KAIQUE DIAS GALERA

Estudo estrutural comparativo das 20 CDKs humanas utilizando
ferramentas de bioinformática

Dissertação apresentada ao Programa de Pós-Graduação em Química do Instituto de Química de São Carlos, Universidade de São Paulo, como parte integrante dos requisitos para obtenção do grau de mestre em Química, área de concentração em Química Orgânica e Biológica.

Orientadora: Profa. Dra. Fernanda Canduri

São Carlos

Agosto/2020

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE QUÍMICA DE SÃO CARLOS
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA ORGÂNICA E BIOLÓGICA

KAIQUE DIAS GALERA

Estudo estrutural comparativo das 20 CDKs humanas utilizando
ferramentas de bioinformática

Orientadora: Profa. Dra. Fernanda Canduri

São Carlos
Agosto/2020

AGRADECIMENTOS

- À minha orientadora, Prof. Dr. Fernanda Canduri, pela oportunidade de orientação, esforço e paciência depositados em mim.
- A todos os colegas do grupo Nequimed e de outros, que contribuíram direta e indiretamente para o desenvolvimento deste trabalho.
- À minha família, pelo apoio e incentivo, que pude contar em todos os momentos.
- Aos meus amigos e minha namorada que, por mais da distância, sempre estiveram dando o suporte necessário para o meu melhor desempenho.
- Às agências de fomento, CNPq e Fapesp, pelo auxílio com bolsa e financiamento do projeto de pesquisa, respectivamente.
- A todos os brasileiros que acreditam num mundo melhor através da ciência e na pesquisa científica.

RESUMO

As Quinases dependentes de Ciclinas (CDKs) são serina/treonina quinases, cuja atividade é modulada pela interação com ciclinas e sucessivas fosforilações que ocorrem no complexo. Com estudos desde 1980, já é claro que as CDKs atuam como uma das principais famílias de proteínas quinases eucarióticas envolvidas na integração de sinais extra e intracelulares para modular a transcrição gênica e a divisão celular. Relacionado com uma série de disfunções carcinogênicas, mutações podem resultar no mal funcionamento dessa família. O estudo e caracterização das CDKs pode contribuir para o progresso da terapia contra o câncer, desde que abre a possibilidade de encontrar inibidores cada vez mais específicos. O projeto tem como objetivo estudar os membros da família das CDKs, explorando os aspectos estruturais de forma a correlacionar com sua importância fisiológica, desempenhada em cada tecido. Para tanto, um estudo comparativo está sendo realizado, com enfoque na arquitetura do sítio ativo, na extensão da cadeia polipeptídica de cada membro, e na interface de interação com a ciclina. Dessa forma, o projeto foi fundamentado na busca por informações disponíveis nos bancos de dados e bibliografia, e na modelagem molecular comparativa das CDKs cujas estruturas não foram determinadas experimentalmente até o momento. Resultados mostram que os aminoácidos que efetuam as interações com os ligantes no domínio quinase possuem em média 80 % de analogia, e 55% de identidade quando considerados os resíduos que interagem com o ATP. De forma parcial, o sítio ativo pode ser explorado para o desenho de inibidores, quando considerados alguns membros da família, como a CDK4 e CDK6. Esta informação foi fundamentada na análise comparativa do sítio de ligação do ATP, na presença dos fármacos: abemaciclibe, palbociclibe e ribociclibe. Além disso, estudos da interface CDK-ciclina demonstraram que o mecanismo de interação/ativação varia entre os membros da família, bem como a forma de recrutamento do substrato a ser fosforilado. Portanto, esta interface tem grande potencial de ser explorada para o desenvolvimento de inibidores.

ABSTRACT

Cyclin-dependent kinases (CDKs) are serine/threonine kinase, whose activity is modulated by interaction with cyclins and successive phosphorylations that occur in the complex. With studies since 1980, it is already known that CDKs act as one of the main families of eukaryotic protein kinases involved in the integration of extra and intracellular signals to modulate gene transcription and cell division. Related to a series of carcinogenic dysfunctions, mutations can result in the malfunction of this family. The study and characterization of CDKs can contribute to the progress of cancer therapy, since it opens the possibility of finding more and more specific inhibitors. The project aims to study the members of the CDK family, exploring the structural aspects in order to correlate with their physiological importance, played in each tissue. For this purpose, a comparative study is being carried out, focusing on the architecture of the active site, the extension of the polypeptide chain of each member, and the interface of interaction with the cyclin. Thus, the project was based on the search for information available in the databases and bibliography, and on the comparative molecular modeling of CDKs whose structures have not been experimentally determined so far. Results show that the amino acids that effect interactions with ligands in the kinase domain have an average of 80% analogy, and 55% identity when considering the residues that interact with ATP. Partially, the active site can be explored for the design of inhibitors, when considering some members of the family, with a CDK4 and CDK6. This information was based on the comparative analysis of the ATP binding site, in the presence of drugs: abemaciclib, palbociclib and ribociclib. In addition, studies of the CDK-cyclin interface have shown that the mechanism of interaction / activation varies between family members, as well as the form of recruitment of the substrate to be phosphorylated. Therefore, this interface has great potential to be explored for the development of inhibitors.

LISTA DE FIGURA

Figura 1. Visão geral de todos os membros da família das CDKs.....	10
Figura 2. Uma visão da estrutura terciária da CDK1.....	11
Figura 3. Uma visão geral das funções e localização de membros da família das CDKs na célula.....	12
Figura 4. Alinhamento de sequências realizado pelo Skylign (Wheeler et al., 2014).....	27
Figura 5. Alinhamento de sequências realizado pelo ESPript (Robert & Gouet, 2014).....	28
Figura 6. Árvore filogenética mostrando a origem e agrupamento de cada membro da família das CDKs humanas.....	29
Figura 7. Modo de interação dos fármacos de terceira geração.....	32
Figura 8. Análise do sítio catalítico da CDK6 em complexo com ribociclibe.....	34
Figura 9. Sítio catalítico da CDK6 em duas imagens distintas.....	36
Figura 10. Sobreposição dos complexos de pCDK2 com as ciclinas A, B e E.....	38
Figura 11. Alinhamento entre as ciclinas mostrando resíduos em contato nas interfaces pCDK2-ciclina.....	39
Figura 12. Contatos da hélice N-terminal da ciclina para três regiões distintas do pCDK2.....	41
Figura 13. Contatos realizados pelas hélices $\alpha 2$ e $\alpha 3$ das ciclinas com o segmento de ativação da pCDK2.....	43
Figura 14. Contatos das hélices $\alpha 3$, $\alpha 4$ e $\alpha 5$ das ciclinas para a região de voltas 40 e 70 da pCDK2.....	45
Figura 15. A esquerda na figura temos os sítios de ligação RLX para as ciclinas A, B e E respectivamente de cima para baixo.....	47
Figura 16. Comparação entre os complexos: CDK1-ciclina B (4Y72), pCDK2-ciclina A (1FIN) e CDK4-ciclina D (3G33)	50
Figura 17. (A). complexo pCDK2-ciclina A (1QMZ). (B) complexo CDK4-ciclina D (3G33). (C) complexo CDK2-ciclina A (1FIN)	51
Figura 18. Comparação da estrutura de CDK4-ciclina D com outros complexos de CDK-ciclina. (A) CDK4-ciclina D. (B) CDK9-ciclina T1.....	53

SUMÁRIO

1. INTRODUÇÃO.....	9
1.1. CDK: Evolução, estrutura e função.....	9
1.2. Papel da ciclina na ativação das CDKs.....	13
1.3. Bancos de dados e bioferramentas	14
1.4. Os inibidores de CDK	16
2. OBJETIVOS.....	19
3. METODOLOGIA	20
3.1. Exploração nos bancos de dados.....	20
3.2. Análise das sequências e alinhamentos múltiplos	20
3.2.1 Multalin e ClustalOmega.....	20
3.2.2 Skyline (Wheeler et al., 2014).....	21
3.3. Análise estrutural das CDKs	21
3.4. Modelagem molecular do domínio quinase das CDKs.....	22
3.4.1 Swiss-Model	22
3.4.2 Modeller (Webb & Sali 2016)	22
3.4.3 Validação dos modelos gerados	23
3.5. Investigação do N- e C-terminal das CDKs	23
3.6. Análise das interações proteína-ligante	24
3.7. Análise das interações proteína-proteína	24
3.8. Análise da superfície eletrostática	25
4. RESULTADOS E DISCUSSÕES	26
4.1. Alinhamentos Múltiplos.....	26
4.1.1. Skyline.....	26
4.1.2 Análise comparativa dos elementos de estrutura secundária utilizando a ferramenta ESript.....	27
4.2. Modelagem molecular comparativa	30
4.3. Investigação do N- e C-terminal das CDKs	31
4.4. Interações proteína-ligante	31
4.4.1 Modo de interação entre os ligantes: abemaciclib, palbociclib e ribociclib e a CDK6.	32
4.4.2 Interação entre ribociclib e a CDK6.	33
4.4.3 Análise da superfície eletrostática	35
4.5. Interface CDK-ciclinas	37
4.5.1 pCDK2 em complexo com as ciclinas: A, B, E.....	37
4.5.2 Aspectos estruturais: Diferenças e semelhanças entre os complexos de pCDK2....	39

4.5.3 Interações com o N-terminal das ciclinas.	40
4.5.4 Interações com a alça de ativação	42
4.5.5 Interações com a volta do resíduo 40 e volta do resíduo 70	44
4.5.6 Sítio de recrutamento RXL Φ	46
4.6 Análise dos demais complexos CDK/ciclina: um novo ‘ramo’ na família das CDKs. ...	48
4.6.1 Complexos ‘tipo’ pCDK4-ciclina D: estruturas e funções.	52
5. CONCLUSÃO	55
PERSPECTIVAS PARA TRABALHOS FUTUROS	57
REFERÊNCIAS	58

1. INTRODUÇÃO

1.1. *CDK: Evolução, estrutura e função*

As Quinases dependentes de Ciclinas (Cyclin dependent kinases - CDK) são enzimas serina/treonina quinases, e possuem sua atividade catalítica modulada pela interação com ciclinas e sucessivas fosforilações que ocorrem no complexo. As CDKs desempenham importantes papéis regulatórios no controle da divisão celular, modulam a transcrição em resposta a alterações extra e intracelular, além de desempenharem uma série de outros papéis, envolvidas com: metabolismo, renovação de células-tronco, funções neuronais e espermatogênese (Singh & Wu, 2019). Um dos primeiros trabalhos que fundamentou a importância das CDKs envolveu estudos genéticos em organismos modelo, como leveduras e rãs (Malumbres, 2014). Este trabalho estabeleceu a importância das CDKs na promoção de transições ao longo do ciclo celular, bem como foi capaz de demonstrar que a subunidade catalítica, a CDK, deve se associar a uma subunidade reguladora, a ciclina, para desempenhar sua função quinase. Dos estudos pioneiros realizados na década de 1980, até o momento, ficou claro que as CDKs atuam como uma das principais famílias de proteínas quinases eucarióticas envolvidas na integração de sinais extracelulares e intracelulares para modular a transcrição gênica e a divisão celular.

O genoma humano é capaz de codificar 21 CDKs considerando duas isoformas da CDK11, denominadas de CDK11A e CDK11B (Malumbres & Barbacid, 2009). É possível agrupar as CDKs em sub-famílias, de acordo com a função desempenhada e a semelhança estrutural, como o grupo CDK1, CDK2, CDK4 e CDK6 envolvidas diretamente na regulação do ciclo celular que apresentam múltiplos sítios de ligação de ciclinas, e as relacionadas com a regulação transcricional: CDK7, CDK8, CDK9, CDK11, CDK12 e CDK13, que apresentam apenas um sítio de ligação de ciclina. As CDKs variam em tamanho de aproximadamente 250 resíduos de aminoácidos, abrangendo apenas o domínio catalítico, até proteínas que podem chegar a mais de 1.500 resíduos, com extensões nos terminais amino e/ou carboxi de comprimentos variáveis (Figura 1). Como as demais proteínas quinases, as CDKs possuem uma estrutura de dois lobos. O lobo amino-terminal contém β -folhas, enquanto o lobo

carboxi-terminal é rico em α -hélices, e o sítio ativo está localizado entre os dois lobos (Li et al., 2015). O lóbulo N-terminal contém um elemento inibidor rico em glicina (G-loop) e uma única hélice principal - a hélice C (contendo a sequência PSTAIRE em CDK1).

Figura 1. Visão geral de todos os membros da família das CDKs, distribuídas de acordo com as características estruturais conservadas ao longo do processo evolutivo. Em vermelho é mostrado o domínio quinase conservado em todos os membros.

Fonte: (Malumbres, 2014)

O processo de regulação das CDKs ocorre predominantemente por meio do controle de produção e destruição da ciclina, já que a ligação da ciclina, na maioria das CDKs estudadas até o momento, desloca um elemento denominado alça de ativação (figura 2), expondo o local de ligação do substrato, realinhando os resíduos fundamentais do sítio ativo que inicia a atividade quinase (Zhao et al., 2019). Bem como os efeitos da ligação com a ciclina, a fosforilação também coordena a atividade das CDKs em resposta a vários estímulos. Na forma monomérica livre de ciclina, a fenda catalítica da CDK é fechada pela alça de ativação, impedindo a atividade enzimática. Além disso, o segmento de ativação no lóbulo C é parcialmente desordenado. Quando a ciclina se liga, há uma mudança conformacional que afasta

a volta T para longe do sítio ativo, reorientando a hélice amino-terminal. A fosforilação do resíduo Thr da volta T é estabilizada por 3 resíduos de Arginina, que a mantêm longe do sítio de ligação do substrato (Nelson & Cox, 2014). Como ilustrado na figura 2 os membros da família CDK tem um núcleo catalítico que contém um bolsão de ligação ao ATP, nas proximidades da volta T160, um domínio de ligação a ciclina ilustrado pela hélice α C, e o alça de ativação. Os conjuntos dessas unidades são responsáveis pela ativação e desativação das CDKs.

