Quanto à função destes motivos de poliglutamato, ela permanece sendo ignorada desde a descoberta da CDK11.

Tabela 5 – Sequências de DNA relacionadas à CDK9. A Tabela resume as sequências de DNA que são mencionadas no texto, indicando seu tamanho em pares de bases e a massa molecular do polipeptídeo codificado - quando for o caso - estimada pelo *ProtParam* (ExPASy). A proteína expressa a partir do pET28a::cdk9 corresponde, na realidade, à His-tag-CDK9. Contudo, passaremos a omitir a His-tag por conveniência algumas vezes, nos referindo à proteína resultante simplesmente como CDK9.

Sequência	Observação	DNA	Proteína
cdk9		1.119 pb	42,78 kDa
	+ Cauda de poli-histidina	+ 102 pb	+ 3,54 kDa
His-tag-cdk9		1.221 pb	46,32 kDa
pET28a		5.369 pb	-
	- Trecho entre <i>BamHI</i> e <i>EcoRI</i>	- 0 pb	-
pET28a- <i>[BamHI/EcoRI]</i>		5.369 pb	-
	+ cdk9	+ 1.119 pb	-
pET28a::cdk9		6.488 pb	-
pET23a		3.666 pb	-
	- Trecho entre <i>NdeI</i> e <i>BamHI</i>	- 33 pb	-
pET23a- <i>[NdeI/BamHI]</i>		3.633 pb	-
	+ cdk9	+ 1.119 pb	-
	- (sobreposição entre <i>NdeI</i> e ATG iniciação)	- 3 pb	-
pET23a::cdk9		4.749 pb	-

Além da His-tag propriamente dita, a cauda N-terminal introduzida na CDK9 pelo vetor pET28a inclui ainda um sítio de clivagem pela trombina e o epitopo (tag) T7 dentre os seus 34 resíduos de aminoácidos. Por sua vez, a proteína expressa a partir do plasmídeo pET23a::cdk9 corresponde à CDK9₄₂ humana original, sem nenhuma modificação (na realidade o pET23a também codifica uma His-tag C-terminal, mas a sequência nucleotídica inserida no vetor contém seu códon de terminação próprio). A presença da cauda de histidina completa fornece um incremento de 3,54 kDa na massa molecular da CDK9 codificada pelo pET28a::cdk9. Esta diferença de massa entre as versões com e sem His-tag da CDK9 pode ser facilmente visualizada nos géis SDS-PAGE, como mostra a Figura 28. A partir deste gel foi construída uma curva de calibração para os padrões de massa molecular, cuja regressão linear permitiu calcular a massa aparente das bandas referentes à CDK9 com o auxílio do software *Origin* 6.0 (Microcal). Os valores obtidos foram de $42,9 \pm 0,5$ kDa para a banda de superexpressão produzida pelo pET23a::cdk9 (CDK9 original) e $47,2 \pm 0,5$ kDa para a banda correspondente do pET28a::cdk9 (His-tag-CDK9). Existe assim uma boa concordância destes resultados com os valores teóricos indicados na Tabela 5.

Figura 28 – Diferença de massa molecular ocasionada pela introdução da cauda de histidina na CDK9. Este gel compara os perfis de expressão dos plasmídeos pET23a::cdk9 e pET28a::cdk9 em BL21(DE3), evidenciando a diferença na posição das bandas referentes à CDK9 devido à cauda de histidina introduzida pelo vetor pET28a. Para cada plasmídeo, na raia (I) foi aplicada uma amostra da cultura bacteriana imediatamente antes da adição do IPTG, enquanto a raia (II) corresponde à amostra coletada ao final da indução que empregou 0,2 mmol/L de IPTG a 30 °C durante 3:00 horas. As raia (III) e (IV) consistem, respectivamente, na fração sobrenadante (solúvel) e sedimentada (insolúvel) coletadas após a centrifugação do lisado bruto das células induzidas. A distância real entre as bandas de superexpressão observadas no gel foi de apenas 0,25 cm. Por meio de uma análise de regressão linear foram determinados os valores de massa molecular aparente da CDK9, resultando em 42,9 kDa na ausência e 47,2 kDa na presença da His-tag.

Um detalhe importante a ser observado na Figura 28 é que a banda correspondente à His-tag-CDK9 está situada abaixo do marcador de 45,0 kDa. Esta desconformidade nas posições relativas das duas bandas está presente em todos os géis SDS-PAGE que contêm a His-tag-CDK9 e empregam o marcador de massa molecular indicado. É preciso esclarecer que o problema com a localização da banda da His-tag-CDK9 se refere à sua divergência em relação ao valor estequiométrico teórico de 46,32 kDa, e não ao valor de 47,2 kDa estimado pela regressão linear. Já havia sido descrita uma ligeira diferença de 1,8 kDa entre os valores aparente e real para a massa molecular da CDK9 humana (PRICE, 2000). Todavia, a His-tag adicionada parece ter intensificado consideravelmente o efeito de migração anômala da

proteína no gel. Embora os fatores que levam à migração anômala de proteínas em SDS-PAGE não sejam conhecidos em profundidade, sabe-se que algumas vezes a adição ou remoção de um único resíduo de aminoácido pode gerar uma diferença de vários kilodaltons na massa molecular aparente da proteína (SHI et al., 2012).

O emprego de um novo plasmídeo para obtenção da CDK9 exige que as condições de expressão estabelecidas para o pET23a::cdk9 sejam revistas. A princípio não esperávamos observar diferenças significativas na expressão a partir do pET28a::cdk9. Neste sentido, a experiência prévia adquirida com o pET23a::cdk9 havia demonstrado uma relativa facilidade para expressão da CDK9 no sistema bacteriano, até mesmo sua indução espontânea em meio LB na ausência de IPTG. Esta observação é muito importante, uma vez que concentrações de IPTG maiores que 0,1 mmol/L ocasionaram a diminuição da quantidade de proteína solúvel (LEOPOLDINO et al., 2006). Para o novo plasmídeo pET28a::cdk9, entretanto, os resultados dos experimentos de indução foram ligeiramente diferentes. Como é possível constatar na Figura 29, não se observa nenhuma expressão da CDK9 na ausência de IPTG adicionado à cultura. Também nota-se que os níveis de expressão alcançados nestas condições são menores que aqueles obtidos anteriormente para o pET23a::cdk9, mesmo quando empregada uma temperatura mais alta durante a indução (35 °C).

A razão para a diferença de comportamento dos dois plasmídeos frente ao IPTG deve ser encontrada nas sequências regulatórias que controlam a expressão do inserto em cada vetor. A Figura 30 apresenta as regiões de clonagem e expressão dos vetores pET-28a(+) e pET-23a(+), destacando os principais elementos presentes na sequência nucleotídica. Os vetores do sistema pET (Novagen) utilizam o promotor do bacteriófago T7 para controlar a transcrição do gene inserido no seu sítio múltiplo de clonagem. Todavia, devemos notar que o promotor T7 no pET-28a(+) encontra-se associado ao operador *lac*, diferentemente do que ocorre no caso do pET-23a(+). Deste modo, a sequência introduzida no pET-23a(+) é transcrita livremente pela T7 RNA polimerase, enquanto no pET-28a(+) a transcrição é impedida na ausência de IPTG pela atuação da proteína repressora *lac* (*lacI*) ligada ao operador adjacente. No entanto, é preciso considerar ainda que o próprio gene da T7 RNA polimerase presente nas células da linhagem BL21(DE3) está sob controle do promotor-operador *lacUV5*, e assim também depende da adição de IPTG para liberar sua expressão.

Figura 29– Dependência do IPTG para expressão da proteína a partir do pET28a::cdk9. Neste experimento testamos a expressão da CDK9 na linhagem BL21(DE3) a 35 °C na presença e na ausência de IPTG. Ambas as culturas foram mantidas nas mesmas condições, com a diferença que em uma delas (representada à direita) foi adicionado IPTG suficiente para uma concentração final de 0,1 mmol/L quando a DO_{600} atingiu 0,5 unidades. As raias indicam o tempo decorrido após a adição do indutor na cultura correspondente. Observamos que a expressão da CDK9 ocorreu apenas na presença de IPTG, e parece ter atingido seu máximo depois de aproximadamente 3h de indução (outro experimento em condições idênticas mostrou que os níveis de expressão permanecem constantes neste ponto mesmo após 5h de indução).

Provavelmente, a expressão que observamos para o pET23a::cdk9 na ausência do indutor IPTG deve ser causada pelos níveis basais da T7 RNA polimerase nas células B21(DE3). A síntese de pequenas quantidades da T7 RNA polimerase pelas bactérias é inevitável, pois a eficiência da repressão transcricional exercida pelo sistema *lac* é sabidamente limitada. Como o promotor T7 do pET-23a(+) encontra-se desimpedido, uma única molécula da T7 RNA polimerase poderá transcrever diversas cópias do RNAm da CDK9, cada uma das quais será traduzida em diversas cópias do polipeptídeo. Assim, ocorre uma amplificação do efeito de “vazamento” (*leakage*) através do promotor-operador *lacUV5*, resultando em níveis de expressão basais para a CDK9 muito maiores que os da própria T7 RNA polimerase. No caso do pET-28a(+), porém, a presença de um segundo sistema repressor *lac* no vetor impede o acesso da polimerase ao promotor T7. Mesmo que a T7 RNA polimerase seja muito ativa e a eficiência do repressor *lac* limitada, os níveis basais de expressão da polimerase devem ser insuficientes para transcrever uma quantidade significativa de RNAm a partir de um promotor sob bloqueio da *lacI*.

Figura 30 – Comparação entre as regiões de clonagem e expressão dos vetores pET-28a(+) e pET-23a(+) (Novagen). Analisando os elementos presentes na sequência nucleotídica de cada um deles, fica evidente que os dois vetores são muito semelhantes. Em ambos os casos, notamos que a transcrição é controlada pelo promotor do bacteriófago T7. A principal diferença entre eles corresponde à ausência do operador *lac* na região adjacente ao sítio promotor no pET-23a(+). Por este motivo, o pET-23a(+) também não possui o gene que codifica a proteína repressora *lac* (*lacI*), ao contrário do pET-28a(+). Além disso, os dois vetores diferem quanto aos marcadores seletivos que carregam (genes de resistência a antibióticos), sendo que o pET-23a(+) confere resistência à ampicilina enquanto o pET28a(+) confere resistência à canamicina.

Fonte: Adaptado do Catálogo Novagen pET-28a-c(+) e pET-23a-d(+).

Uma consequência importante da mudança no vetor de expressão para o pET28a parece ter sido o agravamento da insolubilidade da CDK9 no sistema bacteriano. De fato, uma influência negativa na solubilidade de proteínas humanas expressas em *E. coli* já foi correlacionada com a utilização de caudas de histidina N- e C-terminais (WOESTENENK et al., 2004). O elevado grau de insolubilidade pode ser constatado no gel da Figura 31, que demonstra também uma característica recorrente nos testes de expressão: A quantidade

muito menor da proteína observada na raia em que foi aplicada uma amostra da cultura pós-indução (II) em comparação com a raia que contém a fração insolúvel do lisado (IV). Esta observação não poderia ser uma consequência da diluição das amostras, pois neste caso a fração insolúvel foi ressuspensa em uma quantidade de tampão proporcional àquela da cultura induzida para aplicação das amostras no gel. Em seguida, foi percebido que o aquecimento prolongado da amostra a 95 °C no tampão SDS-PAGE leva ao ressurgimento da proteína na raia que corresponde à cultura induzida. A sonicação da amostra e a adição de ureia 8,0 mol/L no tampão produziram o mesmo efeito de realçar a banda, tornando-a equivalente àquela presente na fração insolúvel do lisado. Portanto, apenas o tratamento térmico ou mecânico drástico e a adição de grandes quantidades de agentes caotrópicos foram capazes de solubilizar a His-tag-CDK9 no tampão de amostra SDS-PAGE, promovendo o aparecimento da banda correspondente no gel.

Figura 31 – Expressão da His-tag-CDK9 predominantemente na fração insolúvel do lisado bacteriano.

Foram aplicadas no gel de poliacrilamida as amostras referentes a um experimento de indução da His-tag-CDK9 na cepa BL21(DE3), com 0,6 mmol/L de IPTG a 30 °C durante 5h20min. A raia (I) corresponde à amostra não induzida (retirada da cultura antes da adição do IPTG), e a raia (II) à amostra induzida (coletada no final do processo para avaliar a expressão). A raia (III) corresponde à fração sobrenadante (solúvel) do lisado bacteriano bruto, e a raia (IV) à sua fração sedimentada (insolúvel). O resultado demonstra uma tendência muito acentuada de localização da proteína na fração insolúvel do lisado. Também se evidencia neste gel a discrepância entre a quantidade de His-tag-CDK9 presente nas raia (II) e (IV), que é discutida no texto.

A insolubilidade acentuada da His-tag-CDK9 no sistema bacteriano, resistente ao tratamento brando com desnaturantes, indica que a proteína expressa está sendo depositada na forma de corpos de inclusão. Os corpos de inclusão são agregados proteicos refratários, comumente gerados pelas células bacterianas em resposta à superexpressão ectópica de proteínas (CARRIÓ et al., 2001). A formação de corpos de inclusão é um dos problemas mais frequentes da expressão heteróloga em sistemas bacterianos. Uma grande variedade de fatores pode influenciar o grau de solubilidade das proteínas expressas pela bactéria e sua tendência a formar corpos de inclusão. Consequentemente, diversas são as estratégias utilizadas para tentar reverter a insolubilidade de uma proteína ectopicamente expressa no sistema bacteriano. A primeira alternativa consiste em ajustar as variáveis do processo de indução – concentração de IPTG, temperatura e duração – e a composição do meio de cultura (SAHDEV et al., 2008).

Os resultados obtidos pela variação das condições de indução utilizando temperaturas mais baixas (até 20 °C) e menores concentrações de IPTG (até 0,05 mmol/L) não foram satisfatórios, pois levaram à diminuição da quantidade de proteína expressa sem aumentar perceptivelmente a sua ocorrência na fração solúvel. O próximo passo foi trocar o meio LB (*Lysogen Broth*) por um meio de composição semelhante ao SOB (*Super Optimal Broth*), mais conveniente para a superexpressão de proteínas por ser mais rico em nutrientes, principalmente aminoácidos. Além de possuir o dobro do conteúdo de triptona do meio LB, o meio de indução alternativo também inclui 15 mmol/L de MgCl₂, visto que o cátion magnésio é um co-fator essencial para a atividade enzimática da CDK9. Este também passou a ser tamponado com 50 mmol/L de tris, para evitar a queda no pH inerente ao processo de indução, a qual representa um importante sinal de deterioração do meio. A tentativa de mudança do meio de cultura, no entanto, não foi capaz de incrementar a solubilidade da proteína no sistema heterólogo. Apesar disso, o meio enriquecido aumentou significativamente a quantidade de sedimento bacteriano obtido – de menos de 4,0 g no meio LB para mais de 10 g por litro de cultura induzida – sem que fosse modificado o valor da densidade óptica no princípio da indução. Devido ao seu potencial de aumentar o rendimento na indução e favorecer a obtenção de proteína solúvel em combinação com outros métodos, o novo meio de cultura foi mantido.

Quando o simples ajuste das variáveis nos experimentos de indução não pode reverter a insolubilidade da proteína de interesse, a alternativa seguinte consiste na utilização de cepas de expressão que contornam problemas específicos do sistema heterólogo associados à formação de corpos de inclusão. Uma vez que a cepa pLysE se mostrou eficaz para promover a solubilização da CDK9 expressa a partir do pET23a::cdk9 (LEOPOLDINO et al., 2006), a primeira opção foi testar a expressão do pET28a::cdk9 nas cepas pLysE e pLysS. Estas linhagens derivadas da BL21(DE3) expressam a T7 lisozima, uma proteína do bacteriófago T7 que inibe naturalmente a T7 RNA polimerase, reduzindo a taxa de transcrição de genes sob controle de um promotor T7. Deste modo, o efeito da T7 lisozima é semelhante à diminuição da quantidade de IPTG adicionada ao meio ou, em menor grau, ao abaixamento da temperatura de indução. A diferença entre as cepas pLysE e pLysS está na quantidade de T7 lisozima expressa pela célula, que é substancialmente maior no caso da pLysE. Além de funcionar como um regulador negativo da T7 RNA polimerase, a T7 lisozima também atua enfraquecendo a parede celular bacteriana e facilitando o procedimento de lise.

Outra possibilidade considerada para solubilização da His-tag-CDK9 foi o emprego da cepa de expressão pT-GroE. A pT-GroE é uma linhagem derivada da BL21(DE3) que expressa a chaperona GroEL e sua co-chaperona GroES a partir de um plasmídeo contendo o operon *groE* controlado pelo promotor T7. O complexo GroESL é essencial para o enovelamento de muitas proteínas bacterianas, especialmente em condições de stress conformacional, como no caso de choque térmico ou superexpressão ectópica de proteínas na célula. GroEL e GroES são funcional e estruturalmente muito semelhantes às chaperonas eucarióticas Hsp60 e Hsp10, respectivamente. A co-expressão de GroEL e GroES com proteínas heterólogas é capaz de promover em muitos casos a sua transferência para a fração solúvel no sistema bacteriano (YASUKAWA et al., 1995). Foi mencionado na introdução que a CDK9 depende de uma via de enovelamento assistido da qual participam as chaperonas Hsp70 e Hsp90/Cdc37 para assumir sua estrutura nativa (O'KEEFFE et al., 2000). Entretanto, é importante esclarecer que estas chaperonas que auxiliam o enovelamento da CDK9 na célula em condições normais pertencem a classes totalmente distintas em relação às chaperoninas GroESL e Hsp60/Hsp10.

Os testes realizados com as novas cepas de expressão, porém, não permitiram avaliar diretamente o nível de sucesso na obtenção de proteína solúvel. A causa deste problema está na inconveniente sobreposição entre as bandas da His-tag-CDK9 e do fator de alongação ribossomal EF-Tu, que possui massa molecular de 43,2 kDa. O fator EF-Tu é a proteína citosólica mais abundante em *E. coli*, podendo representar até 10% do total de proteínas da célula (WEIJLAND et al., 1992). A banda correspondente ao EF-Tu é bastante evidente no perfil de expressão da bactéria, e se localiza pouco abaixo do marcador de 45,0 kDa nos diversos géis apresentados que mostram a expressão da CDK9 e CDK11. A Figura 31 fornece um exemplo nítido, onde é possível reconhecer a banda do EF-Tu na raia (I) e a banda da His-tag-CDK9 surgindo logo abaixo dela, na raia (II), após a indução. Devido à sua estreita proximidade, a banda do EF-Tu frequentemente se sobrepõe à banda da His-tag-CDK9, dificultando a avaliação dos níveis de expressão da quinase (principalmente para as amostras da fração solúvel). Ainda, como o EF-Tu constitui um fator de tradução, é natural que ele próprio seja induzido em situações de superexpressão de proteínas, confundindo-se com a indução da His-tag-CDK9.

Para confirmar a expressão da His-tag-CDK9 nas amostras induzidas e a sua presença na fração solúvel do lisado, foi realizado um *Western blot* empregando um anticorpo que reconhece a cauda de histidina, cujo resultado é apresentado na Figura 32. Infelizmente, o experimento revelou que a proteína não é expressa pelas cepas pLysE e pLysS, além de estar definitivamente ausente na fração solúvel em BL21(DE3). É importante destacar que, neste caso, foram utilizadas condições brandas na indução em BL21(DE3) para favorecer a solubilidade da proteína, e condições mais drásticas em pLysS e pLysE para maximizar os níveis de expressão. A razão presumida para a dificuldade de expressão nas cepas pLysE e pLysS é que a inibição da T7 RNA polimerase ocorreu sobre níveis de atividade já reduzidos da enzima, suprimindo totalmente a transcrição a partir do promotor T7. Em relação à cepa pT-GroE, a expressão da proteína já havia sido descartada por diversas tentativas malsucedidas de se obter a sua ligação a uma resina *IMAC* de níquel. É provável que a superexpressão de GroEL e GroES tenha sobrecarregado a T7 RNA polimerase e a maquinaria de tradução bacteriana, inviabilizando a síntese de His-tag-CDK9 em uma competição desigual pelos recursos da célula.

Figura 32 – Resultado do *Western Blot* para avaliar a expressão e solubilidade da His-tag-CDK9. A figura mostra o resultado do experimento de imunoblot utilizando um anticorpo primário que reconhece a cauda de histidina. Na raia (I) foi aplicada a cultura induzida em BL21(DE3) com 0,2 mmol/L de IPTG a 30 °C por 3h, e na raia (II) a correspondente fração sobrenadante do lisado. Nas raias (III) e (IV) foram aplicadas as culturas induzidas em pLysS e pLysE, respectivamente, com 0,6 mmol/L de IPTG a 37 °C por 5h. A raia (V) foi pulada, e na raia (VI) foi aplicada uma amostra de uma proteína genérica (*SUMO* protease) contendo His-tag, para servir como controle positivo. O resultado demonstra que apenas a BL21(DE3) foi capaz de expressar a His-tag-CDK9 e que a proteína se particionou completamente na fração insolúvel do lisado.

O resultado do imunoblot remove a esperança de se obter a proteína na sua forma solúvel e enzimaticamente ativa através da expressão do plasmídeo pET28a::cdk9 associada a métodos de purificação em condições não desnaturantes. Tendo em vista as possibilidades já exploradas, as únicas alternativas restantes para obtenção da proteína diretamente em uma forma solúvel consistiriam em modificações na sua sequência (como truncamentos da cadeia e passagem da His-tag para o C-terminal), substituição do vetor de expressão, ou a mudança definitiva para um sistema eucariótico. Porém, tendo em vista a quantidade de trabalho envolvida nestas operações, o tempo disponível, e a probabilidade de sucesso, decidimos que seria mais conveniente realizar algumas tentativas de recuperação da proteína a partir dos corpos de inclusão expressos em BL21(DE3). Existem até certas vantagens na obtenção de proteínas de corpos de inclusão, como os níveis abundantes de expressão e a facilidade de isolamento do produto insolúvel com elevada pureza.