Figura 2. Uma visão da estrutura terciária da CDK1, com as principais estruturas secundárias destacadas da seguinte forma: Azul: Alça rica em glicina, amarelo: Alça móvel, roxo: hélice α C (PSTAIRE), vermelho: alça de ativação, volta T (T160).

Fonte: (Wood e Endicott, 2018)

Apesar das diversas correlações, e mesmo que estreitamente relacionadas, há peculiaridades estruturais e sequências distintas entre as CDKs, além de diferentes formas de interações com suas parceiras regulatórias. As diferenças podem ser entendidas pela função, localização, modo de regulação, todos intrínsecos à sua estrutura e a rede de sinalização que cada CDK especificamente está conectada. A figura 3 ilustra a atuação de uma série de CDKs na célula, com suas respectivas ciclinas.

A figura 3 demonstra as diversas atuações, bem como algumas vias que as CDKs participam. As CDK1, CDK2, CDK4 e CDK6 são conhecidas pela regulação das diferentes fases de divisão do ciclo celular. Essas atividades são pelo menos parcialmente mediadas pelo controle de múltiplos fatores de transcrição (TFs) ou elementos regulatórios, como a proteína do retinoblastoma (Rb) (McGrath et al., 2013). No controle da transcrição estão presentes as CDKs 10 e 11, que atuam na fosforilação dos TFs, receptores hormonais e outros reguladores associados. A CDK10 tem sido relatada como o regulador do fator de transcrição Ets2 e modula sua atividade de transativação. Além disso, a sinalização CDK10/Ets2/c-RAF tem sido demonstrada como um ponto importante, determinante na resistência à terapia endócrina para câncer de mama. As CDKs 7, 9 e a CDK12, que também regulam a transcrição, fosforilam diretamente o C- terminal da RNA polimerase II, modulando as diferentes fases de geração de transcritos. Os membros da subfamília da CDK14, como a própria CDK14 ou CDK16, são ativadas na membrana pela ciclina Y e também participam em muitas vias diferentes, como na sinalização dependente ou transdução de sinal (Cao et al., 2014). Importante ressaltar que a figura 3 apenas ilustra alguns

dos possíveis parceiros das CDKs, pois muitas CDKs, como a CDK1 podem se ligar a mais de uma ciclina, além de fosforilar mais de 100 substratos durante sua atuação na fase mitótica (Liao et al., 2016).

1.2. Papel da ciclina na ativação das CDKs

As ciclinas, além de executarem importante papel no controle progressivo da divisão celular, possuem habilidade de ligar, ativar e aumentar a afinidade dos substratos que se associam às CDKs. A transição das células no início do estágio G1 da divisão celular é coordenada pelas atividades dos complexos CDK4 e CDK6 que são formados após a expressão da ciclina D. CDK4-ciclinaD e CDK6-ciclinaD estimulam a expressão dos genes necessários para a progressão de G1 através da fosforilação do produto do gene do retinoblastoma, pRb, sendo este o principal substrato das CDKs nessa etapa da divisão celular. Sequenciais fosforilações do pRb, dependentes de CDK, resultam na dissociação de complexos que regulam a atividade de transcrição do pRb. Ao mesmo tempo, a expressão de genes cujos produtos são essenciais para a progressão de G1, são estimulados (Macaluso et al., 2006). A fosforilação do pRb é inicialmente realizada pelas CDK4 e CDK6, e ocorrem em múltiplos sítios. Posteriormente, a CDK2, nos complexos com as ciclinas parceiras, A e E, também realizam fosforilações no pRb (Kitagawa et al., 1996).

Um modelo de regulação da CDK por ligação à ciclina, CKI e por fosforilação, foi desenvolvido a partir de uma série de estruturas de complexos CDK2-ciclina A (Brown et al., 1999). Essas estruturas também revelaram as interações moleculares que determinam os requisitos necessários à sequência do substrato, nos locais de ligação do substrato-peptídeo à CDK, ao redor do local da fosfotransferência e no local de recrutamento da ciclina. Como já discutido, a CDK2 em sua forma monomérica é inativa como resultado da disposição da hélice C e do segmento de ativação. A ligação à ciclina A e a fosforilação em T160 resultam em rearranjos no local ativo da CDK2 que orientam corretamente interações chaves entre o ATP e os resíduos catalíticos, e criam o local de ligação ao substrato peptídico. Para as CDKs que se ligam às ciclinas A, B, E e D, interações adicionais na ciclina na região conhecida como "local de recrutamento" contribuem para a seletividade do substrato (Adams et al., 1999).

Essa sequência de resíduos que compõem o "local de recrutamento" é um trecho hidrofóbico que liga o motivo R/KXLX[0-1] Φ (sendo o X a representação para qualquer resíduo e Φ para qualquer resíduo hidrofóbico), presente em determinados

substratos recrutados e na classe de proteínas CKIs, sigla em inglês para designar proteínas que interagem com CDKs e/ou proteínas inibidoras de quinases (Cip/Kip). Todavia, fatores adicionais executam um importante papel na atividade das CDKs, e para alguns membros dessa família, como a CDK4, parecem ter uma influência maior, visto a especificidade de seus substratos. A determinação de estruturas como CDK5-p25, CDK6-ciclina V, e CDK9-cyclin T revelaram diferenças estruturais significativas entre as CDKs e o modo de interação com as ciclinas, sugerindo diferentes mecanismos de ativação e regulação com a expansão da família (Baumli et al., 2008; Schulze-Gahmen & Kim, 2002; Tarricone et al., 2001).

A CDK4 é uma das diversas quinases cujo sistema chaperona Hsp90 é um passo essencial na ativação (Pearl, 2005). Além disso, diferente de complexos como CDK1-ciclina e CDK2-ciclina, a CDK4 e a ciclina D não se associam de imediato, necessitando de uma etapa a mais para que o complexo se torne ativo. Outros estudos também propõem uma relação mais íntima entre o complexo CDK4-ciclina D e o grupo de moléculas CKIs. A proposta feita é que membros da família Cip/Kip: p21, p27, p57, promovam a associação entre a CDK4 e a ciclina D (Errico et al., 2010).

Com tudo, fica evidente que a análise estrutural da família das CDKs deve levar em consideração não somente sua forma monomérica, mas como também seus parceiros regulatórios. Fica cada vez mais evidente também, que as ciclinas desempenham papel fundamental na diferenciação de substratos, e para entender de forma aprofundada a maneira que os substratos são recrutados e fosforilados pela atuação das CDKs, é importante entender como relacionam-se as diferentes proteínas presentes neste complexo.

1.3. Bancos de dados e bioferramentas

O avanço da computação aliado as descobertas do campo bioquímico trouxeram diversas ferramentas e acesso a dados de uma forma muito facilitada. Entre as diversidades que abrange esse campo de estudo, como assuntos que se destacam, cabe o armazenamento e o acesso aos dados de biologia molecular em bancos de dados, em particular as sequências de ácidos nucleicos e aminoácidos e suas respectivas anotações, e os algoritmos para análises destes dados (Gupta, 2017). Os bancos de dados de genoma representam hoje uma das principais ferramentas de suporte para os biólogos moleculares e geneticistas. A pesquisa nesta área é alimentada pelos cadastros de sequências e de algumas anotações

relacionadas, possibilitando consultas nestes bancos a fim de levantar dados para análises biológicas. Destaca-se entre as análises a comparação de sequências e o descobrimento de novos genes, funções e características de uma nova sequência. Além disso, o desenvolvimento e estudos de novos algoritmos que façam análises eficientes em todo este volume de dados é um assunto que tem se tornado cada vez mais explorado.

Atualmente existem não somente os bancos de dados de sequências de nucleotídeos (DNA) e de aminoácidos (proteínas), mas também inúmeros outros com informações bem específicas, como organismos especiais (Eukariotic, *Escherichia coli* e *Drosophila*), biossequências específicas, como exemplo: tRNA e rRNA, enzimas, mutações, famílias de biossequências (filogenia), etc. Além disso, bancos que guardam estruturas tridimensionais das biossequências, como por exemplo o Protein Data Bank (PDB) (Risler, 2002). É possível destacar os seguintes BDBM (Blast data base manager) como os maiores atualmente: GenBank Sequence Database, EMBL Nucleotide Sequence Database, Genome Sequence Database (GSDB), Genome Database (GDB), etc. Nestes bancos de dados estão armazenadas anotações relevantes, além das próprias sequências, o que dá um acesso e possibilidade imensurável de exploração e análise de dados.

O alinhamento de sequências é uma forma de organizar estruturas primárias de DNA, RNA ou proteína para identificar regiões similares que possam ser consequência de relações funcionais, estruturais ou evolucionárias entre elas (Diniz & Canduri, 2017). A evolução das bioferramentas e seus algoritmos possibilitam alinhamentos múltiplos envolvendo até 4 mil sequências, no caso do Clustal Omega. Nesse tipo de alinhamento, as sequências são correlacionadas de forma a encontrar a melhor correspondência, seja local ou global. Alinhamentos múltiplos são frequentemente usados na identificação de regiões de sequências conservadas durante a evolução. Essas sequências, ou regiões conservadas podem ser usadas em conjunto com informações estruturais e mecânicas para localizar os sítios catalíticos de enzimas, e aprofundar seus estudos. Todas as variações do software Clustal alinham sequências usando uma heurística que progressivamente constrói um alinhamento de múltiplas sequências a partir de uma série de alinhamentos em pares. Esse método funciona analisando as sequências como um todo. Uma árvore guia é então calculada a partir dos escores das sequências na matriz e, em seguida, usada

para construir o alinhamento de múltiplas sequências, pareando progressivamente as sequências em ordem de similaridade (Sievers et al., 2011).

Uma outra importante ferramenta para a construção de estruturas moleculares, a modelagem molecular por métodos computacionais, surgiu como uma alternativa, especialmente após o desenvolvimento de programas capazes de calcular a estrutura com uma fidelidade adequada entre velocidade e precisão. A caracterização funcional, uma frequente tarefa no campo da bioquímica, pode ser dificultada pela falta de uma estrutura tridimensional precisa (3-D) da proteína estudada. Na ausência de uma estrutura determinada experimentalmente, a modelagem comparativa ou de homologia geralmente fornece um modelo 3D útil para uma proteína relacionada a pelo menos uma estrutura de proteína conhecida (Chojnacki et al., 2017). Dessa forma, podemos definir brevemente a modelagem molecular como um conjunto de ferramentas para a construção, edição e análise de sistemas moleculares complexos. Um programa que se destaca na área é o Modeller, um programa de computador para modelagem comparativa de estruturas proteicas. Através de cálculos automáticos é capaz de criar um modelo de proteína, bem como: definir dobras, alinhar duas ou mais sequências de proteínas, gerar árvores filogenéticas, refinar alças e voltas, bem como inserir ligantes em estruturas proteicas (Webb & Sali, 2017). Outro programa, que também se destaca na modelagem homóloga de proteínas, é o Swiss-model. Uma das principais diferenças entre o Modeller e o Swiss-model é que o último é um serviço oferecido pela web, totalmente automatizado. Mais especificamente, ele orienta o usuário na construção de modelos de homologia de proteínas através de quatro etapas. A primeira seria a busca por um modelo que servirá de base para a modelagem. O segundo passo é o alinhamento da sequência alvo com o modelo selecionado, seguido pelo terceiro e quarto passos que tratam da criação do modelo e a avaliação da qualidade do mesmo, respectivamente (Biasini et al., 2014).

1.4. Os inibidores de CDK

Atualmente projetar pequenas moléculas biologicamente ativas, como inibidores de CDKs, está se tornando uma solução cada vez mais legítima para o controle da proliferação celular. Até o momento, alguns inibidores de CDKs foram descobertos, e podem ser subdivididos em inibidores competitivos e não competitivos, de acordo com seus locais de ligação ao sítio alvo (Esposito et al., 2013). Entre eles,

os inibidores competitivos do ATP são mais amplamente investigados, e alguns deles já foram ou estão sendo avaliados, a partir das várias etapas dos ensaios clínicos.

Os inibidores das enzimas CDK4 e CDK6 são utilizados para o tratamento do subtipo mais comum de câncer de mama, chamado HR+/HER2- (receptor hormonal positivo/receptor do fator de crescimento epidérmico humano 2 negativo). Este tipo de tratamento interrompe a atividade de enzimas, no caso as CDKs, que neste caso, apresentam-se em maior concentração do que numa célula sadia, ocasionando na proliferação de células. Entre os inibidores já aprovados pela ANVISA estão, o palbociclibe, o abemaciclibe e o ribociclibe (Paula et al., 2019).