5.2.3 Purificação e reenovelamento da CDK9

Por diversas razões, o emprego de caudas de histidina N-terminais é aconselhável como primeira escolha na obtenção de proteínas recombinantes para aplicações em escala laboratorial. Estas incluem a eficácia do método de purificação (*IMAC*), e a compatibilidade da sequência N-terminal inserida com altas taxas de transcrição, tradução e estabilidade da proteína (GRÄSLUND et al., 2008). No presente estudo, porém, a introdução de uma His-tag N-terminal na CDK9 humana foi relacionada a um severo comprometimento da solubilidade da proteína expressa pelo sistema bacteriano. Existe um número muito pequeno de referências que mencionam especificamente a utilização de His-tags para a CDK9. Por exemplo, a expressão da CDK9 de *Drosophila* contendo uma His₆-tag C-terminal em células de inseto Sf9, e sua subsequente purificação por cromatografia *IMAC* em resina de níquel, já foi realizada com sucesso (PENG et al., 1998b). Outros autores relataram a purificação do complexo P-TEFb ativo pela co-expressão da ciclina T1 e His₆-tag-CDK9 em células Sf9, embora neste caso a posição da cauda de histidina introduzida na proteína não tenha sido especificada (PINHEIRO et al., 2004).

Também há muito pouca informação disponível a respeito da expressão heteróloga da CDK9. Antes do trabalho que descreveu a obtenção da CDK9 humana biologicamente ativa a partir do sistema bacteriano (LEOPOLDINO et al., 2006), a expressão da CDK9 de *Drosophila* já havia sido realizada em *E. coli*, mas a proteína obtida se mostrou insolúvel e foi utilizada apenas para a geração de anticorpos (ZHU et al., 1997). Em outro estudo, a CDK9 humana fusionada a uma cauda GST foi expressa na forma solúvel pela bactéria, mas sua atividade enzimática não foi testada, pois o objetivo era apenas avaliar o potencial da CDK9 como substrato da histona acetiltransferase p300HAT (FU et al., 2007). Uma observação importante deste trabalho é a co-purificação de uma forma N-terminal truncada da GST-CDK9, que foi atribuída à degradação espontânea da proteína no sistema bacteriano. Todavia, dentre todas as referências consultadas não há nenhum exemplo que relate a expressão da CDK9 fusionada a uma cauda de histidina em *E. coli*.

Antes que o experimento de imunoblot revelasse a total ausência da His-tag-CDK9 na fração solúvel, foram realizadas algumas tentativas de purificação da proteína utilizando uma resina *IMAC* de níquel. Conforme os resultados da cromatografia se mostravam

Uma vez que o *Western blot* não detectou nenhuma proteína contendo His-tag na fração solúvel do lisado, as bandas isoladas destas amostras pela cromatografia *IMAC* não podem corresponder à CDK9. Mesmo assim, a banda observada por volta de 60 kDa esteve presente em praticamente todas as aplicações da fração sobrenadante na coluna de níquel. É sabido que algumas proteínas endógenas de *E. coli* são muitas vezes co-purificadas com proteínas recombinantes contendo caudas de histidina, por interagirem especificamente com a resina *IMAC* (GRÄSLUND et al., 2008). A GroEL é uma das principais proteínas de *E. coli* co-purificadas por *IMAC* e, devido à sua massa molecular de 57,3 kDa, é a candidata mais provável para justificar a banda de ~60 kDa. Curiosamente, a GroEL possui uma prevalência muito pequena de resíduos de histidina (0,2%) em relação às demais proteínas bacterianas que são contaminantes frequentes na purificação por *IMAC*. Foi determinado que a GroEL também representa um subproduto comum da purificação por cromatografia de afinidade em resinas de glutationa-sefarose, o que sugere que a proteína interage com o componente polimérico da resina *IMAC* e não com o cátion metálico (BOLANOS-GARCIA et al., 2006).

Quanto à outra banda purificada na Figura 33, é possível que se trate da DnaK (MM 69,1 kDa), uma chaperona homóloga à Hsp70 eucariótica que também figura entre os principais contaminantes da purificação por *IMAC*. Um detalhe importante das tentativas de purificação da proteína nativa é que, na ausência do coquetel de inibidores de proteases, um número muito maior de bandas se liga à resina *IMAC*. Estas bandas apresentam massa molecular menor que a His-tag-CDK9, e provavelmente correspondem a fragmentos N-terminais gerados durante a lise celular, de modo semelhante ao caso relatado da GST-CDK9 (FU et al., 2007). Os produtos de degradação proteolítica não puderam ser detectados no *Western blot* porque a amostra aplicada no ensaio foi lisada na presença do coquetel de inibidores. Assim, as tentativas de purificação da proteína a partir da fração solúvel do lisado em condições nativas foram definitivamente abandonadas.

A His-tag-CDK9 só pôde ser purificada com sucesso empregando condições desnaturantes, como mostrado na Figura 34. Neste caso, a lise foi realizada em tampão fosfato acrescido de 8,0 mol/L de ureia, para permitir que a proteína contida nos corpos de inclusão fosse solubilizada. Em condições desnaturantes, a ligação de proteínas inespecíficas à resina *IMAC* é grandemente reduzida, normalmente resultando na purificação completa

dos alvos que contêm a cauda de histidina em uma única etapa. Outra vantagem da utilização de condições desnaturantes é a possibilidade de se utilizar o tampão fosfato, o mais indicado para *IMAC*, uma vez que a forte interação do ânion fosfato com o cátion magnésio – *Kps* do $\text{Mg}_3(\text{PO}_4)_2$ é da ordem de 10^{-24} – inviabiliza o seu emprego na purificação da CDK9 nativa. A quantificação das frações purificadas demonstrou que a His-tag-CDK9 é adsorvida pela resina *IMAC* próximo do limite de retenção nominal, de 15 mg de proteína por mL de resina. Contudo, ao realizar a diafiltração da amostra em um concentrador para remoção do desnaturante, a proteína precipitou quando a concentração de ureia ficou abaixo de 4,0 mol/L.

Figura 34 – Perfil das amostras eluídas na cromatografia de afinidade ao Ni^{2+} (*IMAC*) em condições desnaturantes. Neste experimento, células BL21 (DE3) contendo o plasmídeo pET28a::cdk9 foram induzidas com 0,2 mmol/L de IPTG a 30 °C por 4,5 horas. Após a lise das bactérias em tampão fosfato (pH 7,5) contendo 8,0 mol/L de ureia, o extrato bruto de *E. coli* foi incubado com a resina de níquel sob agitação por 4 horas à temperatura ambiente. Em seguida, foi realizada a eluição das amostras na ordem indicada na Figura. A primeira fração, E_0 , corresponde à amostra eluída inicialmente (*flowthrough*). As frações L_1 , L_2 e L_3 correspondem a lavagens consecutivas da resina com um tampão contendo 20 mmol/L de imidazol. Por fim, as frações I_1 a I_4 foram eluídas com concentrações crescentes de imidazol (50, 100, 250 e 500 mmol/L) em volumes individuais de 2,0 mL. Observa-se que a separação foi bastante eficiente, produzindo amostras concentradas da His-tag-CDK9 quase pura. A quantificação da mistura (*pool*) das frações eluídas com imidazol forneceu cerca de 3,0 mg de proteína por mL.

Embora a purificação em condições desnaturantes seja vantajosa em termos de eficiência e rendimento, a perda total de atividade biológica da proteína pela desnaturação limita muito o seu emprego. Modernamente, proteínas expressas como corpos de inclusão em *E. coli* têm encontrado vasta aplicação industrial, graças ao progresso nas técnicas de reenovelamento *in vitro* (SAHDEV et al., 2008). No entanto, as dificuldades normalmente enfrentadas para se desenvolver protocolos específicos de reenovelamento levam à sua indicação apenas quando todas as chances de expressão da proteína solúvel foram esgotadas (GRÄSLUND et al., 2008). Como este provou ser o caso da His-tag-CDK9, era muito importante saber se a proteína seria capaz de recuperar em algum grau a sua atividade enzimática após a completa desnaturação. Por se tratar de uma proteína contendo His-tag, a alternativa mais óbvia seria o reenovelamento assistido por matriz, no qual o agente desnaturante é removido enquanto a proteína permanece ligada à resina cromatográfica. A interação estabilizadora da proteína com a matriz e a separação espacial dos intermediários de reenovelamento favorecem bastante a recuperação da conformação ativa do polipeptídeo, evitando a formação de agregados (SAHDEV et al., 2008). Porém, a eficiência do processo depende da remoção lenta do desnaturante, e a aplicação de um gradiente de 8,0 a 0 mol/L de ureia para reenovelamento de uma proteína ligada à resina *IMAC* deve ser realizada ao longo de muitas horas (REHM et al., 2001).

Assim, pelo fato de utilizarmos um sistema cromatográfico operado manualmente, o reenovelamento assistido por matriz seria de difícil execução. Uma alternativa mais simples e viável poderia ser a recuperação da conformação ativa da proteína a partir do estado desnaturado por SDS em vez de ureia. De fato, a desnaturação de proteínas por SDS apresenta várias diferenças em relação à ureia, a começar pela concentração necessária do agente desnaturante – que é milimolar para o SDS, mas pode chegar a 8,0 mol/L para a ureia. O reenovelamento de uma proteína desnaturada por SDS pode ser grandemente facilitado na presença de determinados co-solventes anfipáticos, como o 2-metil-2,4-pentanodiol (MPD), que diminuem muito a capacidade do SDS de interagir com a cadeia polipeptídica (MICHAX et al., 2008; BOISSELIER et al., 2011). Para promover o reenovelamento, simplesmente adiciona-se o MPD a uma amostra da proteína desnaturada em SDS e a solução é incubada por 24 horas a 25 °C. Entretanto, uma etapa prévia de cromatografia *IMAC* continua sendo necessária para purificar a His-tag-CDK9 e remover a

ureia 8,0 mol/L empregada na solubilização dos corpos de inclusão. Após a ligação da proteína à resina em um tampão contendo 8,0 mol/L de ureia e 0,5% de SDS, foi aplicado um gradiente negativo de 8,0 a 0 mol/L de ureia para remover o agente caotrópico do meio. Em seguida, um gradiente positivo de 15 a 500 mmol/L de imidazol foi empregado para eluir a His-tag-CDK9 retida na coluna cromatográfica. Apesar da remoção da ureia, a proteína continua desnaturada e mantém a solubilidade no tampão devido à presença do SDS 0,5%. O ensaio de reenovelamento com MPD utilizou esta amostra de His-tag-CDK9 purificada em condições desnaturantes, e o produto foi submetido a um teste de atividade enzimática cujo resultado é apresentado na Figura 35.

Figura 35 – Resultado do teste de atividade com a resina de ATP. A amostra obtida ao final do experimento de reenovelamento foi testada quanto à sua capacidade de interagir com uma resina de afinidade ao ATP. A primeira raia, A_0 , corresponde à amostra que foi aplicada na resina em tampão Hepes (pH 7,4) na presença de 10 mmol/L de $MgCl_2$. A raia seguinte, E_0 , contém a fração eluída diretamente (*flowthrough*) ao final de 24 horas de incubação da amostra a 4 °C. Após três lavagens consecutivas com o próprio tampão de ligação (L_1 a L_3), a última raia (A_1) representa a proteína adsorvida pela resina e eluída com 50 mmol/L de ATP. A quantificação das amostras A_0 e A_1 obteve as concentrações de 0,9 e 0,4 mg/mL, respectivamente. Entretanto, é preciso considerar que a amostra A_0 possuía um volume total de 5,0 mL, enquanto a amostra A_1 foi eluída com apenas 1,0 mL da solução de ATP, o que fornece um rendimento aproximado de 9,0% para o processo de reenovelamento *in vitro* da proteína.

A ligação de uma fração do produto de reenovelamento à resina de ATP demonstra que é possível obter a renaturação parcial da proteína a partir do estado desnaturado por SDS. Embora o rendimento do processo seja pequeno (inferior a 10%), ele indica uma relativa facilidade de recuperação da conformação ativa, e pode ser contrabalançado pelos altos níveis de expressão da proteína na forma insolúvel. Alguns fatores devem ter colaborado muito para a eficácia do reenovelamento *in vitro*, como a ausência de ligações dissulfeto nativas na CDK9, o que poupa a necessidade de adicionar e remover um agente redutor de cistinas durante o processo. Também deve ser importante o fato de a CDK9 apresentar um elevado *aliphatic index* – estimado em 90,16 pela ferramenta *ProtParam* (ExPASy) – índice que mede a estabilidade térmica de proteínas globulares e correlaciona positivamente com a espontaneidade do enovelamento (IDICULA-THOMAS et al., 2005). Ainda, o valor próximo de 9,0 para o ponto isoelétrico da CDK9 implica em um excesso considerável de cargas positivas ao longo da molécula, o que presumivelmente dificultaria a formação de agregados porque a repulsão eletrostática das cadeias polipeptídicas favorece processos intramoleculares (que possuem cinética de primeira ordem), como o reenovelamento.

O papel de substâncias como o MPD no processo de reenovelamento decorre essencialmente do seu modo de interação com o desnaturante SDS e não com a proteína. Tanto o MPD quanto as cadeias polipeptídicas desnaturadas parecem interagir preferencialmente com a superfície das micelas de SDS, que contém os grupos-cabeça sulfato. Assim, o efeito do MPD deve ser ao menos em parte devido à competição por sítios de ligação na superfície das micelas de SDS (MICHAUX et al., 2008). Pelo fato de não influenciar diretamente a estabilidade conformacional da proteína, o MPD não pode ser considerado um exemplo de chaperona química, ou seja, uma substância que promove de modo genérico a interação favorável de polipeptídeos com a água (PAPP et al., 2006). No entanto, o tampão de reenovelamento contém 10% de glicerol – a chaperona química mais comumente empregada – o que certamente deve ter contribuído para o êxito na renaturação parcial da proteína. A presença de 10 mmol/L de $MgCl_2$ também influencia no reenovelamento, uma vez que o cátion Mg^{2+} se liga à conformação nativa da proteína e representa um co-fator essencial para a atividade enzimática da quinase. Porém, a necessidade de adicionar íons magnésio ao meio impede a utilização de tampão fosfato na

etapa anterior, de purificação cromatográfica por *IMAC*. O emprego do tampão Hepes 30 mM contendo SDS 0,5% prejudicou seriamente a ligação da proteína à resina de níquel, fornecendo apenas 4,5 mg de His-tag-CDK9 purificada (o que equivale a 19% dos 24 mg obtidos na aplicação em tampão fosfato que é mostrada na Figura 34). Esta perda de eficiência na purificação compromete muito o rendimento global no processo de obtenção da proteína ativa.

A fração proteica retida pela resina de afinidade ao ATP foi capaz de recuperar a sua conformação nativa a partir do estado completamente desnaturado. Embora esta amostra seja referida como proteína ativa, é preciso lembrar que, a rigor, a forma biologicamente ativa da CDK9 é o P-TEFb, pois a atividade enzimática da quinase depende essencialmente da sua ligação à ciclina. Foi explicado na Introdução que a associação à ciclina e a fosforilação do segmento de ativação (alça T) das CDKs são responsáveis por modificar a conformação local do sítio ativo para permitir a ligação específica do substrato peptídico (BAUMLI et al., 2008). Desta maneira, mesmo tendo alcançado a conformação nativa, a CDK9 produzida pelo sistema bacteriano não possui atividade enzimática propriamente dita, porque não interage com os substratos da quinase na ausência de uma ciclina ativadora. Entretanto, a capacidade de interagir com o ATP evidentemente não é condicionada pela ativação completa da CDK. O fato de o ATP possuir um sítio de ligação profundo – ao contrário da ligação superficial do substrato peptídico – deve ter relação com a maior afinidade da enzima pelo ATP. Neste sentido, um estudo recente elucidou o mecanismo enzimático da CDK9 e descobriu a ligação ordenada dos seus substratos ao sítio ativo (BAUMLI et al., 2012). O modelo proposto para o ciclo catalítico da CDK9 está ilustrado na Figura 36.

A primeira pista sobre o mecanismo de associação sequencial dos substratos à CDK9 foi a observação de que a remoção do segmento C-terminal (330-372), que não pertence ao domínio quinase, faz com que a atividade enzimática da proteína seja reduzida a cerca de 30% da atividade da CDK9 contendo a sequência completa de aminoácidos. Estudos comparativos de cinética enzimática investigaram o motivo da diminuição na atividade catalítica da CDK9 pela perda de sua cauda C-terminal. Empregando uma concentração fixada para o segundo substrato, foram determinados valores aparentes para a constante de Michaelis (K_{Map}) em relação aos dois substratos da quinase, ATP e GST-CTD, para as versões

original e truncada da enzima (BAUMLI et al., 2012). O resultado revelou que o valor de K_{Map} referente ao substrato peptídico GST-CTD não é alterado pelo encurtamento da CDK9. Por outro lado, o K_{Map} relativo ao ATP apresenta um valor bem mais elevado para a versão truncada em comparação com a CDK9 original. Isto significa que a afinidade da CDK9 pelo ATP é bastante diminuída na ausência do segmento C-terminal. Em seguida, foi realizado um experimento de inibição por baixas concentrações de ADP para examinar a ordem de ligação e liberação dos substratos no mecanismo enzimático. Se a ligação dos substratos à CDK9 ocorresse em uma ordem aleatória, o ADP deveria agir como um inibidor competitivo com respeito a ambos. Porém, se a ligação dos substratos procedesse de forma ordenada, o ADP não deveria competir com ambos, produzindo diferentes modalidades de inibição relativas a cada substrato (BAUMLI et al., 2012).

Figura 36 – Ciclo catalítico da CDK9. A cauda C-terminal da CDK9 pode assumir uma conformação na qual se dobra sobre o sítio ativo da quinase, embora não deva realizar contatos diretos com os substratos. Esta conformação fechada - que denotamos por CDK9* - é induzida pela ligação do ATP ao seu sítio. Com a ligação do substrato peptídico, o acesso ao sítio de ligação do ATP é completamente obstruído, impedindo a efetuação de trocas ATP/ADP com o meio. Desta forma, o mecanismo catalítico exige que o ATP seja o primeiro substrato a se ligar (I) e o ADP o segundo produto a ser liberado da enzima (V). O modo ordenado de associação dos ligantes permite que a CDK9 interaja com o ATP na ausência do substrato peptídico fosforilável, representado pelo CTD neste diagrama.

Fonte: Adaptado de Baumli et al. (2012, p. 1793).

De fato, os resultados da inibição da CDK9 revelaram que o ADP se comporta como um inibidor competitivo apenas em relação ao ATP, o que sugere um modelo de associação ordenada no qual o ATP é o primeiro substrato a se ligar e o ADP o segundo produto a ser liberado pela enzima. Porém, este ordenamento não ocorre para a reação catalisada pela forma truncada da CDK9, pois neste caso o ADP mostrou atuar como um inibidor competitivo em relação a ambos os substratos. Assim, ficou demonstrado que a cauda C-terminal da CDK9 não só contribui para a sua afinidade de ligação ao ATP, como também é responsável por estabelecer um mecanismo de associação ordenada dos substratos à enzima (BAUMLI et al., 2012). A ligação da CDK9 à resina de ATP somente é possível em razão deste mecanismo particular de associação dos ligantes, no qual a interação com o ATP não requer a ativação completa da quinase e independe da presença de um substrato peptídico.

A renaturação parcial da His-tag-CDK9 mostrou que a introdução da cauda de histidina não prejudica a atividade da proteína recombinante da mesma maneira que a sua solubilidade na expressão heteróloga. O reenovelamento *in vitro* do produto fornece um novo método para obtenção da CDK9 humana ativa a partir do sistema bacteriano. Embora o método original de Leopoldino et al. (2006) de expressão da CDK9 solúvel e purificação em condições nativas ofereça um rendimento cerca de 25 vezes maior, os resultados deste trabalho são preliminares e certamente existem muitas variáveis experimentais que podem ser otimizadas no processo. Até onde sabemos, a capacidade de renaturação da CDK9 humana nunca havia sido demonstrada por outros autores.

6 CONCLUSÃO

A obtenção de quinases humanas ativas a partir de um sistema bacteriano é dificultada por uma série de obstáculos comumente relacionados à expressão de proteínas em sistemas heterólogos, os quais refletem fundamentalmente as diferenças entre células eucarióticas e procarióticas. No caso da CDK9 humana, os níveis de expressão da proteína se mostraram satisfatórios desde o início, mas a sua acentuada insolubilidade, agravada pela introdução da cauda de histidina, não pôde ser revertida por nenhuma estratégia convencional. Assim, a alternativa encontrada foi a solubilização e purificação da His-tag-CDK9 em condições desnaturantes. Um ensaio de reenovelamento *in vitro* mostrou que é possível a renaturação da quinase a partir do estado desnaturado por SDS. Embora o rendimento obtido de proteína ativa ainda seja muito pequeno em comparação com o método original de Leopoldino et al. (2006), ele representa a primeira demonstração da capacidade de renaturação da CDK9 humana.

Para a CDK11 humana foi observada uma séria dificuldade de expressão no sistema bacteriano. Mesmo com o emprego de condições muito favoráveis para a indução, nenhuma expressão da proteína pela bactéria pôde ser detectada. Foi proposta então a tentativa de expressar uma versão encurtada da proteína que contivesse apenas o seu domínio quinase. Após a modificação da sequência contida no vetor, a proteína truncada, designada CDK11(Δ), foi prontamente expressa pelo sistema bacteriano. Para investigar o motivo pelo qual a sequência removida inibe severamente a expressão da proteína, alguns experimentos utilizaram cepas enriquecidas com genes de RNAs transportadores que decodificam códons raros. Porém, a hipótese de que a prevalência de códons raros no trecho eliminado é o principal fator responsável pelo efeito negativo sobre a expressão da proteína foi descartada, indicando que o fenômeno deve estar relacionado às repetições em série de glutamato na porção N-terminal da CDK11. A expressão da CDK11(Δ) pelo sistema bacteriano assume especial interesse porque não são conhecidos relatos prévios de obtenção da CDK11 humana em *E. coli*.

Assim, embora nossos resultados sejam preliminares e necessitem de mais dados empíricos para sua confirmação, eles fornecem exemplos de como é possível contornar certas adversidades da expressão heteróloga de modo a obter CDKs humanas a partir do sistema bacteriano. Os problemas superados indicam que o aperfeiçoamento das condições experimentais deverá em breve tornar viável a obtenção das CDKs humanas em quantidade e pureza suficientes para a realização de estudos estruturais.