O palbociclibe foi aprovado em primeira linha e segunda linha metastática, após estudos de câncer de mama receptores hormonais positivos/Her-2 negativo. Associado à hormonioterapia (letrozol ou fulvestranto), em primeira e segunda linha metastática em Câncer de mama, demonstrou excelentes taxas de controle tumoral e mais do que duplicou o tempo de vida livre de doença. O palbociclibe já está aprovado nos EUA desde fevereiro de 2015 e foi aprovado no Brasil pela ANVISA em 2018. O ribociclibe é também outro representante desta nova classe de medicamento. Ele pode ser usado em pacientes com Câncer de mama metastático HR+/HER2- associado a inibidor de aromatase ou fulvestranto, em 1ª e 2ª linhas. Combinado com tratamento endócrino convencional apresentou melhora significativa a sobrevida global de mulheres mais jovens com câncer de mama avançado positivo para receptor hormonal HR+ em comparação com o tratamento endócrino isolado. Assim como o palbociclibe, o ribociclibe está aprovado no Brasil desde 2018. O abemaciclibe foi lançado nos EUA 2017 e logo incorporado a essas novas opções terapêuticas pelo Brasil também em 2018. O tratamento oral é indicado em combinação com um inibidor da aromatase como terapia endócrina inicial, ou após terapia endócrina, no cenário de doença avançada/metastática. Sua administração pode ser feita isolada, após progressão da doença depois do uso de terapia endócrina e quimioterápicos anteriores. Sua segurança e eficácia como tratamento autônomo foram analisadas em um estudo com 132 pacientes com câncer de mama HR+ para HER2- que progrediu após tratamento com terapia endócrina e quimioterapia após o câncer ter se tornado metastático (O'Leary et al., 2016). Infelizmente, devido ao alto custo, estes fármacos ainda são pouco prescritos no Brasil, realidade que deve ser mudada nos próximos

anos com o avanço dos estudos em torno destes alvos potenciais e descoberta de novas alternativas para o tratamento do câncer.

2. OBJETIVOS

Analisar individualmente e comparativamente os membros da família das quinases dependentes de ciclinas (CDK1-CDK20), com o objetivo de traçar o diferencial, a partir do detalhamento estrutural de cada membro, e analisar interações intramoleculares que possam colaborar para o estabelecimento da especificidade estrutural de cada uma delas.

Quanto aos objetivos específicos:

- Modelar as CDKs cujas estruturas cristalográficas não estão determinadas, na ausência e na presença do ATP, e dos fármacos abemaciclibe, palbociclibe e ribociclibe. Esta análise irá colaborar para o entendimento da especificidade destes fármacos pelas CDK4 e pela CDK6 (depositada no PDB), e dessa forma, oferecer pistas que ofereçam subsídios para o planejamento de inibidores das demais CDKs;
- Realizar análises estruturais de cada membro (CDK1-CDK20), em busca de informações que colabore para o estabelecimento de características individuais, principalmente no entorno do sítio catalítico;
- Investigar a superfície de interação entre os complexos CDKs-ciclinas com estruturas cristalográficas resolvidas e depositadas no PDB, em busca de informações sobre o modo de ativação e de recrutamento de substratos, e aspectos estruturais importantes para o estabelecimento de distinções dentre os membros da família.

3. METODOLOGIA

3.1. *Explorando os bancos de dados*

Inicialmente realizou-se uma busca pelas sequências de aminoácidos, estruturas tridimensionais e informações sobre as funções relacionadas à família das quinases dependentes de ciclinas humanas (CDKs). No decorrer do trabalho, também foram analisadas as sequências de nucleotídeos das CDKs humanas, a fim de suprir informações do N- e C-terminais das CDKs de maior massa molecular. Os dados foram obtidos a partir de bancos de dados internacionais, com informações disponibilizadas gratuitamente por toda extensão da web. No NCBI (*National Center for Biotechnology information*), foram pesquisados por todos os membros da família, desde a CDK1 até a CDK20 humana. Como contraponto, o Expasy foi também utilizado para a busca dos dados, servindo de base comparativa, visto que são dois bancos de dados bem completos e conceituados, e a mesma sequência/estrutura pode ser depositada por diferentes pesquisadores. O critério de seleção das estruturas primárias das CDKs levou em consideração as proteínas depositadas que possuíam maior número de aminoácidos, ou seja, que possuísem a sequência completa das proteínas, e as principais informações obtidas foram a sequência FASTA e a localização celular de cada membro da família, resultando assim em 20 códigos/sequências diferentes.

As estruturas terciárias e informações correlacionadas foram obtidas no PDB (rcsb.org), e como critério, foram consideradas as diferentes conformações de cada CDK depositada. Dessa forma deu-se prioridade para estruturas que estivessem tipicamente na conformação ativa, ou seja, CDKs em complexo com a sua ciclina, preferencialmente com a alça de “ativação” fosforilada. Importante ressaltar que algumas CDKs possuem poucas ou até mesmo uma única estrutura depositada, e para esses casos, o critério não foi atendido.

3.2. *Análise das sequências e alinhamentos múltiplos*

3.2.1 *Multalin e ClustalOmega*

Quando as sequências primárias estão disponíveis em um banco de dados, um alinhamento múltiplo torna-se a melhor opção. Levando isso em consideração, foi utilizado o “software” Multalin by Florence Corpet, para a realização, primordialmente,

do alinhamento múltiplo das 20 CDKs, cujas sequências FASTA foram previamente obtidas, como descrito na seção anterior. Nessa ferramenta também houve a possibilidade de selecionar o formato da sequência trabalhada, permitindo: Genbank, Multialin-Fasta, EMBL-Swissprot além do Auto, opção utilizada nesse projeto.

Outra ferramenta utilizada para a realização do alinhamento múltiplo foi o ClustalOmega (ebi.ac.uk/Tools/msa/clustalo/). Nesta, foi possível, pelos dados gerados no alinhamento múltiplo, produzir uma árvore filogenética, que pôde ser utilizada para produzir novos agrupamentos, de acordo com uma familiaridade decorrente da relação evolutiva dessas proteínas, muito importante na análise comparativa.

3.2.2 *Skyalign (Wheeler et al., 2014)*

O alinhamento múltiplo pode ser realizado utilizando, tanto a sequência primária na íntegra, dito alinhamento global, quanto com somente determinada região dessa sequência, o alinhamento local. Utilizando o Skyalign, um alinhamento local da região do domínio quinase das CDKs foi realizado. Este alinhamento demonstra de uma maneira mais simplificada, a conservação de cada posição da cadeia polipeptídica.

Utilizando os formatos HMMER, alinhamentos Stockholm ou FASTA, é possível obter o alinhamento, o qual roda no próprio site (Skyalign.org), e pode ser baixado no formato de pdf ou imagem, tanto dinâmica quanto estática.

3.3. *Análise estrutural das CDKs*

Para a visualização e análise das estruturas secundárias e terciárias das CDKs, foi utilizado o programa Pymol 2.0.7 (pymol.org/2/). Nele é possível explorar a estrutura tridimensional e realizar análises da interação proteína-ligante, proteína-proteína, bem como sobrepor diferentes CDKs para realizar análises comparativas, importante principalmente no entorno do sítio ativo de ligação do ATP e na interface de interação com a ciclina, que poderá oferecer informações cruciais para o presente trabalho.

Uma das formas de analisar um alinhamento de sequências é correlacionar o que se observa na estrutura primária com as estruturas secundárias correspondentes.

Uma ferramenta de análise que permite fazer esse tipo de comparação é a ESPript 3.0 (<http://esprict.ibcp.fr/ESPript/ESPript/>). Para gerar os alinhamentos foi necessário fazer o upload de dois arquivos de entrada. O primeiro arquivo é correspondente a um alinhamento múltiplo, no caso, foi utilizado o alinhamento múltiplo local, do domínio quinase das 20 CDKs. O segundo arquivo necessário é uma estrutura tridimensional da primeira proteína que aparece no alinhamento do primeiro arquivo. Ajustes finos podem ser realizados pela escolha de parâmetros, que no caso do presente trabalho, foram escolhidos de forma automática.

3.4 Modelagem molecular do domínio quinase das CDKs

3.4.1 Swiss-Model

Levando em conta o objetivo do trabalho, e considerando que nem todas CDKs apresentam estrutura cristalográfica resolvida e depositada no Protein Data Bank (PDB), ferramentas de modelagem molecular comparativa foram utilizadas para gerar modelos estruturais, e assim ser possível realizar elucidações envolvendo todas as 20 CDKs humanas. O ponto de partida foi realizar uma seleção das CDKs que não possuem estruturas tridimensionais elucidadas, bem como avaliar as estruturas cristalográficas das CDKs já existentes, considerando a resolução e interações intra e intermoleculares realizadas. Seleccionadas as estruturas cristalográficas, suas coordenadas foram então usadas para a realização da modelagem molecular comparativa, utilizando inicialmente a ferramenta SwissModel do Expasy (swissmodel.expasy.org), um servidor de modelagem por homologia totalmente automatizado e eficiente. Adicionalmente, este programa oferece uma ferramenta de busca por templates (moldes), que é listado em ordem decrescente. Importante mencionar que os moldes selecionados possuem identidade igual ou maior que 30%, sendo desta forma considerada pelo programa SwissModel, como um potencial molde.

3.4.2 Modeller (Webb & Sali 2016)

Ainda utilizando os mesmos critérios de seleção, uma segunda ferramenta para realização de modelagem molecular comparativa, Modeller, foi empregada a fim de fundamentar a qualidade dos modelos obtidos. Essa ferramenta não só permitiu a modelagem molecular comparativa das CDKs sem estruturas no PDB, como permitiu

também a inclusão de diferentes ligantes nas CDKs, colaborando para uma melhor abordagem do estudo estrutural. Todos os modelos gerados foram visualizados e analisados utilizando Pymol, ferramenta já citada anteriormente.

3.4.3 Validação dos modelos gerados

Para a validação dos modelos foi utilizada a ferramenta SAVES v. 5.0 (<https://servicesn.mbi.ucla.edu/SAVES/>), que permitiu avaliar a qualidade dos modelos gerados a partir da obtenção de um gráfico de Ramachandran. Para isso, foram selecionados para cada CDK, 10 candidatos de um total de 100 modelos gerados, de acordo com a energia de Gibbs de cada modelo. Os modelos selecionados foram analisados pela ferramenta citada acima, e dessa forma, avaliando os desvios dos ângulos psi e fi, foi possível selecionar os melhores modelos baseados na qualidade estereoquímica e menor energia.

3.5 Investigação do N- e C-terminal das CDKs

Considerando que as CDK7, CDK8, CDK9, CDK11B e de CDK12 à CDK20 apresentam maior massa molecular, devido aos seus N- e/ou C-terminais serem extensos com relação às demais CDKs, foi realizada uma busca por informações estruturais destas regiões, as quais não pertencem ao domínio quinase. Desde que nenhuma informação foi obtida a partir das sequências primárias das proteínas utilizando a ferramenta BLASTp do NCBI, estas análises basearam-se na sequência dos RNAm das CDKs.

Assim, utilizando a ferramenta BLASTn do NCBI, e a partir das sequências do DNA correspondente a expressão das regiões N- e C-terminais das CDK7, CDK8, CDK9, CDK11B e de CDK12 à CDK20, buscou-se por sequências de nucleotídeos distintas das CDKs. Importante ressaltar que para a busca, foi utilizada a opção “expressed sequence tags” (est), a qual compreende uma sequência curta de cDNA que ocorre apenas uma vez no genoma e cuja localização e sequência base são conhecidas. Esta foi uma opção, pois nenhuma outra resultou em resultados passíveis de análises.

3.6 *Análise das interações proteína-ligante*

Inicialmente foi realizada a análise das interações intermoleculares das CDKs depositadas no PDB, porém, devido a diversidade lá encontrada, foi realizada análises de CDKs ligadas a diferentes ligantes, o que pode levar a falhas de interpretação do modo de interação, uma das propostas do presente trabalho.

Neste sentido, as análises da interação proteína-ligante foram conduzidas para os modelos das CDKs com ATP, e com os fármacos abemaciclibe, palbociclibe e ribociclibe, os quais foram gerados utilizando o Modeller versão 9.22. Esta abordagem possibilitou traçar melhores comparativos do modo de interação proteína-ligantes. Os modelos foram analisados pelo Pymol, as quais permitiram inferir interações intra e intermoleculares no sítio catalítico das CDKs na presença dos diferentes ligantes agora padronizados.

3.7 *Análise das interações proteína-proteína*

Estudos da interação entre as CDKs e suas respectivas parceiras ciclinas foram realizados. Alguns substratos peptídicos foram incluídos nas análises como substratos proteicos, a fim de enriquecer e embasar a discussão sobre as interações entre cada CDK e sua ciclina. Diferente da abordagem do estudo proteína-ligante, não foi objetivo do trabalho modelar as ciclinas, e portanto, foram analisadas estruturas cristalográficas depositadas nos bancos de dados, e portanto, as análises foram focadas naquelas CDKs cuja estruturas estavam disponíveis no PDB.

Selecionadas as estruturas e suas coordenadas no PDB, o estudo foi conduzido seguindo as análises anteriores já citadas, porém agora focadas nas ciclinas. No caso, foram realizados alinhamentos entre ciclinas homólogas, e diferentes tipos de ciclina. Foram analisados os resíduos que compõem a superfície de interação CDK-ciclina e comparados em dois conjuntos diferentes: mesma CDK com diferentes ciclinas, e para os casos de diferentes CDKs com a mesma ciclina. O estudo envolveu a análise do modo de ativação que cada complexo CDK-ciclina pode desencadear, procurando classificar cada CDK de acordo com os diferentes modos de ativação, observados e descritos no presente trabalho.

3.8 Análise da superfície eletrostática

Com o objetivo de analisar a superfície molecular e eletrostática das proteínas, e compará-las principalmente no entorno do sítio catalítico e da superfície de contato CDK-ciclina, foi utilizado o programa Pymol. Em paralelo, foram consideradas as interações dos resíduos do sítio com os ligantes, bem como interações intermoleculares das diferentes CDKs com suas respectivas ciclinas.

Outra importante informação estrutural que se pode observar em uma estrutura proteica é a superfície eletrostática, capaz de indicar cavidades e regiões mais ou menos carregadas, mostrando possíveis posições que os ligantes e/ou substratos podem ocupar. Levando isso em consideração, imagens da superfície molecular foram geradas pelo Pymol, e traçou-se um paralelo entre as cargas encontradas na região do sítio ativo e quais aminoácidos ali presentes participam das interações com os ligantes utilizados no trabalho.

4. RESULTADOS E DISCUSSÕES

4.1. Alinhamentos Múltiplos

4.1.1. Skylign

Como já mencionado, o alinhamento de sequências de proteína ou de nucleotídeos é uma importante ferramenta de análise para traçar comparativos, a fim de observar semelhanças e diferenças entre as estruturas. Com o intuito de tornar mais fácil a visualização e consequentemente, as análises, alinhamentos já realizados pelo ClustalOmega foram utilizados para gerar a imagem ilustrada na figura 4. A imagem torna o alinhamento menos poluído, por representar a conservação de cada posição por letras de tamanho correspondente à probabilidade de ocupação. Quanto maior, mais conservado. Assim sendo, no caso do alinhamento local do domínio quinase das 20 CDKs, a posição do aminoácido alanina (A) da região **VALK**, por exemplo, é idêntica nas 20 CDKs, assim como a lisina (K, **VALK**) (Figura 4). Dessa forma, é possível observar com maior clareza os resíduos de aminoácidos de cada posição, bem como fazer estimativas a respeito da ocupação da mesma. Estes resultados mostram que o domínio quinase de todos os membros são cerca de 28% conservados, com grande contribuição dos resíduos do sítio de ligação ao ATP.

Em alguns casos a posição pode conter mais de uma letra, como exemplo a valina (V) da mesma região **VALK**. Nela, encontra-se tanto a valina, como a tirosina (Y). Isso é devido ao fato de que ambas, CDK8 e CDK19 possuem tirosina no lugar da valina, a qual é conservada nas demais CDKs. Neste caso é possível dizer então, que a valina é predominante nesta região (Figura 4). Neste sentido, pode-se observar que ambas, CDK8 e CDK19, estão relacionadas evolutivamente, o que também pode ser observado pela presença da extensa alça em torno do resíduo K120, comum entre ambas (Figuras 4 e 5).

Figura 4. Alinhamento de sequências realizado pelo Skylign (Wheeler et al., 2014). O trecho apresentado corresponde ao alinhamento múltiplo das 20 CDKs. A seta laranja indica o resíduo de alanina da região VALK (caixa). A segunda caixa destaca a região da hélice C, conservada em todas as CDKs, e no círculo amarelo, a região de alça comum entre as CDK8 e CDK19.

4.1.2 Análise comparativa dos elementos de estrutura secundária utilizando a ferramenta ESPrpt

Com o objetivo de analisar a analogia dos elementos de estrutura secundária das 20 CDKs, um alinhamento múltiplo foi realizado utilizando o programa ESPrpt (Robert & Gouet, 2014), que através do alinhamento local do domínio quinase (cerca de 300 resíduos de aminoácidos) de todas as CDKs, e uma estrutura de referência (CDK8, PDB 3rgf), foi possível correlacionar a estrutura primária com a estrutura secundária, e explorar as regiões que divergem dentre todas, ou parcialmente (Figura 5).

Figura 5. Alinhamento de sequências realizado pelo ESPrict (Robert & Gouet, 2014). Alinhamento múltiplo correspondente ao domínio quinase das 20 CDKs, tendo como referência a CDK8. Em vermelho estão destacados os resíduos de aminoácidos conservados e idênticos (caixas vermelhas). O círculo em amarelo indica a região de alça comum entre as CDKs 8 e 19, e o círculo em verde, os resíduos comuns conservados nas CDK4 e CDK6.

A figura 5 mostra que a região N-terminal das CDKs, tendo como referência a CDK8, apresenta uma hélice conservada ($\alpha 1$) para as CDK1, CDK6, CDK7, CDK8, CDK9, CDK12 e CDK13, e parcialmente conservada para as demais, mesmo não havendo identidade da sequência primária. A hélice $\alpha 2$ é exclusiva para as CDK8 e CDK19, enquanto as fitas $\beta 1$ e $\beta 2$ são conservadas em todas as CDKs, com alta identidade sequencial (mostrado em vermelho na Fig. 5), característica do domínio N-terminal das CDKs, predominantemente constituído de folhas β . Lembrando que apenas a região N-terminal está sendo mostrada. A análise foi realizada para todo o domínio quinase e foi observado que apesar dos elementos de estrutura secundária serem conservados em todo o domínio quinase, há regiões distintas, principalmente de alças, que podem ser exploradas. Como exemplos estão as CDKs 8 e 19, e as CDKs 4 e 6, que apresentam alças mais extensas localizadas nas proximidades do sítio de ligação ao ATP (indicadas nos círculos da Figura 5), diferente das demais CDKs.

A Figura 6 mostra a árvore filogenética feita a partir do ClustalOmega, que corrobora para a análise do alinhamento da Figura 5. Observa-se, portanto, o agrupamento entre os membros da família das CDKs, como exemplo, vemos a proximidade entre a CDK8 e a CDK19, assim como entre as CDKs 4 e 6 e entre as CDK9, CDK12 e 13. Neste dendograma está representado à direita, 'scores' referentes a distância evolucionária das sequências de cada CDK em relação ao ancestral comum. Nesse sentido observa-se que a CDK8, por exemplo, possui uma relação mais estreita com o ancestral dessa família, quando comparado com a CDK9, mais distante.

Figura 6. Árvore filogenética dos membros da família das CDKs humanas. Os scores à direita indicam a distância evolucionária entre os membros. Análise feita pelo programa ClustalOmega.

4.2. Modelagem molecular comparativa

Inicialmente, a modelagem molecular comparativa foi realizada utilizando o SwissModel (Expasy), uma ferramenta online, que permitiu gerar modelos das CDKs cujas estruturas cristalográficas não estão reportadas no PDB.

Vendo a necessidade de gerar múltiplos modelos, inclusive na presença de ligantes, passamos a utilizar o Modeller v. 9.23. Portanto, foi possível a obtenção dos modelos dos complexos das CDKs com o ATP, ribociclibe, palbociclibe e abemaciclibe. Importante mencionar que foi possível a obtenção dos modelos apenas do domínio quinase das CDKs, desde que todas as estruturas depositadas no PDB correspondem ao próprio domínio quinase. Nenhuma informação estrutural está disponível para as regiões N- e C-terminal das CDKs de maior massa molecular. Possivelmente devido a estas regiões serem naturalmente desordenadas, enquanto na sua forma inativa.

Outra consideração importante é que cada modelo, das 9 CDKs que não possuem estruturas reportadas (CDK3, CDK10, CDK11B, CDK14, CDK15, da CDK17 à CDK20), foi gerado a partir de 5 moldes diferentes. Desses moldes, 3 deles foram de CDK2 complexadas com algum ligante (Código PDB: 1aq1, 1gz8, 5k4j), os quais foram utilizados para todas as CDKs modeladas. Houve a preocupação de utilizar um molde que estivesse na forma “ligada”, desde que é sabido que há um movimento da volta T, onde situa-se o resíduo fosforilável T160 na CDK2. Os outros 2 moldes utilizados foram específicos para cada CDK, a fim de garantir a especificidade estrutural de cada uma, baseada na porcentagem de identidade obtida pelos alinhamentos de sequência.

Almejando obter o modelo mais representativo, foram gerados 100 modelos para cada uma das CDKs, dos quais, o eleito foi aquele de menor energia livre de Gibbs e de melhor qualidade estereoquímica, baseado na análise do gráfico de Ramachandran (modelos com maior porcentagem de resíduos nas regiões permitidas e adicionalmente permitidas, e zero ou menos de 1% dos resíduos em regiões não permitidas), obtido pelo SAVES v5.0 (<https://servicesn.mbi.ucla.edu/SAVES/>).

4.3. *Investigação do N- e C-terminal das CDKs*

Conforme mencionado no item anterior, nenhuma informação foi encontrada que pudesse representar a estrutura secundária ou terciária das regiões N- e C-terminal das CDKs de maior massa molecular. Sendo assim, na ausência de informações sobre a estrutura e função, optou-se por realizar uma análise da sequência gênica, baseada então no cDNA dessas CDKs. Para isso, foi utilizado o BLASTn do NCBI, e a partir da opção EST, buscou-se por correlações com outras sequências. Os resultados mostraram que há a expressão em diferentes tecidos, porém ainda assim, apontam para as mesmas CDKs. Ou seja, a análise converge para as CDKs expressas em diferentes tecidos, demonstrando que não há outras correlações. Esta análise foi importante para confirmarmos a ausência de informações disponíveis sobre estas sequências, e que provavelmente são exclusivas das CDKs, além de reforçar a hipótese de que são regiões desordenadas, que assumem um arranjo estrutural momentâneo, durante a interação com outras proteínas da célula, para desempenharem sua função.

4.4. *Interações proteína-ligante*

Esta etapa do projeto culminou na obtenção dos complexos CDK-ligante ATP-competitivo, e desde que as estruturas cristalográficas da CDK6 na presença de ribociclibe, palbociclibe e abemaciclibe (PDB 5L2T, 5L2I, 5L2S, respectivamente), estão disponíveis no PDB, surgiu o interesse de investigar o diferencial que torna estes ligantes específicos para a CDK6 e também para a CDK4, mas não para as demais CDKs. Para realização das análises, de posse dos modelos gerados, foram consideradas tanto as interações intermoleculares, quanto as intramoleculares no entorno do sítio ativo. Estas análises oferecem mais detalhes estruturais sobre o modo de interação com os ligantes, considerando que as interações por ligação de hidrogênio se fazem majoritariamente com a cadeia principal, e as interações hidrofóbicas, com as cadeias laterais apolares presentes no bolsão mais profundo do sítio de ligação.

Nesse relatório serão exploradas as análises realizadas com as CDKs ligadas ao ribociclibe, focando nos pontos divergentes observados dentre as CDKs, e propondo justificativas sobre a especificidade já citada. Para justificar o foco em somente um dos ligantes, o próximo tópico aborda e mostra que o modo de interação

entre os ligantes abemaciclibe, palbociclibe e ribociclibe com a CDK6 são muito semelhantes, até mesmo devido a semelhança entre os possíveis inibidores de CDK4/6.

4.4.1 Modo de interação entre os ligantes: abemaciclibe, palbociclibe e ribociclibe e a CDK6

Conhecidos como a terceira geração de inibidores para as CDKs, os fármacos abemaciclibe, palbociclibe e ribociclibe foram reportados, inicialmente, como inibidores específicos para a CDK4 e a CDK6. Essas moléculas ganharam atenção da comunidade por apresentarem um impacto significativo no combate ao câncer de mama além de receberem aprovação da agência reguladora estadunidense FDA. Esta seção é destinada à análise das estruturas do complexo da CDK6 com tais inibidores, onde dar-se-á maior atenção ao modo de interação com o sítio catalítico, e quais são as principais diferenças entre tais interações, que suportam a ideia da especificidade desses fármacos com ambas CDKs. Importante lembrar que não há estrutura tridimensional reportada para a CDK4 na presença de tais inibidores.

Figura 7. Modo de interação dos fármacos de terceira geração. Estruturas de CDK6 com: abemaciclibe (A), palbociclibe (B) e ribociclibe (C), em uma mesma posição (PDB 5L2S, 5L2I, 5L2T, respectivamente). As ligações de hidrogênio são representadas como linhas pontilhadas e os principais resíduos do sítio ativo foram rotulados.

A figura 7 ilustra os fármacos abemaciclibe (5L2S), palbociclibe (5L2I) e ribociclibe (5L2T) e os resíduos de aminoácidos da CDK6 que realizam interações. As estruturas foram determinadas na ausência da ciclina D, e portanto, se configuram na conformação inativa, com a hélice C deslocada do sítio ativo, e a alça de ativação desordenada. Contudo, os contatos realizados entre os inibidores de terceira geração e a CDK6 não são afetados por não estarem localizados diretamente nessas regiões

que sofrem mudança conformacional com a alça de ativação das CDKs (Chen et al., 2016).

Cada fármaco interage com a região de alça usando o mesmo grupo 2-aminopirimidina. Além disso, existem várias interações exclusivas para CDK6. As estruturas cristalográficas revelam que o anel de piperazina carregado positivamente, de cada fármaco é estabilizado apoiando-se contra uma estrutura exposta ao solvente consistindo de D104 e T107. Em CDK1, CDK2, CDK3 e CDK5, o resíduo análogo a CDK6, D107 é lisina (K), que deve causar repulsão eletrostática com a piperazina e, portanto, menor afinidade com essas CDKs. No complexo de CDK6/abemaciclib, uma molécula de água ordenada é observada ligando o imidazol do resíduo de alça H100 e o ligante nitrogênio da piridina. A densidade de elétrons para esta molécula de água não é observada nos complexos palbociclibe e ribociclibe, no entanto, há espaço para uma molécula de água nesta posição. Essa interação de ponte com H100 poderia contribuir para a seletividade, uma vez que esse resíduo de histidina é encontrado em apenas 8 quinases, observados com base em alinhamento de sequência (Chen et al., 2016). Dentro da família CDK, H100 é encontrado apenas em CDK4/6. Em outro estudo, foi proposto que uma interação inibidor-H100 pode contribuir para a seletividade observada de 26 vezes para CDK4/6 sobre outras CDKs (Cho et al., 2010).