REFERÊNCIAS BIBLIOGRÁFICAS

- ADAMS, J. A. Kinetic and catalytic mechanisms of protein kinases. **Chemical Reviews**, v. 101, n. 8, p. 2271-2290, 2001.
- AHN, S. H.; KIM, M.; BURATOWSKI, S. Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. **Molecular Cell**, v. 13, p. 67-76, 2004.
- ALBERTS, B.; JOHNSON, A.; LEWIS, J.; RAFF, M.; ROBERTS, K.; WALTER, P. **Biologia molecular da célula**. Porto Alegre: Artmed, 2010. 1396 p.
- AMMOSOVA, T.; OBUKHOV, Y.; KOTELKIN, A.; BREUER, D.; BEULLENS, M.; GORDEUK, V. R.; BOLLEN, M.; NEKHAI, S. Protein phosphatase-1 activates CDK9 by dephosphorylating Ser175. **PLoS ONE**, v. 6, n. 4, 2011. Disponível em: <<http://www.plosone.org/article/fetchObject.action?uri=info%3Adoi%2F10.1371%2Fjournal.pone.0018985&representation=PDF>>. Acesso em: 18 jul. 2014.
- AMMOSOVA, T.; WASHINGTON, K.; DEBEDE, Z.; BRADY, J.; NEKHAI, S. Dephosphorylation of CDK9 by protein phosphatase 2A and protein phosphatase-1 in Tat-activated HIV-1 transcription. **Retrovirology**, v. 2, n. 47, 2005a. Disponível em: <<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1187922/pdf/1742-4690-2-47.pdf>>. Acesso em: 18 jul. 2014.
- AMMOSOVA, T.; BERRO, R.; KASHANCHI, F.; NEKHAI, S. RNA interference directed to CDK2 inhibits HIV-1 transcription. **Virology**, n. 341 p. 171-178, 2005b.
- ARIZA, M. E.; BROOME-POWELL, J.; LAHTI, M.; KIDD, V. J.; NELSON, M. A. Fas-induced apoptosis in human malignant melanoma cell lines is associated with the activation of p34^{cdc2}-related PITSLRE protein kinases. **The Journal of Biological Chemistry**, v. 274, n. 40, p. 28505-28513, 1999.
- AZEVEDO JR, W. F.; CANDURI, F.; SILVEIRA, N. J. F. Structural basis for inhibition of cyclin-dependent kinase 9 by flavopiridol. **Biochemical and Biophysical Research Communications**, v. 293, p. 566-571, 2002.
- BAEK, K.; BROWN, R. S.; BIRrane, G.; LADIAS, J. A. Crystal structure of human cyclin K, a positive regulator of cyclin-dependent kinase 9. **Journal of Molecular Biology**, n. 366, v. 2, p. 563-573, 2007.
- BARBORIC, M.; ZHANG, F.; BESENICAR, M.; PLEMENITAS, A.; PETERLIN, B. M. Ubiquitylation of Cdk9 by Skp2 facilitates optimal Tat transactivation. **Journal of Virology**, v. 79, n. 17, p. 11135-11141, 2005.
- BARRÔCO, R. M.; DE VEYLDER, L.; MAGYAR, Z.; ENGLER, G.; INZÉ, D.; MIRONOV, V. Novel complexes of cyclin-dependent kinases and a cyclin-like protein from *Arabidopsis thaliana* with a function unrelated to cell division. **Cellular and Molecular Life Sciences**, v. 60, p. 401-412, 2003.
- BARTIKOWIAK, B.; LIU, P.; PHATNANI, H. P.; FUDA, N. J.; COOPER, J. J.; PRICE, D. H.; ADELMAN, K.; LIS, J. T.; GREENLEAF, A. L. CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. **Genes & Development**, v. 24, p. 2303-2316, 2010.

BAUMLI, S.; ENDICOTT, J. A.; JOHNSON, L. N. Halogen bonds form the basis for selective P-TEFb inhibition by DRB. **Chemistry & Biology**, v. 17, p. 931-936, 2010.

BAUMLI, S.; HOLE, A. J.; WANG, L.; NOBLE, M. E. M.; ENDICOTT, J. A. The CDK9 tail determines the reaction pathway of positive transcription elongation factor b. **Structure**, v. 20, p. 1788-1795, 2012.

BAUMLI, S.; LOLLI, G.; LOWE, E. D.; TROIANI, S.; RUSCONI, L.; BULLOCK, A. N.; DEBRECZENI, J. É.; KNAPP, S.; JOHNSON, L. N. The structure of P-TEFb (CDK9/cyclin T1), its complex with flavopiridol and regulation by phosphorylation. **The EMBO Journal**, v. 27, p. 1907-1918, 2008.

BEYAERT, R.; KIDD, V. J.; CORNELIS, S.; CRAEN, M. V.; DENECKER, G.; LAHTI, J. M.; GURURAJAN, R.; VANDENABEELE, P.; FIERS, W. Cleavage of PITSLRE kinases by ICE/CASP-1 and CPP32/CASP-3 during apoptosis induced by tumor necrosis factor. **The Journal of Biological Chemistry**, v. 272, n. 18, p. 11694-11697, 1997.

BOISSELIER, E.; AUDET, M.; CANTIN, L.; SALESSE, C. A strategy for purifying glutathione S-transferase in the presence of sodium dodecyl sulfate. **BioTechniques**, v. 51, n. 3, p. 193-194, 2011.

BOLANOS-GARCIA, V. M.; DAVIES, O. R. Structural analysis and classification of native proteins from E. coli commonly co-purified by immobilized metal affinity chromatography. **Biochimica et Biophysica Acta**, v. 1760, n. 9, p. 1304-1313, 2006.

BOSSEMEYER, D. Protein kinases – structure and function. **FEBS Letters**, v. 369, p. 57-61, 1995.

BÖSKEN, C. A.; FARNUNG, L.; HINTERMAIR, C.; SCHACHTER, M. M.; VOGEL-BACHMAYR, K.; BLAZEK, D.; ANAND, K.; FISHER, R. P.; EICK, D.; GEYER, M. The structure and substrate specificity of human Cdk12/Cyclin K. **Nature Communications**, v. 5, n. 3505, p. 1-14, 2014.

BRESOLIN, I. T. L.; MIRANDA, E. A.; BUENO, S. M. A. Cromatografia de afinidade por íons metálicos imobilizados (IMAC) de biomoléculas: Aspectos fundamentais e aplicações tecnológicas. **Química Nova**, v. 32, n. 5, p. 1288-1296, 2009.

BREUER, D.; KOTELKIN, A.; AMMOSOVA, T.; KUMARI, N.; IVANOV, A.; ILATOVSKIY, A. V.; BEULLENS, M.; ROANE, P. R.; BOLLEN, M.; PETUKHOV, M. G.; KASHANCHI, F.; NEKHAI, S. CDK2 regulates HIV-1 transcription by phosphorylation of CDK9 on serine 90. **Retrovirology**, v. 9, n. 94, 2012. Disponível em: <http://www.retrovirology.com/content/pdf/1742-4690-9-94.pdf>. Acesso em: 18 jul. 2014.

BURATOWSKI, S. Progression through the RNA polymerase II CTD cycle. **Molecular Cell**, v. 36, p. 541-547, 2009.

BUNNELL, B. A.; HEATH, L. S.; ADAMS, D. E.; LAHTI, J. M.; KIDD, V. J. Increased expression of a 58-kDa protein kinase leads to changes in the CHO cell cycle. **Proceedings of the National Academy of Sciences**, v. 87, p. 7467-7471, 1990.

CANDURI, F.; PERES, P. C.; CACERES, R. A.; AZEVEDO JR., W. F. CDK9 a potential target for drug development. **Medicinal Chemistry**, v. 4, n. 3, p. 210-218, 2008.

CARRIÓ, M. M.; VILLAVARDE, A. Protein aggregation as bacterial inclusion bodies is reversible. **FEBS Letters**, v. 489, p. 29-33, 2001.

CASALI, N.; PRESTON, A. **Methods in Molecular Biology**. Totowa, N.J.: Humana Press, 2003. 304 p.

CHEN, S.; YIN, X.; ZHU, X.; YAN, J.; JI, S.; CHEN, C.; CAI, M.; ZHANG, S.; ZONG, H.; HU, Y.; YUAN, Z.; SHEN, Z.; GU, J. The C-terminal kinase domain of the p34cdc2-related PITSLRE protein kinase (p110C) associates with p21-activated kinase 1 and inhibits its activity during anoikis. **The Journal of Biological Chemistry**, v. 278, n. 22, p. 20029-20036, 2003.

CHI, Y.; ZHANG, C.; ZONG, H.; HONG, Y.; KONG, X.; LIU, H.; ZOU, W.; WANG, Y.; YUN, X.; GU, J. Thr-370 is responsible for CDK11^{p58} autophosphorylation, dimerization, and kinase activity. **The Journal of Biological Chemistry**, v. 286, n. 3, p. 1748-1757, 2011.

CORNELIS, S.; BRUYNOOGHE, Y.; DENECKER, G.; HUFFEL, S. V.; TINTON, S.; BEYAERT, R. Identification and characterization of a novel cell cycle-regulated internal ribosome entry site. **Molecular Cell**, v. 5, p. 597-605, 2000.

CZUDNOCHOWSKI, N.; BÖSKEN, C. A.; GEYER, M. Serine-7 but not serine-5 phosphorylation primes RNA polymerase II CTD for P-TEFb recognition. **Nature Communications**, v. 3, n. 842, 2012. Disponível em: <<http://www.nature.com/ncomms/journal/v3/n5/pdf/ncomms1846.pdf>>. Acesso em: 18 jul. 2014.

DICKINSON, L. A.; EDGAR, A. J.; EHLEY, J.; GOTTESFELD, J. M. Cyclin L is an RS domain protein involved in pre-mRNA splicing. **The Journal of Biological Chemistry**, v. 277, n. 28, p. 25465-25473, 2002.

DINGWALL, C.; DILWORTH, S. M.; BLACK, S. J.; KEARSEY, S. E.; COX, L. S.; LASKEY, R. A. Nucleoplasmin cDNA sequence reveals polyglutamic acid tracts and a cluster of sequences homologous to putative nuclear localization signals. **The EMBO Journal**, v. 6, n. 1, p. 69-74, 1987.

DRAETTA, G.; LUCA, F.; WESTENDORF, J.; BRIZUELA, L.; RUDERMAN, J. Cdc2 protein kinase is complexed with both cyclin A and B: Evidence for proteolytic inactivation of MPF. **Cell**, v. 56, n. 5, p. 829-838, 1989.

DROGAT, J.; MIGEOT, V.; MOMMAERTS, E.; MULLIER, C.; DIEU, M.; BAKEL, H.; HERMAND, D. Cdk11-CyclinL controls the assembly of the RNA polymerase II Mediator complex. **Cell reports**, v. 2, n. 5, p. 1078-1086, 2012.

DUNPHY, W. G.; BRIZUELA, L.; BEACH, D.; NEWPORT, J. The *Xenopus* cdc2 protein is a component of MPF, a cytoplasmic regulator of mitosis. **Cell**, v. 54, n. 3, p. 423-431, 1988.

EDWARDS, M. C.; WONG, C.; ELLEDGE, S. J. Human cyclin K, a novel RNA Polymerase II-associated cyclin possessing both carboxy-terminal domain kinase and Cdk-activating kinase activity. **Molecular and Cellular Biology**, v. 18, n. 7, p. 4291-4300, 1998.

EVANS, T.; ROSENTHAL, E. T.; YOUNGBLOM, J.; DISTEL, D.; HUNT, T. Cyclin: A protein specified by maternal mRNA in sea urchin eggs that is destroyed at each cleavage division. **Cell**, v. 33, n. 2, p. 389-396, 1983.

FENG, Y.; ARIZA, M. E.; GOULET, A.; SHI, J.; NELSON, M. A. Death-signal-induced relocalization of cyclin-dependent kinase 11 to mitochondria. **Biochemical Journal**, v. 392, p. 65-73, 2005.