4.4.2 Interação entre ribociclibe e a CDK6

A Figura 8A mostra as interações intermoleculares realizadas pelos resíduos do sítio catalítico da CDK6 e o ribociclibe e a figura 8B as interações intramoleculares. Dois tipos de interações foram encontrados: hidrofóbicas, realizadas, em sua maioria, pela cadeia lateral, e hidrofílicas, ligações de hidrogênio realizadas pela cadeia principal dos resíduos de aminoácidos. Podemos listar os resíduos que efetuam interação hidrofóbica sendo eles: I19, V27, A41, V77, F98, C β D104, L152, A162 e C β D163, e considerando que cada resíduo, salvo L152, realiza uma interação, temos um total de 10 interações hidrofóbicas feitas com o ribociclibe. As ligações de hidrogênio (interações hidrofílicas) estão presentes em menor número, e sempre estabelecidas com a cadeia principal, sendo os resíduos participantes, V101 que realiza 2 interações, uma de doação e outra de recepção de hidrogênio, e o D163.

Figura 8. Análise do sítio catalítico da CDK6 em complexo com ribociclíbe. Em A estão representadas as ligações intermoleculares, e em B as ligações intramoleculares, realizadas pelos resíduos do entorno do sítio catalítico. Imagem gerada pelo PyMol.

A respeito dos aminoácidos citados, é importante ressaltar os que fazem parte de elementos que se repetem nessa família, como exemplo a A41 que pertence a extensão **VALK**, além dos aminoácidos A162 e D163, que constituem a região **KLADFGLAR**. Traçando um comparativo geral das interações intermoleculares encontradas na CDK6 com as outras CDKs, não foi possível constatar pontos de divergência, nos casos que os aminoácidos não foram idênticos, as substituições foram feitas de modo a conservar as características do aminoácido, como as trocas que ocorrem na CDK7: I10 por L6 e V77 por I63, conservadas. Outra troca recorrente é a substituição da V101 por C ou M, entretanto isso não influencia na interação intermolecular realizada, uma vez que as ligações de hidrogênio partem da cadeia principal, comum entre todos os resíduos de aminoácidos. Todavia, as interações intramoleculares, ou seja, aquelas realizadas entre os resíduos, são mais instáveis, e dependem do posicionamento de alças e voltas que por natureza são regiões mais expostas ao solvente e possuem uma maior mobilidade. São cerca de 15 interações intramoleculares que a CDK6 realiza quando em complexo com o ribociclíbe, dentre elas 5 são de caráter hidrofílico e o restante hidrofóbico. Quando comparada as interações reportadas pela estrutura 5L2T com as demais CDKs modeladas, somente a CDK4 e CDK6 possuem as mesmas 15 interações demonstradas na figura 8. Como principal ponto de divergência, podemos citar a alteração já citada da V101, que além de realizar as interações de hidrogênio, possui 2 interações intramoleculares com o resíduo L152. A troca da V101 por C ou M, impede a formação das interações citadas com o L152 devido ao fato da valina possuir uma cadeia lateral ramificada, apolar e

curta, diferente dos resíduos encontrados nessa mesma posição nas outras CDKs que não a CDK6 e CDK4. Além disso, das 5 ligações de hidrogênio, somente uma foi observada nas demais CDKs, com exceção da CDK4. Importante mencionar que o resíduo de T107 não é conservado em nenhuma outra CDK, que não sejam aquelas que possuem especificidade pelo fármaco estudado. Sabe-se que dentre as interações fracas de Van der Waals, as ligações de hidrogênio são aquelas que contribuem mais para a estabilização, por possuírem maior proximidade e direcionamento ocasionado pela própria doação/recepção de átomos de hidrogênio. Contudo, elas não proveem especificidade suficiente para manter as interações com o ribociclib. Nota-se isso quando observado que a rede de interações formada entre os resíduos do sítio catalítico é tão importante quanto, senão mais, por serem mais numerosas, do que as realizadas entre proteína-ligante. Sendo assim, é fundamental que as interações intramoleculares sejam realizadas para abrigar de forma energeticamente favorável os ligantes, fato que não é observado para as CDKs que não interagem com esse fármaco.

4.4.3 Análise da superfície eletrostática

Outra importante análise que fundamenta os modos de interação do ligante à proteína, é aquela feita sobre a superfície eletrostática das CDKs. Na figura 9 estão destacadas duas regiões laterais, uma a esquerda e outra à direita do sítio, que possuem uma maior densidade de carga.

A figura 9A mostra a superfície eletrostática do sítio catalítico da CDK6 representada na ausência do ligante. Na figura 9B está representado as cadeias laterais dos resíduos do sítio catalítico da CDK6 (PDB 5L2T), e as interações com o ligante ribociclibe. Podemos observar que as regiões destacadas, que possuem uma maior densidade de carga, são as regiões que abrigam resíduos e/ou grupos polares, passíveis de realizar ligação de hidrogênio, mais especificamente os resíduos V101 e D163. Importante lembrar que estas ligações se fazem pela cadeia principal e não lateral. Esta região também concentra resíduos como a H100 e o D102. No caso da região em destaque do lado direito, o D163 realiza ligação de hidrogênio com o ribociclibe, porém na sua maior extensão, esta região da cavidade não possui carga, sendo que 21 interações de um total de 27 (inter e intramoleculares), possuem caráter hidrofóbico.

4.5 Interface CDK-ciclinas

Diferente da abordagem com as CDKs, não foram realizadas modelagens para aquelas ciclinas que não possuem estruturas resolvidas. O objetivo aqui foi buscar entender um pouco mais das características das CDKs no complexo com suas ciclinas, considerando aqueles depositados no PDB, traçando características que possam distinguir as CDKs uma das outras no que tange a forma de ativação, regulação e recrutamento de substratos. Quando não foi possível distinguir cada CDK de forma individual, ao menos, grupos distintos foram estabelecidos de acordo com as semelhanças e diferenças das características brevemente citadas.

4.5.1 *pCDK2 em complexo com as ciclinas: A, B, E.*

Uma busca inicial no banco de dados PDB constatou uma série de estruturas da CDK2 em complexo com a parceira ciclina, inclusive na forma fosforilada (pCDK2), das quais três estruturas, CDK2-ciclina A (2cch), CDK2-ciclina B (2jgz) e CDK2-ciclina E (1w98), foram utilizadas para as análises abordadas neste tópico. A princípio, comparar a mesma CDK com diferentes ciclinas pareceu intuitivo para as análises comparativas, justificando a escolha dessas estruturas.

A figura 10 apresenta a sobreposição dos 3 complexos citados acima. Essa expõe de uma forma mais generalista as principais regiões das CDKs e das ciclinas, cujos nomes estão destacadas em preto. Os complexos ilustrados encontram-se na forma ativa, caracterizada pelo deslocamento da hélice-C (PSTAIRE) e posicionamento correto do segmento de ativação. Em relação às diferentes CDK2 ilustradas, não foi encontrada nenhuma região de grande divergência, salvo a volta do resíduo 40, que possui diferente orientação no complexo CDK2-ciclina A.

Figura 10. Sobreposição dos complexos de pCDK2 com as ciclinas A, B e E. Para comparação estão sobrepostos os complexos: pCDK2-ciclina A (PDB: 1QMZ - amarelo e magenta); pCDK2-ciclina B (PDB: 2JGZ - musgo e marrom); pCDK2-ciclina E (PDB: 1W98 - salmão e ciano). As estruturas secundárias mais relevantes estão designadas pelo nome. Estão também presentes um heptapeptídeo de sequência (HHASPRK em verde) e o ATP.

Apesar da similaridade entre os complexos, o alinhamento trazido pela figura 10 demonstra diferenças significativas entre as sequências de aminoácidos das ciclinas. Um dos reflexos da divergência é a maior desordem da região do C-terminal observado na ciclina B quando comparado com a ciclina A, o maior número de interações adicionais que o N-terminal da ciclina A e E realizam em comparação com a ciclina B. Essa e outras características que serão aprofundadas a seguir nos revelam para além da natureza regulatória das ciclinas, uma habilidade de coordenar o recrutamento de substratos específicos para cada CDK.

4.5.2 Aspectos estruturais: Diferenças e semelhanças entre os complexos de pCDK2.

Por se tratar de uma extensa interface de interação, é possível dividir as regiões de contato entre a CDK2 e as ciclinas, de acordo com os elementos estruturais que estão interagindo. Dessa forma, temos: região N-terminal das ciclinas, região da alça de ativação das CDKs e região da hélice C. No geral, para os 3 complexos da CDK2, a interface de interação com as ciclinas é muito similar e as diferenças observadas é resultado das mudanças de sequência de aminoácidos, como a figura 11, busca ilustrar.

Figura 11. Alinhamento entre as ciclinas mostrando resíduos em contato nas interfaces pCDK2-ciclina. Ciclina E1, ciclina A2 e ciclina B1 alinhadas. Os resíduos coloridos em ciano da ciclina E1 estão em contato com pCDK2 no complexo pCDK2-ciclina E1, os resíduos coloridos em magenta da ciclina A2 está em contato com pCDK2 no complexo pCDK2-ciclina A2 e os resíduos coloridos em marrom da ciclina B, estão em contato com pCDK2, no complexo pCDK2-ciclina B1. Contato é definido como qualquer átomo em uma distância até 3.5 Å na superfície pCDK2-ciclina.

CYCLIN_E1_(1W98)	IIAPSRGSP L PVL	SWANRE	EVWKIMLN	KEKTYLRDQHFL	EQHPLLQPKMRAILLD	WLM	EV	60	
CYCLIN_A2_(2CCH)	-----NE--VP--	DYHEDIHTYL	REMEVKCKPKVGYMKKQ	PDITNSMRAILVD	WL	VEV		49	
CYCLIN_B1_(2JGZ)	-----CS--EY	KDIYAYLRQLEEEQAVRPKYLL	-GREVTGNMRAILID	WL	VQV			46	
		:::	: : *		:		.*****:***::~*		
CYCLIN_E1_(1W98)	CEVYKL	HRET	FYLAQDF	FRYMATQENVV	KTLQLIGISSL	FIAAKLEEIY	PPKLHQFAY	120	
CYCLIN_A2_(2CCH)	GEEYKL	QNETLHLAVNYIDRFLSSM	-SVLRGKLQLVGTAAML	LASKFEEIY	PPPEVAEFVY			108	
CYCLIN_B1_(2JGZ)	QMKFRL	LQETMYMTVSIIDRFMQNN	-CVPKKMLQLVGVTAMFI	ASKYEEMY	PPPEIGDFAF			105	
		::*	.**::: .	::*: .	* : ***:*	:::.*:	**::*: .*: .		
CYCLIN_E1_(1W98)	VTDGACSGDEI	LTMLMIMK	ALKWRLSPL	TIVSWL	NVYMQVAYLNDLHEVLL	PQYPQQIF		180	
CYCLIN_A2_(2CCH)	ITDDTYT	KKQVLRMEHLVLRKVL	TFDLAAPT	VNQFL	TQYFL	HQQPANCKVESL	AMF----	163	
CYCLIN_B1_(2JGZ)	VTDNTYTKHQIR	QMEMKILRALNFG	LGRPLPLHFL	RRASKI	-GEVDVEQHTL	AKY----		159	
		::*:	: .:	**	:::.*:	*	:*	: . *	
CYCLIN_E1_(1W98)	IQIAELLDLC	VLVDVD	-CLEFPYG	ILAAASAL	----	YHFSS	---SELMQKVSGYQW	CDIENC	232
CYCLIN_A2_(2CCH)	-----LG	ELSLIDADPYLKYLP	SVIAGAAFHLALYTV	TGQSWPESLIRKTGYT	LES	LKPC		218	
CYCLIN_B1_(2JGZ)	-----LM	ELTMLD	YD-MVHFPPSQIAAGAF	CLALKILDNGEW	TPTLQHYLSY	TEESLLPV		213	
		*	:* :*	* : .	:*..*	.	: : .*	: :	
CYCLIN_E1_(1W98)	VKWMVPFAM	VIRETGSS	KLKHFRGV	-ADEDAH	NIQTHRDS	LDLLDKARAKA	---	283	
CYCLIN_A2_(2CCH)	LMDLHQTYL	KA-PQHA	--QQSIREKYKNSKYHGV	-----	SLLNP	PET---	LNL	260	
CYCLIN_B1_(2JGZ)	MQHLAKNV	VMV-NQGL	TKHMTVKNKYATSKHAKI	-----	STLPQL	NLSALVQDL		260	
	:	:	:	:	:	:	:		

Como observado, o complexo pCDK2-ciclina E possui um maior número de contatos dentro da interface comparado ao complexo com ciclina A e ciclina B, principalmente na região do C-terminal, onde demonstra ser a única a realizar interações. Um dos reflexos desse maior número de interações é o tamanho da superfície molecular enterrada entre a CDK e a ciclina. Essa assume maiores valores para complexos que apresentam maior número de interações e interações mais energéticas. Com a maior área temos a pCDK2-ciclina E (3250 Å), seguida da pCDK2-

ciclina A (2911 Å) e pCDK2-ciclina B (2758 Å), valores esses condizentes com a quantidade de interações realizadas por cada dímero e com os encontrados na literatura (Honda et al., 2005).

É de conhecimento que a CDK1 é encontrada, normalmente, em complexos contendo ciclina A e ciclina B, enquanto CDK2 é encontrado em complexos que contêm ciclina A e ciclina E. Essa seletividade pode ser explicada com base nas alterações das sequências de aminoácidos observadas na CDK2. Cerca de 20 resíduos da ciclina B compõem a interface com a CDK. Destes, 11 são idênticos aos aminoácidos equivalentes na ciclina A, enquanto apenas 7 são idênticos aos equivalente na ciclina E. Em particular, os resíduos Y170, Y177 e Y258 da ciclina B são conservada como aromático na ciclina A e apresenta grandes cadeias laterais que se acumulam na superfície da CDK1, enquanto a ciclina E apresenta aminoácidos menores nessas posições (N112, I119 e L202). Essas alterações indicam um possível motivo pelo qual a CDK1 liga-se preferencialmente às ciclinas B e A, ao em vez da ciclina E. Ainda mais, a CDK2 pode suprir esses efeitos gerados pela baixa eficiência de interação dos resíduos pequenos da ciclina E com a sua extensa interface, realizando um maior número de interações, inclusive com o C-terminal, como indicado na figura 11.