FISCHER, P. M. The design of drug candidate molecules as selective inhibitors of therapeutically relevant protein kinases. **Current Medicinal Chemistry**, v. 11, p. 1563-1583, 2004.

FRANCK, N.; MONTEBAULT, E.; ROMÉ, P.; PASCAL, A.; CREMET, J.; GIET, R. CDK11^{p58} is required for centriole duplication and Plk4 recruitment to mitotic centrosomes. **PLoS ONE**, v. 6, n. 1, 2011. Disponível em: <<http://www.plosone.org/article/fetchObject.action?uri=info%3Adoi%2F10.1371%2Fjournal.pone.0014600&representation=PDF>>. Acesso em: 18 jul. 2014.

FU, J.; YOON, H.; QIN, J.; WONG, J. Regulation of P-TEFb elongation complex activity by CDK9 acetylation. **Molecular and Cellular Biology**, v. 27, n. 13, p. 4641-4651, 2007.

FU, T.; PENG, J.; LEE, G.; PRICE, D. H.; FLORES, O. Cyclin K functions as a CDK9 regulatory subunit and participates in RNA polymerase II transcription. **The Journal of Biological Chemistry**, v. 274, n. 49, p. 34527-34530, 1999.

FUJINAGA, K.; CUJEC, T. P.; PENG, J.; GARRIGA, J.; PRICE, D. H.; GRAÑA, X.; PETERLIN, B. M. The ability of positive transcription elongation factor b to transactivate human immunodeficiency virus transcription depends on a functional kinase domain, cyclin T1 and Tat. **Journal of Virology**, v. 72, n. 9, p. 7154-7159, 1998.

FUJINAGA, K.; IRWIN, D.; HUANG, Y.; TAUBE, R.; KUROSU, T.; PETERLIN, B. M. Dynamics of human immunodeficiency virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the transactivation response element. **Molecular and Cellular Biology**, v. 24, n. 2, p. 787-795, 2004.

GARBER, M. E.; WEI, P.; KEWALRAMANI, V. N.; MAYALL, T. P.; HERMANN, C. H.; RICE, A. P.; LITTMAN, D. R.; JONES, K. A. The interaction between HIV-1 Tat and human cyclin T1 requires zinc and a critical cysteine residue that is not conserved in the murine CycT1 protein. **Genes & Development**, v. 12, p. 3512-3527, 1998.

GARRIGA, J.; BHATTACHARYA, S.; CALBO, J.; MARSHALL, R. M.; TRUONGCAO, M.; HAINES, D. S.; GRAÑA, X. CDK9 is constitutively expressed throughout the cell cycle, and its steady-state expression is independent of SKP2. **Molecular and Cellular Biology**, v. 23, n. 15, p. 5165-5173, 2003.

GARRIGA, J.; GRAÑA, X. Cellular control of gene expression by T-type cyclin/CDK9 complexes. **Gene**, v. 337, p. 15-23, 2004.

GARRIGA, J.; XIE, H.; OBRADOVIC, Z.; GRAÑA, X. Selective control of gene expression by CDK9 in human cells. **Journal of Cellular Physiology**, v. 222, p. 200-208, 2010.

GAUTIER, J.; NORBURY, C.; LOHKA, M.; NURSE, P. MALLER, J. Purified maturation-promoting factor contains the product of a *Xenopus* homolog of the fission yeast cell cycle control gene *cdc2+*. **Cell**, v. 54, n. 3, p. 433-439, 1988.

GLOVER-CUTTER, K.; LAROCHELLE, S.; ERICKSON, B.; ZHANG, C.; SHOKAT, K.; FISHER, R. P.; BENTLEY, D. L. TFIIH-associated CDK7 kinase functions in phosphorylation of C-terminal domain ser7 residues, promoter proximal-pausing, and termination by RNA polymerase II. **Molecular and Cellular Biology**, v. 29, n. 20, p. 5455-5464, 2009.

GRAÑA, X.; DE LUCA, A.; SANG, N.; FU, Y.; CLAUDIO, P. P.; ROSENBLATT, J.; MORGAN, D. O.; GIORDANO, A. PITALRE, a nuclear CDC2-related protein kinase that phosphorylates the retinoblastoma protein in vitro. **Proceedings of the National Academy of Sciences**, v. 91, p. 3834-3838, 1994.

GRÄSLUND, S.; NORDLUND, P.; WEIGELT, J.; HALLBERG, B. M.; BRAY, J.; GILEADI, O.; KNAPP, S.; OPPERMAN, U.; ARROWSMITH, C.; HUI, R.; MING, J.; DHE-PAGANON, S.; PARK, H.; SAVCHENKO, A.; YEE, A.; EDWARDS, A.; VINCENTELLI, R.; CAMBILLAU, C.; KIM, R.; BÜSSOW, K.; GUNSALUS, K. C. Protein production and purification. **Nature Methods**, v. 5, n. 2, p. 135-146, 2008.

GREEN, H.; WANG, N. Codon reiteration and the evolution of proteins. **Proceedings of the National Academy of Sciences**, v. 91, n. 10, p. 4298-4302, 1994.

HANKS, S. K.; HUNTER, T. The eukaryotic protein kinase superfamily: Kinase (catalytic) domain structure and classification. **The FASEB Journal**, v. 9, p. 576-596, 1995.

HARTWELL, L. H.; CULOTTI, J.; PRINGLE, J. R.; REID, B. Genetic control of the cell division cycle in yeast. **Science**, v. 183, n. 4120, p. 46-51, 1974.

HERMANN, C. H.; CARROLL, R. G.; WEI, P.; JONES, K. A.; RICE, A. P. Tat-associated kinase, TAK, activity is regulated by distinct mechanisms in peripheral blood lymphocytes and promonocytic cell lines. **Journal of Virology**, v. 72, n.12, p. 9881-9888, 1998.

HOFFMANN, D.; SCHWARCK, D.; BANNING, C.; BRENNER, M.; MARIYANNA, L.; KREPSTAKIES, M.; SCHINDLER, M.; MILLAR, D.; HAUBER, J. Formation of trans-activation competent HIV-1 Rev:RRE complexes require the recruitment of multiple protein activation domains. **PLoS ONE**, v. 7, n. 6, 2012. Disponível em: <<http://www.plosone.org/article/fetchObject.action?uri=info%3Adoi%2F10.1371%2Fjournal.pone.0038305&representation=PDF>>. Acesso em: 18 jul. 2014.

HU, D.; MAYEDA, A.; TREMBLEY, J. H.; LAHTI, J. M.; KIDD, V. J. CDK11 complexes promote pre-mRNA splicing. **The Journal of Biological Chemistry**, v. 278, n. 10, p. 8623-8629, 2003.

HU, D.; VALENTINE, M.; KIDD, V. J.; LAHTI, J. M. CDK11^{p58} is required for the maintenance of sister chromatid cohesion. **Journal of Cell Science**, v. 120, n. 14, p. 2424-2434, 2007.

HUNT, T. Maturation promoting factor, cyclin and the control of M-phase. **Current Opinion in Cell Biology**, v. 1, p. 268-274, 1989.

IDICULA-THOMAS, S.; BALAJI, P. V. Understanding the relationship between the primary structure of proteins and its propensity to be soluble on overexpression in *Escherichia coli*. **Protein Science**, v. 14, p. 582-592, 2005.

JANG, M. K.; MOCHIZUKI, K.; ZHOU, M.; JEONG, H.; BRADY, J. N.; OZATO, K. The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. **Molecular Cell**, v. 19, p. 523-534, 2005.

KAO, S.; CALMAN, A. F.; LUCIW, P.; PETERLIN, B. M. Antitermination of transcription within the long terminal repeat of HIV-1 by tat gene product. **Nature**, v. 330, p. 489-493, 1987.

KATTI, M. V.; SAMI-SUBBU, R.; RANJEKAR, P. K.; GUPTA, V. S. Amino acid repeat patterns in protein sequences: Their diversity and structural-functional implications. **Protein Science**, v. 9, n. 6, p. 1203-1209, 2000.

KIERNAN, R. E.; EMILIANI, S.; NAKAYAMA, K.; CASTRO, A.; LABBÉ, J. C.; LORCA, T.; NAKAYAMA, K. BENKIRANE, M. Interaction between cyclin T1 and SCF^{SKP2} targets CDK9 for ubiquitination and degradation by the proteasome. **Molecular and Cellular Biology**, v. 21, n. 23, p. 7956-7970, 2001.

KIM, J. B.; SHARP, P. A. Positive transcription elongation factor b phosphorylates hSPT5 and RNA polymerase II carboxyl-terminal domain independently of cyclin-dependent kinase-activating kinase. **The Journal of Biological Chemistry**, v. 276, n. 15, p. 12317-12323, 2001.

KOHOUTEK, J. P-TEFb the final frontier. **Cell Division**, v. 4, n. 19, 2009. Disponível em: <<http://www.celldiv.com/content/pdf/1747-1028-4-19.pdf>>. Acesso em: 18 jul. 2014.

KONG, X.; GAN, H.; HAO, Y.; CHENG, C.; JIANG, J.; HONG, Y.; YANG, J.; ZHU, H.; CHI, Y.; YUN, X.; GU, J. CDK11^{p58} phosphorylation of PAK1 Ser174 promotes DLC2 binding and roles on cell cycle progression. **Journal of Biochemistry**, v. 146, n. 3, p. 417-427, 2009.

LAHTI, J. M.; XIANG, J.; HEATH, L. S.; CAMPANA, D.; KIDD, V. J. PITSLRE protein kinase activity is associated with apoptosis. **Molecular and Cellular Biology**, v. 15, n. 1, p. 1-11, 1995.

LAROCHELLE, S.; AMAT, R.; GLOVER-CUTTER, K.; SANSÓ, M.; ZHANG, C.; ALLEN, J. J.; SHOKAT, K. M.; BENTLEY, D. L.; FISHER, R. P. Cyclin-dependent kinase control of the initiation-to-elongation switch of RNA polymerase II. **Nature Structural & Molecular Biology**, v. 19, n. 11, p. 1108-1115, 2012.

LEE, M. G.; NURSE, P. Complementation used to clone a human homologue of the fission yeast cell cycle control gene cdc2. **Nature**, v. 327, n. 6117, p. 31-35, 1987.

LEE, C.; PARK, Y.; KIM, Y.; PETERKOFISKY, A.; SEOK, Y. Phosphorylation-dependent mobility shift of proteins on SDS-PAGE is due to decreased binding of SDS. **Bulletin of the Korean Chemical Society**, v. 34, n. 7, p. 2063-2066, 2013.

LEOPOLDINO, A. M.; CANDURI, F.; CABRAL, H.; JUNQUEIRA, M.; MARQUI, A. B. T.; APPONI, L. H.; FONSECA, I. O.; DOMONT, G. B.; SANTOS, D. S.; VALENTINI, S.; BONILLA-RODRIGUEZ, G. O.; FOSSEY, M. A.; AZEVEDO JR, W. F.; TAJARA, E. H. Expression, purification and circular dichroism analysis of human CDK9. **Protein Expression and Purification**, v. 47, p. 614-620, 2006.

LI, T.; INOUE, A.; LAHTI, J. M.; KIDD, V. J. Failure to proliferate and mitotic arrest of CDK11^{p110/p58}-null mutant mice at the blastocyst stage of embryonic cell development. **Molecular and Cellular Biology**, v. 24, n. 8, p. 3188-3197, 2004.

LI, M.; POLIAKOV, A.; DANIELSON, U. H.; SU, Z.; JANSON, J. Refolding of a recombinant full-length non-structural (NS3) protein from hepatitis C virus by chromatographic procedures. **Biotechnology Letters**, v. 25, n. 20, p. 1729-1734, 2003.

- LI, Q.; PRICE, J. P.; BYERS, S. A.; CHENG, D.; PENG, J.; PRICE, D. H. Analysis of the large inactive P-TEFb complex indicates that it contains one 7SK molecule, a dimer of HEXIM1 or HEXIM2, and two P-TEFb molecules containing Cdk9 phosphorylated at threonine 186. **The Journal of Biological Chemistry**, v. 280, n. 31, p. 28819-28826, 2005.
- LIN, X.; TAUBE, R.; FUJINAGA, K.; PETERLIN, B. M. P-TEFb containing Cyclin K and Cdk9 can activate transcription via RNA. **The Journal of Biological Chemistry**, v. 277, n. 19, p. 16873-16878, 2002.
- LIU, H.; RICE, A. P. Genomic organization and characterization of promoter function of the human CDK9 gene. **Gene**, v. 252, p. 51-59, 2000.
- LOLLI, G.; LOWE, E. D.; BROWN, N. R.; JOHNSON, L. N. The crystal structure of human CDK7 and its protein recognition properties. **Structure**, v. 12, p. 2067-2079, 2004.
- LOYER, P.; TREMBLEY, J. H.; GRENET, J. A.; BUSSON, A.; CORLU, A.; ZHAO, W.; KOCAK, M.; KIDD, V. J.; LAHTI, J. M. Characterization of cyclin L1 and L2 interactions with CDK11 and splicing factors. **The Journal of Biological Chemistry**, v. 283, n. 12, p. 7721-7732, 2008.
- LOYER, P.; TREMBLEY, J. H.; KATONA, R.; KIDD, V. J.; LAHTI, J. M. Role of CDK/cyclin complexes in transcription and RNA splicing. **Cellular Signaling**, v. 17, p. 1033-1051, 2005.
- LOYER, P.; TREMBLEY, J. H.; LAHTI, J. M.; KIDD, V. J.; The RNP protein RNPS1 associates with specific isoforms of the p34^{cdc2}-related PITSLRE protein kinase *in vivo*. **Journal of Cell Science**, v. 111, p. 1495-1506, 1998.
- LU, H.; ZAWEL, L.; FISHER, L.; EGLY, J.; REINBERG, D.; Human general transcription factor IIH phosphorylates the C-terminal domain of RNA polymerase II. **Nature**, v. 358, p. 641-645 1992.
- MALUMBRES, M.; HARLOW, E.; HUNT, T.; HUNTER, T.; LAHTI, J. M.; MANNING, G.; MORGAN, D. O.; TSAI, L.; WOLGEMUTH, D. J. Cyclin-dependent kinases: a family portrait. **Nature Cell Biology**, v. 11, n. 11, 2009.
- MANNING, G.; WHYTE, D. B.; MARTINEZ, R.; HUNTER, T.; SUDARSANAM, S. The protein kinase complement of the human genome. **Science**, v. 298, p. 1912-1934, 2002.
- MARCOTTE, E. M.; PELLEGRINI, M.; YEATES, T. O.; EISENBERG, D. A census of protein repeats. **Journal of Molecular Biology**, v. 293, n. 1, p. 151-160, 1998.
- MARSHALL, N. F.; PRICE, D. H. Control of formation of two distinct classes of RNA polymerase II elongation complexes. **Molecular and Cellular Biology**, v. 12, n. 5, p. 2078-2090, 1992.
- MARSHALL, N. F.; PRICE, D. H. Purification of P-TEFb, a transcription factor required for the transition into productive elongation. **The Journal of Biological Chemistry**, v. 270, n. 21, p. 12335-12338, 1995.
- MARSHALL, N. F.; PENG, J.; XIE, Z.; PRICE, D. H. Control of RNA Polymerase II elongation potential by a novel carboxyl-terminal domain kinase. **The Journal of Biological Chemistry**, v. 271, n. 43, p. 27176-27183, 1996.

MASUI, Y.; MARKERT, C. L. Cytoplasmic control of nuclear behavior during meiotic maturation of frog oocytes. **Journal of Experimental Zoology**, v. 177, n. 2, p. 129-145, 1971.

MBONYE, U. R.; GOKULRANGAN, G.; DATT, M.; DOBROWOLSKI, C.; COOPER, M.; CHANCE, M. R.; KARN, J. Phosphorylation of CDK9 at Ser175 enhances HIV transcription and is a marker of activated P-TEFb in CD4+ T Lymphocytes. **PLoS Pathogens**, v. 9, n. 5, 2013. Disponível em: <<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3642088/pdf/ppat.1003338.pdf>>. Acesso em: 18 jul. 2014.

MICHAUX, C.; POMROY, N. C.; PRIVÉ, G. G. Refolding SDS-denatured proteins by the addition of amphipathic cosolvents. **Journal of Molecular Biology**, v. 375, n. 5, p. 1477-1488, 2008.

MICHELS, A. A.; NGUYEN, V. T.; FRALDI, A.; LABAS, V.; EDWARDS, M.; BONNET, F.; LANIA, L.; BENS AUDE, O. MAQ1 and 7SK RNA interact with CDK9/Cyclin T complexes in a transcription-dependent manner. **Molecular and Cellular Biology**, v. 23, n. 14, p. 4859-4869, 2003.

MIKOLAJCZYK, M.; NELSON, M. A. Regulation of stability of cyclin-dependent kinase CDK11^{p110} and a caspase-processed form, CDK11^{p46}, by Hsp90. **Biochemical Journal**, v. 384, p. 461-467, 2004.

MIKOLAJCZYK, M.; SHI, J.; VAILLANCOURT, R. R.; SACHS, N. A.; NELSON, M. The cyclin-dependent kinase 11^{p46} isoform interacts with RanBPM. **Biochemical and Biophysical Research**, v. 310, n. 1, p. 14-18, 2003.

MOCHIZUKI, K.; NISHIYAMA, A.; JANG, M. K.; DEY, A.; GHOSH, A.; TAMURA, T.; NATSUME, H.; YAO, H.; OZATO, K. The bromodomain protein Brd4 stimulates G1 gene transcription and promotes progression to S phase. **The Journal of Biological Chemistry**, v. 283, n. 14, p. 9040-9048, 2008.

NELSON, D. L.; COX, M. M. **Lehninger princípios de bioquímica**. Porto Alegre: Artmed, 2011. 1274 p.

NGUYEN, V. T.; KISS, T.; MICHELS, A. A.; BENS AUDE, O. 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. **Nature**, v. 414, p. 322-325, 2001.

O'KEEFFE, B.; FONG, Y.; CHEN, D.; ZHOU, S.; ZHOU Q. Requirement for a kinase-specific chaperone pathway in the production of a Cdk9/Cyclin T1 heterodimer responsible for P-TEFb-mediated Tat stimulation of HIV-1 transcription. **The Journal of Biological Chemistry**, v. 275, n. 1, p. 279-287, 2000.

PAPP, E.; CSERMELY, P. Chemical chaperones: Mechanisms of action and potential use. **Handbook of Experimental Pharmacology**, v. 172, p. 405-416, 2006.

PENG, J.; MARSHALL, N. F.; PRICE, D. H. Identification of a cyclin subunit required for the function of Drosophila P-TEFb. **The Journal of Biological Chemistry**, v. 273, n. 22, p. 13855-13860, 1998a.

PENG, J.; ZHU, Y.; MILTON, J. T.; PRICE, D. H. Identification of multiple cyclin subunits of human P-TEFb. **Genes & Development**, v. 12, p. 755-762, 1998b.

PETERLIN, B. M.; PRICE, D. H. Controlling the elongation phase of transcription with P-TEFb. **Molecular Cell**, v. 23, p. 297-305, 2006.

PETRETTI, C.; SAVOIAN, M.; MONTEBAULT, E.; GLOVER, D. M.; PRIGENT, C.; GIET, R. The PITSLRE/CDK11^{p58} protein kinase promotes centrosome maturation and bipolar spindle formation. **EMBO reports**, v. 7, n. 4, p. 418-424, 2006.

PINHEIRO, R.; LIAW, P.; YANKULOV, K. A uniform procedure for the purification of CDK7/CycH/MAT1, CDK8/CycC, CDK9/CycT1. **Biological Procedures Online**, v. 6, n. 1, p. 163-172, 2004.

PRICE, D. H. P-TEFb, a cyclin-dependent kinase controlling elongation by RNA polymerase II. **Molecular and Cellular Biology**, v. 20, n. 8, p. 2629-2634, 2000.

PRICE, D. H. Poised polymerases: On your mark... Get set... Go! **Molecular Cell**, v. 30, p. 7-10, 2008.

RADHAKRISHNAN, S. K.; GARTEL, A. L. CDK9 phosphorylates p53 on serine residues 33, 315 and 392. **Cell Cycle**, v. 5, n. 5, p. 519-521, 2006.

REHM, B. H. A.; QI, Q.; BEERMANN, B. B.; HINZ, H. J.; STEINBÜCHEL, A. Matrix-assisted *in vitro* refolding of *Pseudomonas aeruginosa* class II polyhydroxyalkanoate synthase from inclusion bodies produced in recombinant *Escherichia coli*. **Biochemical Journal**, n. 358, v. 1, p. 263-268, 2001.

ROMANO, G.; GIORDANO, A. Role of the cyclin-dependent kinase 9 – related pathway in mammalian gene expression and human diseases. **Cell Cycle**, v. 7, n. 23, p. 3664-3668, 2008.

SABÒ, A.; LUSIC, M.; CERESETO, A.; GIACCA, M. Acetylation of conserved lysines in the catalytic core of cyclin-dependent kinase 9 inhibits kinase activity and regulates transcription. **Molecular and Cellular Biology**, v. 28, n. 7, p. 2201- 2212, 2008.

SAHDEV, S.; KHATTAR, S. K.; SAINI, K. S. Production of active eukaryotic proteins through bacterial expression systems: A review of the existing biotechnology strategies. **Molecular and Cellular Biochemistry**, v. 307, p. 249-264, 2008.

SAMBROOK, J. F.; RUSSEL, D. W. **Molecular cloning: a laboratory manual**. Cold Spring Harbor Laboratory Press, 2001. 2100 p.

SCHNEIDER, E. V.; BÖTTCHER, J.; BLAESSE, M.; NEUMANN, L.; HUBER, R.; MASKOS, K. The structure of the CDK8/CycC implicates specificity in the CDK/Cyclin family and reveals interaction with a deep pocket binder. **Journal of Molecular Biology**, v. 412, n. 2, p. 251-266, 2011.

SCHWARTZ, R.; TING, C. S.; KING, J. Whole proteome pI values correlate with subcellular localizations of proteins for organisms within the Three Domains of life. **Genome Research**, v. 11, p. 703-709, 2001.

SEHGAL, P. B.; DARNELL J. E.; TAMM, I. The inhibition by DRB (5,6-dichloro-1-beta-D-ribofuranosylbenzimidazole) of hnRNA and mRNA production in HeLa cells. **Cell**, v. 9, n. 3, p. 473-480, 1976.

SHARP, P. M.; LI, W. The codon adaptation index – a measure of directional synonymous codon usage bias, and its potential applications. **Nucleic Acids Research**, v. 15, n. 3, p.1281-1295, 1987.