4.5.3 Interações com o N-terminal das ciclinas.

O N-terminal das ciclinas A e B estão em contato com as regiões da hélice C, início de $\beta 7$ e com o segmento de ativação. A ciclina E por outro lado, interage apenas com a hélice C através de seu N-terminal, realizando uma interação adicional com o resíduo E57 (pCDK2). Além disso, os resíduos do N-terminal na ciclina A são mais bem ordenados que na ciclina B e fazem contatos adicionais com a CDK2.

Como as figuras 12 e 13 ilustram, a hélice C se posiciona quase que paralelamente à hélice $\alpha 5$ das ciclinas A, B, E. Nessa região existem extensas interações hidrofóbicas, especialmente com pCDK2 I49 e I52, contudo foram omitidas por serem comuns entre os complexos. A interface é reforçada por ligações de hidrogênio que são semelhantes entre os complexos pCDK2-ciclina A e pCDK2-ciclina B, salvo algumas alterações na sequência.

Figura 12. Contatos da hélice N-terminal da ciclina para três regiões distintas do pCDK2. Em (A) complexo com a ciclina A (magenta), (B) complexo com a ciclina B (marrom) e em (C) complexo com a ciclina E (ciano). A pCDK2 aparece em amarelo, musgo e salmão, respectivamente.

Os contatos da ciclina B (envolvendo Y170, D173 e Y177) são semelhantes aos realizados pelos resíduos correspondentes da ciclina A (Y178, D181 e Y185), todavia, na ciclina E, apenas a tirosina 112 conserva a interação observada nos outros complexos. A ciclina A possui 3 ligações adicionais de hidrogênio através de seus resíduos do N-terminal, E174, V175 e D177, como pode ser observado na figura 12. Ainda, somente no complexo com a ciclina A é observada interação polar entre S120 (CDK2) e D181 (ciclina A). Os resíduos correspondentes não são ordenados nos outros dois complexos. Dentre os complexos de CDK2 analisados, o com a ciclina E, até o momento, mostrou diferentes formas de interação, predominantemente mais apolares e menos conservado, em relação aos outros 2 complexos.

4.5.4 Interações com a alça de ativação

A alça de ativação é o segundo maior elemento estrutural da CDK2 em contato com as ciclinas, e interage com os resíduos presentes nas hélices $\alpha 3$ e $\alpha 4$ das diferentes ciclinas em análise. Nessa região há a formação do cluster de argininas, onde 3 argininas interagem e estabilizam a carga do pT160. No complexo pCDK2-ciclina A, o grupo fosfato da pT160 interage com as argininas (R50, R126 e R150) e atua como um centro organizador para alocar a hélice C e o segmento de ativação nas orientações corretas para reconhecimento de substrato. Para os complexos com ciclina A e ciclina E o mesmo centro organizador, sendo conservado a posição das 3 argininas e suas interações, não só com a pT160, mas como resíduos ao redor, como é observado na figura 13. Essas interações são reforçadas por contribuições do grupo carbonil da cadeia principal do resíduo Y258 (ciclina B), que realiza duas ligações de hidrogênio com R50 da pCDK2. De uma forma muito semelhante, a ciclina A com a cadeia principal do resíduo F267 e a ciclina E com o resíduo L187, também realizam interação com a arginina 50.

Figura 13. Contatos realizados pelas hélices $\alpha 2$ e $\alpha 3$ das ciclinas com o segmento de ativação da pCDK2. A figura segue o mesmo esquema de cores e posições presentes na figura 12. A figura também destaca os resíduos que realizam interação e os casos de perda de contato devido a substituição. Somente os contatos polares foram medidos e possuem valores quantitativos.

Ainda na figura 13, outra alteração significativa encontra-se nas diferentes interações realizadas pelos resíduos da hélice $\alpha 2$ das ciclinas. Para o complexo CDK2-ciclina B, temos que o resíduo E221 do elemento estrutural $\alpha 2$, realiza ligação de hidrogênio com o R157 da pCDK2. O resíduo equivalente no complexo com a ciclina A, o E230 está rotacionado e forma uma interação intramolecular com os resíduos N312 e Q228 (equivalente ao L219 na ciclina B e H147 na ciclina E). Na ciclina B, os resíduos apolares L301 e L303 substitui resíduos polares da ciclina A, e L219 (o equivalente o resíduo Q228 da ciclina A) é voltado de modo a formar contato com eles. As leucinas apolares da ciclina B favorecem E221 ser externo, tornando-o disponível para uma interação iônica com o pCDK2 R157 no complexo pCDK2-ciclina B enquanto na ciclina A o resíduo equivalente (E230) é enterrado e o Q228 polar (equivalente à ciclina B L219) está disponível para entrar em contato com R157. Na ciclina E, o ácido glutâmico presente nas outras duas ciclinas é conservado, na posição E149. Todavia, ele não possui o ambiente, assim como na ciclina B, que o torne externo para interação com R157.

4.5.5 Interações com a volta do resíduo 40 e volta do resíduo 70

A figura 14 ilustra os resíduos das hélices $\alpha 3$, $\alpha 4$ e $\alpha 5$ das ciclinas em contato com as voltas 40, 70 e hélice C da pCDK2. A volta 40 mostra certa variabilidade conformacional dentro dos complexos em análise. No complexo com a ciclina B, a volta 40 é organizada no entorno de 5 ligações de hidrogênio. Essas interações são semelhantes às feitas em pCDK2-ciclina A com a diferença de que a ciclina B R283 é equivalente à ciclina A Leu292 e o contato com a cadeia principal T41 é obtido pela ciclina A K28, que é do início da hélice $\alpha 5$. Alterações nessa região também foram observados no complexo pCDK2-ciclina E, como resultado de alterações na sequência entre a ciclina A e a ciclina E. Pode ser observado a alteração do resíduo R283 (ciclina B) para L212, e subsequente perda de ligação de hidrogênio realizado com T41 (pCDK2). Os resíduos L288 e L289 da ciclina A são alterados para G208 e D209 na ciclina E e isso resulta em uma alteração na conformação da alça.

Também há mudança conformacional na região da volta 70 (pCDK2 71-74). A alteração de H296 (ciclina A) para L216 na ciclina E causa uma mudança na alça para evitar o contato da leucina com o pCDK2 H71 e T72. Já na ciclina B o contato não é reproduzido devido a mudança do resíduo H296 para M287. Todavia, ainda é observado rigidez nas ciclinas dos diferentes complexos devido à formação de um par

de íons internos, na ciclina E, entre K186 e E215. Essa interação também é observada para os complexos com ciclina A (E295-K266) e com a ciclina B (E286-K257), tratando-se de uma marca característica no enovelamento de ciclinas (Brown et al., 2015; Honda et al., 2005).

Figura 14. Contatos das hélices $\alpha 3$, $\alpha 4$ e $\alpha 5$ das ciclinas para a região de voltas do resíduo 40 e do resíduo 70 da pCDK2. Os contatos são semelhantes entre os complexos, com diferenças pontuais, como a diferença de interação da T41 (pCDK2), hora em contato, através de seu oxigênio da cadeia principal, (A) com K288 (ciclina A), hora em contato com a (B) arginina (R283) da ciclina B. Já na ciclina E (C) o mesmo contato não é observado devida a rotação da cadeia lateral de T41 e também pela substituição de K288 (ciclina A) pelo resíduo hidrofóbico L212.

Nos complexos de pCDK2, a hélice C é reorientado pela ciclina principalmente através de contatos hidrofóbicos com os resíduos das hélices $\alpha 5$ e $\alpha 1$. A maior diferença ocorre no resíduo G296 da ciclina B, que é equivalente a asparagina (D305) da ciclina A e R225 na ciclina E. Somente no complexo pCDK2-ciclina A D305 forma um par de íons com a cadeia lateral da K56 (pCDK2), isso devido à falta de conservação da asparagina e mudança de conformação das lisinas dos complexos pCDK2-ciclina B e pCDK2-ciclina E. Contudo, apesar das diferenças de sequência

entre as ciclinas na interface com pCDK2, as alterações na sequência podem ser acomodadas por conformações alternativas das regiões de volta 40 e 70 sem afetar os resíduos da interface catalítica.

4.5.6 Sítio de recrutamento RXL Φ

O complexo pCDK2-ciclina A utiliza uma região denominada como sítio de recrutamento localizada na superfície da ciclina para ligar o motivo estrutural RXL Φ ou RXLX Φ (sendo X um resíduo de aminoácido qualquer e Φ um resíduo hidrofóbico volumoso) em substratos proteicos. Isso resulta além da localização do substrato, num aumento da eficiência enzimática (Cheng et al., 2006). O local correspondente na ciclina B possui alta similaridade, todavia, duas modificações são significativas. Temos que o resíduo E220 da ciclina A, resíduo que reconhece a arginina (R) do motivo RXL Φ , está modificado para uma glutamina (Q211) na ciclina B. A mudança de resíduo para um não carregado provavelmente enfraquece a interação, todavia, não a destrói. A ciclina E mantém conservado o ácido glutâmico na posição de reconhecimento da arginina. A maior contribuição deste sítio de recrutamento está na porção hidrofóbica que aloca a leucina ou o resíduo hidrofóbico adjacente a ela, do motivo RXL Φ . Comparação estrutural e da sequência das 3 ciclinas em estudo mostram que os resíduos que contribuem para a formação desse bolsão hidrofóbico estão conservados. Todavia, é provável que para o complexo com ciclina B, o recrutamento possua certa afinidade pelo motivo RXL Φ , porém, mais fraca do que nas outras duas ciclinas. A outra maior modificação no sítio de recrutamento está na alteração do resíduo E224 (ciclina A) para o K215 (ciclina B) e V143 (ciclina E). A mudança que ocorre na ciclina B tende a ser mais prejudicial para a interação devido a possibilidade de criar um impedimento espacial. A representação da superfície eletrostática dos diferentes complexos de pCDK2 ilustradas nas figuras 15A, 15B e 15C à direita ajuda no entendimento e na observação das mudanças encontradas na região de recrutamento RXL Φ . No complexo pCDK2-ciclina A, o RXL Φ contendo o peptídeo é ligado em uma região majoritariamente carregada com cargas negativas, enquanto que na ciclina B, temos que essa região é neutra, com uma carga positiva devido a presença da lisina (K215). Para o complexo com ciclina E, essa região é ainda mais neutra, sendo a valina (V143) o resíduo que ocupa a posição da K215 (pCDK2-ciclina B).

Figura 15. À esquerda, em A, B e C estão os sítios de ligação RLX para as ciclinas A, B e E respectivamente. O esquema de cores segue o mesmo padrão das imagens 7, 8 e 9. O peptídeo é encontrado originalmente na estrutura de código PDB (2CCI), um complexo pCDK2 ligado a uma ciclina A. Para uma melhor comparação, o peptídeo foi carregado nas estruturas em análise: 1QMZ (pCDK2-ciclina A), 2JGZ (pCDK2-ciclina B) e 1W98 (pCDK2-ciclina E). À direita, em A, B e C estão representadas a superfície de carga dos complexos pCDK2-ciclina A, pCDK2-ciclina B e pCDK2-ciclina E, respectivamente. As cargas negativas receberam a coloração azul, as positivas, vermelho e as neutras cinza. CA – ciclina A, CB – ciclina B e CE – ciclina E.

Outra diferença em relação a carga da superfície ocorre na hélice $\alpha 1'$, uma região carregada positivamente na ciclina B. Essa região possui a sequência R₃₀₇RASK, em destaque na figura 15. Trata-se de uma região conhecida entre os

homólogos da ciclina B, por manter-se conservada em sua totalidade. Todavia, como demonstrado no alinhamento, essa região não é conservada para a ciclina A e para ciclina E. Tal motivo estrutural mostrou-se importante para o reconhecimento do substrato cdc25C pelo complexo CDK1-ciclina B (Goda et al., 2001). A cadeia lateral da R308 aponta na direção oposta à da R307 e isso impossibilita que as duas argininas adjacentes criem um local de reconhecimento de fosfato. Os resíduos A309 e S310 são internos com hidrogênio da cadeia lateral S310 ligado à cadeia lateral H320. O K311 é externo e pode estar disponível para interação, todavia, sua contribuição parece ser menos significativa do que a das duas argininas. Esta região difere muito nas ciclinas A e E, tanto em sequência como em estrutura. Na ciclina A a sequência R₃₀₇RASK é T₃₁₆QYFL e muitas das cadeias laterais tem posições diferentes. Na ciclina E, os resíduos correspondentes: N₂₃₆VYMQ também são diferentes. Este local faz parte de uma estrutura presente ciclina E que está implicado na ligação a centrossomos, de forma independente da CDK. Até o momento, as atividades da CDK1 e CDK2 em relação aos substratos recrutados demonstrou depender da identidade da ciclina associada, das características dos resíduos ao redor do sítio de transferência de fosfato (canônico S/TPXR/K ou não canônico S/TPXX) e da distância entre o RXL e motivos S/TP (Takeda et al., 2001). Estudos ainda indicam que o motivo RXL demonstrou capacidade de melhorar a afinidade de um substrato da CDK, efeito que é ainda mais expressivo se o substrato não tiver um resíduo básico na posição P+3 (Brown et al., 2007; Cheng et al., 2006).

4.6 Análise dos demais complexos CDK-ciclina: um novo 'ramo' na família das CDKs.

Seguindo os estudos da interface CDK-ciclina, foram analisados novos complexos. A seleção desses foi dada com base em informações preliminares da literatura, que indicam divergências significativas entre os membros da família das CDKs, principalmente nas interações com as ciclinas e ativação do complexo. Além disso, é importante ressaltar que, assim como as análises anteriores, estas também foram realizadas somente com os complexos CDK-ciclinas cujas estruturas já foram resolvidas e depositadas no banco de dados PDB. Dessa forma, os complexos: CDK1-ciclina B (4Y72), pCDK2-ciclina A (1FIN), pCDK4-ciclina D3 (3G33), CDK6-ciclina viral (1G3N) e CDK9-ciclina T1 (3BLH), foram empregados nas análises que estão a seguir.

A figura 16 traz 3 diferentes complexos de CDK com diferentes ciclinas. A coloração das estruturas busca distinguir diferentes elementos de estrutura secundária ao mesmo tempo que compara os mesmos entre os complexos ilustrados. Os retângulos desenhados na figura 15 destacam a interface CDK-ciclina, onde podemos ver diferenças significativas, principalmente no complexo CDK4-ciclina D. Ao contrário do que é esperado, a conformação da hélice C e do segmento de ativação do dímero de CDK4 assemelham-se mais à forma monomérica inativa da CDK2, do que a forma parcialmente ativada (ligado a ciclina), ou totalmente ativa (ligação da ciclina e pT160). Embora a ciclina D tenha 'acoplada' à hélice C (PISTVRE) da CDK4, a hélice não passa pela mudança conformacional associada à ativação das CDKs, como observado na grande maioria dos complexos da família CDK resolvidos até então (Pan et al., 2001; Schulze-Gahmen & Kim, 2002). Observa-se que a interação entre a ciclina D e a CDK4 difere-se dos outros dois complexos, devido a ciclina D estar contida numa posição mais elevada, realizando contatos majoritariamente com a hélice C e com o N-terminal da CDK4. Como consequência temos uma interface menor da interação entre a CDK4 e sua ciclina parceira, como ilustrado na figura 16. Ainda que mais semelhante a pCDK2, o complexo CDK1-ciclina B também possui uma interface de interação com a sua ciclina um pouco menor, porém, ainda mais extensa do que a da CDK4-ciclina D. Contudo, é sugerido, que a CDK2, devida sua extensa interface de interação com as ciclinas A e B, possua certa vantagem em relação a CDK1, nas ligações às ciclinas. Ainda mais, o balanço desse suposto desequilíbrio entre a CDK2 e a CDK1, seria suprido pela fosforilação do segmento de ativação da CDK1 (Brown et al., 1999).

Figura 16. Comparação entre os complexos: (A) CDK1-ciclina B (4Y72), (B) pCDK2-ciclina A (1FIN) e (C) CDK4-ciclina D (3G33). As CDKs foram coloridas de forma a destacar os elementos: hélice C (roxo), alça de ativação (magenta), N-terminal (branco) e C-terminal (amarelo). As ciclinas estão coloridas em ciano. Os retângulos circundam a interface CDK-ciclina. Complexo de pCDK2 também mostra em bastonetes o ligante em azul e o heptapeptídeo em verde.

As estruturas CDK2-ciclina A e CDK4-ciclina D3 revelaram padrões de interações diferentes da CDK em relação à ciclina, como ilustrado nas figuras 16 e 17. Como já discutido, a ligação da ciclina A à CDK2 provoca mudanças conformacionais, e o movimento da hélice C e segmento de ativação organiza os principais resíduos do sítio catalítico (Song et al., 2001).

Figura 17. (A). complexo pCDK2-ciclina A (1QMZ). (B) complexo CDK4-ciclina D (3G33). (C) complexo CDK2-ciclina A (1FIN). (D) complexo CDK1-ciclina B (4Y72). O padrão de cores segue o mesmo esquema da figura 16. Um substrato peptídico a partir do PDB 1QMZ é incluído no painel A, o qual ilustra a organização dos resíduos para formação do cluster de arginina na pCDK2. Os painéis B, C e D ilustram as respectivas CDKs e a disposição dos mesmos resíduos participantes do cluster R. Em (D), o grupo β com os resíduos F153 e G154 estão representados pelo círculo.

Interações com o N-terminal da ciclina garantem ao segmento de ativação da CDK2 uma conformação que requer poucos rearranjos na fosforilação da T160 para o reconhecimento de substratos. Essas interações também facilitam a organização dos resíduos R50, R126 e R150, para que possam acomodar o resíduo de fosfotreonina, que é conservado na família CDK. Já para a CDK4, como é possível de observar na figura 17, a ligação da ciclina D não é suficiente para assumir a conformação ativa. A hélice C da CDK4 permanece deslocada do sítio ativo e o

segmento de ativação, embora estruturado, não faz contato com a subunidade ciclina, tornando-se incompatível com a ligação do substrato. Como resultado, há uma dispersão entre os resíduos R55, R139 e R163 e o T172 (figura 17B). Mesmo quando fosforilado e ligado à ciclina, o sítio ativo da CDK4-ciclina D permanece em uma conformação que parece incapaz de suportar catálise, podendo esse então, ser um mecanismo catalítico assistido por substrato. A estrutura da CDK1-ciclina B difere dos outros 2 complexos, e embora mais próximo da estrutura geral da CDK2-ciclina A do que CDK4-ciclina D, possui características que o distingue de seus homólogos mais próximos. O segmento de ativação da CDK1 não se estende para fazer contatos extensos com a ciclina como acontece nos complexos de CDK2. Ao em vez disso, a sequência forma um grampo β com resíduos F153 e G154 na ponta, estrutura destacada por um círculo na figura 17D. Também é observado a formação de uma fenda entre a parte posterior do segmento de ativação e o N-terminal da ciclina B. Essa fenda é preenchida por solvente e é comumente encontrado em estruturas da CDK4-ciclina D e CDK9-ciclina T1, porém, não é característica da CDK2-ciclina B (Baumli et al., 2008).

4.6.1 Complexos 'tipo' pCDK4-ciclina D: estruturas e funções.

Como visto no tópico anterior, a CDK4-ciclina D possui características que diferem significativamente dos complexos de pCDK2 ativado. Na verdade, a CDK4-ciclina D assemelha-se mais às estruturas de formas inativas e não ligadas a ciclina, como a CDK7, e a CDK6 inibida por INK (Baumli et al., 2008; Schulze-Gahmen & Kim, 2002). A volta T da CDK4 é estabilizada na conformação inativa, devido às interações com os lóbulos N- e C-terminais da quinase. Como observado na figura 18, arquiteturas semelhantes da volta T são observadas na CDK9, em complexo com a ciclina T1 e na CDK6 no complexo CDK6-p19INK4d (1G3N).

Figura 18. Comparação da estrutura da CDK4-ciclina D com outros complexos de CDK-ciclina. (A) CDK4-ciclina D. (B) CDK6-ciclina K viral - p18Ink4c. (C) CDK9-ciclina T1. A CDK está representada na cor laranja e a subunidade ciclina em verde. O inibidor de CDK6, p18Ink4c em B está colorido em cinza. Cada foto é tirada do mesmo ponto de vista com relação a CDK. Os códigos PDB são respectivamente: 3G33, 3BLH e 1G3N.

Em relação a disposição entre as subunidades do complexo CDK-ciclina, a estrutura mais análoga a CDK4-ciclina D é a estrutura do complexo CDK6-ciclina K (ciclina viral) ligado ao inibidor p18^{INK4c} e a CDK9-ciclina T1 (Day et al., 2009). A ligação do inibidor INK4c apesar de distante da interface CDK-ciclina, provoca efeito, ao induzir uma rotação do lóbulo N-terminal da CDK6 no sentido do C-terminal da mesma. Como consequência, há uma distorção do sítio catalítico, e ainda, reconfiguração da volta T para assumir uma conformação inibida, o que leva a hélice C a conformação inativa. A mudança que ocorre nesse complexo da CDK6 também altera a sua superfície de contato com a ciclina K de forma que ela somente interage com os lóbulos C- e N- terminais da CDK6, de forma muito parecida com o complexo

CDK4-ciclina D, como demonstrado na figura 18A. Essa conformação pode ser adotada em complexos CDK6-ciclina, mesmo quando a subunidade INK não está presente. A atividade inibitória da subunidade INK seria então resultado da capacidade de se ligar e estabilizar esta conformação, evitando assim a adoção transitória de uma conformação ativa que, possivelmente, deve ocorrer durante o ciclo catalítico das CDK4 e CDK6.

No complexo CDK9-ciclina T a ligação da ciclina leva a hélice C a uma conformação ativa. Todavia, a disposição da ciclina e a área de superfície entre as subunidades catalítica e regulatória não condizem com os outros heterodímeros de CDK-ciclina. A ciclina T1, de forma semelhante à da ciclina D, posiciona-se mais acima, realizando contatos com a CDK9 por uma interface mais reduzida, primordialmente formada pelo N-terminal da CDK e o motivo CBF presente no N-terminal da ciclina T1. Além disso, a ciclina T1 é rotacionada para longe da CDK, provocando uma redução de 60% da área de superfície enterrada em comparação com a CDK2 (Day et al., 2009; Pan et al., 2001; Wang et al., 2020). Portanto, em contraste com todos os outros complexos de CDK-ciclina para os quais as estruturas foram determinadas até o momento, a ligação da ciclina D à CDK4 não impõe uma conformação ativa da CDK4.

5. CONCLUSÕES

Sabendo-se das dificuldades inerentes à obtenção das estruturas tridimensionais das CDKs transcricionais, devido à extensão da cadeia e peculiaridades estruturais relacionadas às suas funções, o presente estudo pôde colaborar com análises estruturais comparativas entre aquelas estruturas de CDKs depositadas no PDB e CDKs obtidas neste trabalho, a partir da modelagem molecular comparativa, utilizando tais estruturas como moldes. Neste sentido, as ferramentas computacionais mostraram-se uma forte aliada na busca por informações que possam trazer respostas e direcionamentos para o desenvolvimento de inibidores da atividade catalítica, que sejam mais específicos para essa importante família de proteínas. Neste estudo podemos observar a necessidade de explorar a superfície de interação com as ciclinas, as proteínas parceiras, reguladoras da atividade das CDKs, muito mais do que o sítio de ligação do ATP (com 55% de identidade quando considerados os resíduos que interagem com o ATP), bastante conservado dentre todas as CDKs, apesar das especificidades encontradas para os fármacos de terceira geração, aqui observadas, inibidores das CDK4 e CDK6.

A interface de interação com as ciclinas mostrou possuir maiores diferenças, com impactos mais significativos, isso é, alteração do modo de ativação e recrutamento de substratos. Dessa forma, é de se esperar que o estudo detalhado dessa região possa levantar informações importantes para a especificidade dos membros da família, levando assim, a obtenção de características incomuns que possam ser utilizadas para o desenho de novos inibidores de CDKs.

Como considerações finais, podemos apontar:

- Que os membros da família podem ser agrupados de acordo com as características evolutivas: as CDKs 1, 2 e 3 por serem as menores CDKs; as CDKs 4 e 6 pelas características distintas do sítio catalítico; as CDKs 10 e 11B, pelas semelhanças na sequência primária, quando comparadas aos demais membros; as CDKs 12 e 13, por serem as maiores CDKs; e as CDKs 8 e 19, devido à presença de uma alça mais extensa próximo ao sítio ativo, presente apenas em ambas.
- Que não foram obtidas informações suficientes a respeito das porções N- e C-terminais das maiores CDKs, impossibilitando o detalhamento estrutural dessas regiões, as quais portanto, devem apresentar-se de forma desordenadas quando da ausência da interação com proteínas parceiras. Apenas os domínios quinases dessas

CDKs foram modelados, assim como as estruturas resolvidas empiricamente, onde também só foi possível determinar o domínio quinase.

- Que os resíduos D104 e T107 possuem papel fundamental na estabilização do anel piperazina dos inibidores palbociclibe, ribociclibe, abemaciclibe com a CDK6. Em outras CDKs, como a CDK1, CDK2, CDK3 e CDK5, o resíduo análogo ao D104 é uma lisina, o que muito provavelmente enfraquece a interação por conta de uma repulsão eletrostática, devido a carga positiva. Portanto, não interage com essas CDKs.
- Que os resíduos H100, D104 e T107 possuem importantes papéis na estabilização dos fármacos em estudo, sendo exclusivos e fundamentais para a especificidade desses com as CDK4 e CDK6.
- Quanto aos complexos CDK-ciclina, foi possível analisar as CDK1-ciclina B, CDK2-ciclina A, CDK2-ciclina B, CDK2-ciclina E, CDK4-ciclina D, CDK6-ciclina K e CDK9-ciclina T1. Dois dos principais motivos pelos quais a ciclina E interage somente com a CDK2, e não interage com a CDK1, uma homóloga muito próxima, são: baixa identidade entre resíduos da ciclina E e aqueles das ciclinas que compõe a superfície CDK1-ciclina B e CDK1-ciclina A, e menor superfície de interação do complexo CDK1-ciclina E, quando comparada com o complexo CDK2-ciclina E.
- Que o local de recrutamento RXL na ciclina B é consideravelmente mais fraco do que o local correspondente na ciclina A, por causa das mudanças de sequência (Glu220, Glu224 na ciclina A para Gln211, Lys215 na ciclina B, respectivamente).
- Que o complexo pCDK2-ciclina A usa o local de recrutamento RXL em substratos que contêm sequências não canônicas para aumentar a eficiência enzimática, ao passo que o complexo pCDK2-ciclina B é capaz de fosforilar sequências não canônicas sem o benefício desse motivo estrutural.
- Que a CDK4-ciclina D se assemelha mais às estruturas de formas inativas e não ligadas à ciclina, e quando ligada à sua parceira, possui uma interface de interação reduzida e localizada mais acima, em relação a CDK2.
- Que os resíduos que participam da formação do cluster de arginina estão desordenados na CDK4, diferente do que é observado na pCDK2 e CDK2. E, ainda que ligada à sua ciclina, não se observa a movimentação da hélice C e do segmento de ativação, necessários para sua ativação. Isso alimenta a hipótese da necessidade de fatores externos para completa ativação da CDK4.