SHI, J.; FENG, Y.; GOULET, A.; VAILLANCOURT, R. R.; SACHS, N. A.; HERSHEY, J. W.; NELSON, M. A. The p34^{cdc2}-related cyclin-dependent kinase 11 interacts with the p47 subunit of eukaryotic initiation factor 3 during apoptosis. **The Journal of Biological Chemistry**, v. 278, n. 7, p. 5062-5071, 2003.

SHI, J.; HERSHEY, J. W. B.; NELSON, M. A. Phosphorylation of the eukaryotic initiation factor 3f by cyclin-dependent kinase 11 during apoptosis. **FEBS Letters**, v. 583, p. 971-977, 2009.

SHI, Y.; MOWERY, R. A.; ASHLEY, J.; HENTZ, M.; RAMIREZ, A. J.; BILGICER, B.; SLUNT-BROWN, H.; BORCHELT, D. R. SHAW, B. F. Abnormal SDS-PAGE migration of cytosolic proteins can identify domains and mechanisms that control surfactant binding. **Protein Science**, v. 21, p. 1197-1209, 2012.

SHI, J.; NELSON, M. A. The cyclin-dependent kinase 11 interacts with NOT2. **Biochemical and Biophysical Research Communications**, v. 334, p. 1310-1316, 2005.

SHORE, S. M.; BYERS, S. A.; MAURY, W.; PRICE, D. H. Identification of a novel isoform of Cdk9. **Gene**, v.307, p. 175-182, 2003.

SHORE, S. M.; BYERS, S. A.; DENT, P.; PRICE, D. H. Characterization of Cdk9₅₅ and differential regulation of two Cdk9 isoforms. **Gene**, v. 350, p. 51-58, 2005.

SMITH, G. K.; KE, Z.; GUO, H.; HENGGE, A. C. Insights into the phosphoryl transfer mechanism of cyclin-dependent protein kinases from ab initio QM/MM free-energy studies. **The Journal of Physical Chemistry B**, v. 115, n. 46, p. 13713-13722, 2011.

TAHIROV, T. H.; BABAYEVA, N. D.; VARZAVAND, K.; COOPER, J. J.; SEDORE, S. C.; PRICE, D. H. Crystal structure of HIV-1 Tat complexed with human P-TEFb. **Nature**, v. 465, p. 747-751, 2010.

TANG, D.; GURURAJAN, R.; KIDD, V. J. Phosphorylation of PITSLRE p110 isoforms accompanies their processing by caspases during Fas-mediated cell death. **The Journal of Biological Chemistry**, v. 273, n. 26, p. 16601-16607, 1998.

TAUBE, R.; LIN, X.; IRWIN, D.; FUJINAGA, K.; PETERLIN, B. M. Interaction between P-TEFb and the C-terminal domain of RNA polymerase II activates transcriptional elongation from sites upstream or downstream of target genes. **Molecular and Cellular Biology**, v. 22, n. 1, p. 321-331, 2002.

TIETJEN, J. R.; ZHANG, D. W.; RODRÍGUEZ-MOLINA, J. B.; WHITE, B. E.; AKHTAR, S.; HEIDEMANN, M.; LI, X.; CHAPMAN, R. D.; SHOKAT, K.; KELES, S.; EICK, D.; ANSARI, A. Z. Chemical-genomic dissection of the CTD code. **Nature Structural & Molecular Biology**, v. 17, p. 1154-1161, 2010.

TINTON, S. A.; SCHEPENS, B.; BRUYNNOOGHE, Y.; BEYAERT, R.; CORNELIS, S. Regulation of the cell-cycle-dependent internal ribosome entry site of the PITSLRE protein kinase: Roles of Unr (upstream of N-ras) protein and phosphorylated translation initiation factor eIF-2 α . **Biochemical Journal**, v. 385, p. 155-163, 2005.

TREMBLEY, J. H.; HU, D.; HSU, L.; YEUNG, C.; SLAUGHTER, C.; LAHTI, J. M.; KIDD, V. J. PITSLRE p110 protein kinases associate with transcription complexes and affect their activity. **The Journal of Biological Chemistry**, v. 277, n. 4, p. 2589-2596, 2002.

TREMBLEY, J. H.; HU, D.; SLAUGHTER, C. A.; LAHTI, J. M.; KIDD, V. J. Casein kinase 2 interacts with cyclin-dependent kinase 11 (CDK11) *in vivo* and phosphorylates both the RNA polymerase II carboxyl-terminal domain and CDK11 *in vitro*. **The Journal of Biological Chemistry**, v. 278, n. 4, p. 2265-2270, 2003.

TYE, C. E.; RATTRAY, K. R.; WARNER, K. J.; GORDON, J. A. R.; SODEK, J.; HUNTER, G. K.; GOLDBERG, H. A. Delineation of the hydroxyapatite-nucleating domains of bone sialoprotein. **The Journal of Biological Chemistry**, v. 278, n. 10, p. 7949-7955, 2003.

VOET, D.; VOET, J. G. **Bioquímica**. Porto Alegre: Artmed, 2006. 1616 p.

WADA, T.; TAKAGI, T.; YAMAGUCHI, Y.; WATANABE, D.; HANDA, H. Evidence that P-TEFb alleviates the negative effect of DSIF on RNA polymerase II-dependent transcription *in vitro*. **The EMBO Journal**, v. 17, n. 24, p. 7395-7403, 1998.

WANG, Y.; DOW, E. C.; LIANG, Y.; RAMAKRISHNAN, R.; LIU, H.; SUNG, T.; LIN, X.; RICE, A. P. Phosphatase PPM1A regulates phosphorylation of Thr-186 in the Cdk9 T-loop. **The Journal of Biological Chemistry**, v. 283, n. 48, p. 33578-33584, 2008.

WEI, P.; GARBER, M. E.; FANG, S. M.; FISCHER, W. H.; JONES, K. A. A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. **Cell**, v. 92, p. 451-462, 1998.

WEIJLAND, A.; HARMARK, K.; COOL, R. H.; ANBORGH, P. H.; PARMEGGIANI, A. Elongation factor Tu: a molecular switch in protein biosynthesis. **Molecular Microbiology**, v. 6, n. 6, p. 683-688, 1992.

WOESTENENK, E. A.; HAMMARSTRÖM, M.; BERG, S.; HÄRD, T.; BERGLUND, H. His tag effect on solubility of human proteins produced in *Escherichia coli*: a comparison between four expression vectors. **Journal of Structural and Functional Genomics**, v. 5, n. 3, p. 217-229, 2004.

WU, S.; CHIANG, C. The double bromodomain-containing chromatin adaptor Brd4 and transcriptional regulation. **The Journal of Biological Chemistry**, v. 282, n. 18, p. 13141-13145, 2007.

XIANG, J.; LAHTI, J. M.; GRENET, J.; EASTON, J.; KIDD, V. J. Molecular cloning and expression of alternatively spliced PITSLRE protein kinase isoforms. **The Journal of Biological Chemistry**, v. 269, n. 22, p. 15786-15794, 1994.

YAMAGUCHI, Y.; TAKAGI, T.; WADA, T.; YANO, K.; FURUYA, A.; SUGIMOTO, S.; HASEGAWA, J.; HANDA, H. NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. **Cell**, v. 97, n. 1, p. 41-51, 1999.

YANG, Z.; HE, N.; ZHOU, Q. Brd4 recruits P-TEFb to chromosomes at late mitosis to promote G1 gene expression and cell cycle progression. **Molecular and Cellular Biology**, v. 28, n. 3, p. 967-976, 2008.

YANG, Z.; YIK, J. H. N.; CHEN, R.; HE, N.; JANG, M. K.; OZATO, K.; ZHOU, Q. Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. **Molecular Cell**, v. 19, p. 535-545, 2005.

YANG, Z.; ZHANG, L.; ZHANG, Y.; ZHANG, T.; FENG, Y.; LU, X.; LAN, W.; WANG, J.; WU, H.; CAO, C.; WANG, X. Highly efficient production of soluble proteins from insoluble inclusion bodies by a two-step-denaturing and refolding method. **PLoS ONE**, v. 6, n. 7, p. 1-8, 2011.

YANG, Z.; ZHU, Q.; LUO, K.; ZHOU, Q. The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription. **Nature**, v. 414, p. 317-322, 2001.

YASUKAWA, T.; KANEI-ISHII, C.; MAEKAWA, T.; FUJIMOTO, J.; YAMAMOTO, T.; ISHII, S. Increase of solubility of foreign proteins in Escherichia coli by coproduction of the bacterial thioredoxin. **The Journal of Biological Chemistry**, v. 270, n. 43, p. 25328-25331, 1995.

YIK, J. H. N.; CHEN, R.; NISHIMURA, R.; JENNINGS, J. L.; LINK, A. J.; ZHOU, Q. Inhibition of P-TEFb (CDK9/Cyclin T) kinase and RNA polymerase II transcription by the coordinated actions of HEXIM1 and 7SK snRNA. **Molecular Cell**, v. 12, p. 971-982, 2003.

YIK, J. H. N.; CHEN, R.; PEZDA, A. C.; ZHOU, Q. Compensatory contributions of HEXIM1 and HEXIM2 in maintaining the balance of active and inactive positive transcription elongation factor b complexes for control of transcription. **The Journal of Biological Chemistry**, v. 280, n. 16, p. 16368-16376, 2005.

YOH, S. M.; CHO, H.; PICKLE, L.; EVANS, R. M.; JONES, K. A. The Spt6 SH2 domain binds Ser2-P RNAP II to direct Iws1-dependent mRNA splicing and export. **Genes & Development**, v. 21, p. 160-174, 2007.

YOKOYAMA, H.; GRUSS, O. J.; RYBINA, S.; CAUDRON, M.; SCHEIDER, M.; WILM, M.; MATTAJ, I. W.; KARSENTI, E. CDK11 is a RanGTP-dependent microtubule stabilization factor that regulates spindle assembly rate. **The Journal of Cell Biology**, v. 180, n. 5, p. 867-875, 2008.

ZHANG, S.; CAI, M.; ZHANG, S.; XU, S.; CHEN, S.; CHEN, X.; CHEN, C.; GU, J. Interaction of p58^{PITSLRE}, a G2/M-specific protein kinase, with cyclin D3. **The Journal of Biological Chemistry**, v. 277, n. 38, p. 35314-35322, 2002.

ZHOU, M.; HUANG, K.; JUNG, K.; CHO, W.; KLASE, Z.; KASHANCHI, F.; PISE-MASISON, C. A.; BRADY, J. N. Bromodomain protein Brd4 regulates human immunodeficiency virus transcription through phosphorylation of CDK9 at threonine 29. **Journal of Virology**, v. 83, n. 2, p. 1036-1044, 2009

ZHOU, M.; NEKHAI, S.; BHARUCHA, D. C.; KUMAR, A.; GE, H.; PRICE, D. H.; EGLY, J.; BRADY, J. N. TFIIH inhibits CDK9 phosphorylation during human immunodeficiency virus type 1 transcription. **The Journal of Biological Chemistry**, v. 276, n. 48, p. 44633-44640, 2001.

ZHU, Y.; PE'ERY, T.; PENG, J.; RAMANATHAN, Y.; MARSHALL, N.; MARSHALL, T.; AMENDT, B.; MATHEWS, M. B.; PRICE, D. H. Transcription elongation factor P-TEFb is required for HIV-1 Tat transactivation in vitro. **Genes & Development**, v. 11, p. 2622-2632, 1997.