PERSPECTIVA PARA TRABALHOS FUTUROS

Como perspectivas, será de extrema importância a continuidade dos estudos estruturais acerca dos complexos CDK-ciclina, principalmente dos membros mais recentemente descobertos e que não possuem estruturas cristalográficas ainda determinadas, ou determinadas parcialmente, considerando apenas o domínio quinase. São estudos desafiadores, devido à complexidade dessas CDKs, e a bioinformática poderá cada vez mais colaborar para a obtenção de informações estruturais cruciais para a descoberta de estratégias para a inibição da atividade dessas importantes proteínas quinases dependentes de ciclinas.

REFERÊNCIAS

- Adams, P. D., Li, X., Sellers, W. R., Baker, K. B., Leng, X., Harper, J. W., Taya, Y., & Kaelin, W. G. (1999). Retinoblastoma Protein Contains a C-terminal Motif That Targets It for Phosphorylation by Cyclin-cdk Complexes. *Molecular and Cellular Biology*. <https://doi.org/10.1128/mcb.19.2.1068>
- Baumli, S., Lolli, G., Lowe, E. D., Troiani, S., Rusconi, L., Bullock, A. N., Debreczeni, J. É., Knapp, S., & Johnson, L. N. (2008). The structure of P-TEFb (CDK9/cyclin T1), its complex with flavopiridol and regulation by phosphorylation. *EMBO Journal*. <https://doi.org/10.1038/emboj.2008.121>
- Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T. G., Bertoni, M., Bordoli, L., & Schwede, T. (2014). SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research*, 42(W1), 252–258. <https://doi.org/10.1093/nar/gku340>
- Brown, Nicholas R., Korolchuk, S., Martin, M. P., Stanley, W. A., Moukhametzianov, R., Noble, M. E. M., & Endicott, J. A. (2015). CDK1 structures reveal conserved and unique features of the essential cell cycle CDK. *Nature Communications*, 6, 1–12. <https://doi.org/10.1038/ncomms7769>
- Brown, Nicholas R., Noble, M. E. M., Lawrie, A. M., Morris, M. C., Tunnah, P., Divita, G., Johnson, L. N., & Endicott, J. A. (1999). Effects of phosphorylation of threonine 160 on cyclin-dependent kinase 2 structure and activity. *Journal of Biological Chemistry*. <https://doi.org/10.1074/jbc.274.13.8746>
- Brown, Nick R., Lowe, E. D., Petri, E., Skamnaki, V., Antrobus, R., & Johnson, L. N. (2007). Cyclin B and cyclin A confer different substrate recognition properties on CDK2. *Cell Cycle*, 6(11), 1350–1359. <https://doi.org/10.4161/cc.6.11.4278>
- Brown, Nick R., Noble, M. E. M., Endicott, J. A., & Johnson, L. N. (1999). The structural basis for specificity of substrate and recruitment peptides for cyclin-dependent kinases. *Nature Cell Biology*. <https://doi.org/10.1038/15674>
- Cao, L., Chen, F., Yang, X., Xu, W., Xie, J., & Yu, L. (2014). Phylogenetic analysis of CDK and cyclin proteins in premetazoan lineages. *BMC Evolutionary Biology*, 14(1). <https://doi.org/10.1186/1471-2148-14-10>
- Chen, P., Lee, N. V., Hu, W., Xu, M., Ferre, R. A., Lam, H., Bergqvist, S., Solowiej, J., Diehl, W., He, Y. A., Yu, X., Nagata, A., Vanarsdale, T., & Murray, B. W. (2016). Spectrum and degree of CDK drug interactions predicts clinical performance. *Molecular Cancer Therapeutics*, 15(10), 2273–2281. <https://doi.org/10.1158/1535-7163.MCT-16-0300>
- Cheng, K. Y., Noble, M. E. M., Skamnaki, V., Brown, N. R., Lowe, E. D., Kontogiannis, L., Shen, K., Cole, P. A., Siligardi, G., & Johnson, L. N. (2006). The role of the phospho-CDK2/cyclin A recruitment site in substrate recognition. *Journal of Biological Chemistry*. <https://doi.org/10.1074/jbc.M600480200>

Cho, Y. S., Borland, M., Brain, C., Chen, C. H. T., Cheng, H., Chopra, R., Chung, K., Groarke, J., He, G., Hou, Y., Kim, S., Kovats, S., Lu, Y., O'Reilly, M., Shen, J., Smith, T., Trakshel, G., Vögtle, M., Xu, M., ... Sung, M. J. (2010). 4-(pyrazol-4-yl)-pyrimidines as selective inhibitors of cyclin-dependent kinase 4/6. *Journal of Medicinal Chemistry*. <https://doi.org/10.1021/jm100571n>

Chojnacki, S., Cowley, A., Lee, J., Foix, A., & Lopez, R. (2017). Programmatic access to bioinformatics tools from EMBL-EBI update: 2017. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkx273>

Day, P. J., Cleasby, A., Tickle, I. J., O'Reilly, M., Coyle, J. E., Holding, F. P., McMenamin, R. L., Yon, J., Chopra, R., Lengauer, C., & Jhoti, H. (2009). Crystal structure of human CDK4 in complex with a D-type cyclin. *Proceedings of the National Academy of Sciences of the United States of America*, 106(11), 4166–4170. <https://doi.org/10.1073/pnas.0809645106>

Diniz, W. J. S., & Canduri, F. (2017). Bioinformatics: An overview and its applications. *Genetics and Molecular Research*, 16(1). <https://doi.org/10.4238/gmr16019645>

Errico, A., Deshmukh, K., Tanaka, Y., Pozniakovsky, A., & Hunt, T. (2010). Identification of substrates for cyclin dependent kinases. *Advances in Enzyme Regulation*, 50(1), 375–399. <https://doi.org/10.1016/j.advenzreg.2009.12.001>

Esposito, L., Indovina, P., Magnotti, F., Conti, D., & Giordano, A. (2013). Anticancer Therapeutic Strategies Based on CDK Inhibitors. *Current Pharmaceutical Design*. <https://doi.org/10.2174/13816128113199990377>

Goda, T., Funakoshi, M., Suhara, H., Nishimoto, T., & Kobayashi, H. (2001). The N-terminal Helix of Xenopus Cyclins A and B Contributes to Binding Specificity of the Cyclin-CDK Complex. *Journal of Biological Chemistry*. <https://doi.org/10.1074/jbc.M011101200>

Gupta, O.P. Study and analysis of various bioinformatics applications using protein BLAST: an overview. *Adv. Comput. Sci. Technol.* 10(8), 2587–2601 (2017)

Honda, R., Lowe, E. D., Dubinina, E., Skamnaki, V., Cook, A., Brown, N. R., & Johnson, L. N. (2005). The structure of cyclin E1/CDK2: Implications for CDK2 activation and CDK2-independent roles. *EMBO Journal*, 24(3), 452–463. <https://doi.org/10.1038/sj.emboj.7600554>

Kitagawa, M., Higashi, H., Jung, H. K., Suzuki-Takahashi, I., Ikeda, M., Tamai, K., Kato, J. Y., Segawa, K., Yoshida, E., Nishimura, S., & Taya, Y. (1996). The consensus motif for phosphorylation by cyclin D1-Cdk4 is different from that for phosphorylation by cyclin A/E-Cdk2. *EMBO Journal*. <https://doi.org/10.1002/j.1460-2075.1996.tb01097.x>

Li, Y., Zhang, J., Gao, W., Zhang, L., Pan, Y., Zhang, S., & Wang, Y. (2015). Insights on structural characteristics and ligand binding mechanisms of CDK2. *International Journal of Molecular Sciences*, 16(5), 9314–9340. <https://doi.org/10.3390/ijms16059314>

Liao, Y., Feng, Y., Shen, J., Hornicek, F. J., & Duan, Z. (2016). The roles and therapeutic potential of cyclin-dependent kinases (CDKs) in sarcoma. *Cancer and Metastasis Reviews*, 35(2), 151–163. <https://doi.org/10.1007/s10555-015-9601-1>

Macaluso, M., Montanari, M., & Giordano, A. (2006). Rb family proteins as modulators of gene expression and new aspects regarding the interaction with chromatin remodeling enzymes. In *Oncogene*. <https://doi.org/10.1038/sj.onc.1209680>

Malumbres, M. (2014). *Cyclin-dependent kinases*. 1–10.

Malumbres, M., & Barbacid, M. (2009). Cell cycle, CDKs and cancer: A changing paradigm. *Nature Reviews Cancer*, 9(3), 153–166. <https://doi.org/10.1038/nrc2602>

McGrath, D. A., Balog, E. R. M., Kõivomägi, M., Lucena, R., Mai, M. V., Hirschi, A., Kellogg, D. R., Loog, M., & Rubin, S. M. (2013). Cks confers specificity to phosphorylation-dependent CDK signaling pathways. *Nature Structural and Molecular Biology*, 20(12), 1407–1414. <https://doi.org/10.1038/nsmb.2707>

Nelson, David L.; COX, Michael M. Princípios de bioquímica de Lehninger. Porto Alegre: Artmed, 2011. 6. ed. Porto Alegre: Artmed, 2014.

O'Leary, B., Finn, R. S., & Turner, N. C. (2016). Treating cancer with selective CDK4/6 inhibitors. In *Nature Reviews Clinical Oncology*. <https://doi.org/10.1038/nrclinonc.2016.26>

Pan, W., Cox, S., Hoess, R. H., & Grafström, R. H. (2001). A cyclin d1/cyclin-dependent kinase 4 binding site within the c domain of the retinoblastoma protein. *Cancer Research*.

Paula, B. de H. R. de, Bandeira, L. L. B., Ramos, A. P., Pinheiro, B. L. M., Ferreira, R. M., Araújo, I. I. de O., Costa, P. C. V. da, & Crocamo, S. (2019). Opções Contemporâneas de Rastreo da Cardiotoxicidade Relacionada a Tratamentos Oncológicos. *Revista Brasileira de Cancerologia*. <https://doi.org/10.32635/2176-9745.rbc.2019v65n3.388>

Pearl, L. H. (2005). Hsp90 and Cdc37 - A chaperone cancer conspiracy. In *Current Opinion in Genetics and Development*. <https://doi.org/10.1016/j.gde.2004.12.011>

Risler, J.-L. (2002). Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins. *Computers & Chemistry*. [https://doi.org/10.1016/s0097-8485\(02\)00027-x](https://doi.org/10.1016/s0097-8485(02)00027-x)

Schulze-Gahmen, U., & Kim, S. H. (2002). Structural basis for CDK6 activation by a virus-encoded cyclin. *Nature Structural Biology*. <https://doi.org/10.1038/nsb756>

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*. <https://doi.org/10.1038/msb.2011.75>

Singh, B., & Wu, P. Y. J. (2019). Regulation of the program of DNA replication by CDK: new findings and perspectives. *Current Genetics*, 65(1), 79–85. <https://doi.org/10.1007/s00294-018-0860-6>

Song, H., Hanlon, N., Brown, N. R., Noble, M. E. M., Johnson, L. N., & Barford, D. (2001). Phosphoprotein-protein interactions revealed by the crystal structure of kinase-associated phosphatase in complex with phosphoCDK2. *Molecular Cell*. [https://doi.org/10.1016/S1097-2765\(01\)00208-8](https://doi.org/10.1016/S1097-2765(01)00208-8)

Takeda, D. Y., Wohlschlegel, J. A., & Dutta, A. (2001). A bipartite substrate recognition motif for cyclin-dependent kinases. *Journal of Biological Chemistry*. <https://doi.org/10.1074/jbc.M005719200>

Tarricone, C., Dhavan, R., Peng, J., Areces, L. B., Tsai, L. H., & Musacchio, A. (2001). Structure and regulation of the CDK5-p25nck5a complex. *Molecular Cell*. [https://doi.org/10.1016/S1097-2765\(01\)00343-4](https://doi.org/10.1016/S1097-2765(01)00343-4)

Wang, X., Deng, K., Wang, C., Li, Y., Wang, T., Huang, Z., Ma, Y., Sun, P., Shi, Y., Yang, S., Fan, Y., & Xiang, R. (2020). Novel CDKs inhibitors for the treatment of solid tumour by simultaneously regulating the cell cycle and transcription control. *Journal of Enzyme Inhibition and Medicinal Chemistry*, 35(1), 414–423. <https://doi.org/10.1080/14756366.2019.1705290>

Webb, B., & Sali, A. (2017). Protein structure modeling with MODELLER. In *Methods in Molecular Biology*. https://doi.org/10.1007/978-1-4939-7231-9_4

Wheeler, T. J., Clements, J., & Finn, R. D. (2014). Skylign: A tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models. *BMC Bioinformatics*. <https://doi.org/10.1186/1471-2105-15-7>

Zhao, L., Yuan, X., Wang, J., Feng, Y., Ji, F., Li, Z., & Bian, J. (2019). A review on flavones targeting serine/threonine protein kinases for potential anticancer drugs. *Bioorganic and Medicinal Chemistry*, 27(5), 677–685. <https://doi.org/10.1016/j.bmc.2019.01.027>