

KARISA KARLA MANHANI

BIOINFORMÁTICA ESTRUTURAL APLICADA AO ESTUDO DE
QUINASES COMPLEXADAS COM INIBIDORES



1910001405



KARISA KARLA MANHANI

**BIOINFORMÁTICA ESTRUTURAL APLICADA AO ESTUDO DE
QUINASES COMPLEXADAS COM INIBIDORES**

Dissertação de Mestrado em Biofísica Molecular –
Área de Concentração em Biofísica Molecular,
Departamento de Física do Instituto de Biociências,
Letras e Ciências Exatas - Universidade Estadual
Paulista "Júlio Mesquita Filho" – UNESP



T. 1405/05

ORIENTADORA: PROF.^A DR.^A FERNANDA CANDURI

São José do Rio Preto-SP
2005

577.112

M278b

13/06/2005

Doação

R\$ 30,00

Autor(a)

076/2005

162/05

20801017

UNIVERSIDADE ESTADUAL PAULISTA
CÂMPUS DE SÃO JOSÉ DO RIO PRETO
INSTITUTO DE BIOCIÊNCIAS, LETRAS E CIÊNCIAS EXATAS

unesp

KARISA KARLA MANHANI

BIOINFORMÁTICA ESTRUTURAL APLICADA AO ESTUDO DE
QUINASES COMPLEXADAS COM INIBIDORES

Manhani, Karisa Karla.

Bioinformática estrutural aplicada ao estudo de quinases complexadas
com inibidores / Karisa Karla Manhani – São José do Rio Preto : [s.n.],
2005

150 f. : il. : 30 cm.

Orientadora: Fernanda Canduri

Dissertação (mestrado) – Universidade Estadual Paulista. Instituto de
Biociências, Letras e Ciências Exatas

1. Proteínas. 2. Ciclo celular. 3. Estrutura molecular. 4. Modelagem
molecular. 5. Drogas - Desenho. 6. Bioinformática. 7. Inibidores
químicos. I. Canduri, Fernanda. II. Universidade Estadual Paulista.
Instituto de Biociências, Letras e Ciências Exatas. III. Título.

CDU – 577.112

São José do Rio Preto-SP
2005



MENSAGEM

*"Pode ser que um dia deixemos de nos falar...
Mas, enquanto houver amizade,
Faremos as pazes de novo.*

*Pode ser que um dia o tempo passe...
Mas, se a amizade permanecer,
Um do outro há de se lembrar.*

*Pode ser que um dia nos afastemos...
Mas, se formos amigos de verdade,
A amizade nos reaproximará.*

*Pode ser que um dia não mais existamos...
Mas, se ainda sobrar amizade,
Nasceremos de novo, um para o outro.*

*Pode ser que um dia tudo acabe...
Mas, com a amizade construiremos tudo novamente,
Cada vez de forma diferente,
Sendo único e inesquecível cada momento.
Que juntos viveremos nos lembraremos pra sempre*

Dedicatória

Dedico este trabalho, aos meus pais Firmino e Doraci, ao meu esposo André Luís, a minha filha que amo muito Jackeline Rany, e a uma amiga muito especial Maria Carolina. Muito obrigada pelo apoio, amor, compreensão e amizade durante esta etapa da minha vida.

MENSAGEM

"Pode ser que um dia deixemos de nos falar...

Mas, enquanto houver amizade,

Faremos as pazes de novo.

Pode ser que um dia o tempo passe...

Mas se a amizade permanecer,

Um do outro há de se lembrar.

Pode ser que um dia nos afastemos...

Mas, se formos amigos de verdade,

A amizade nos reaproximará.

Pode ser que um dia não mais existamos...

Mas, se ainda sobrar amizade,

Nasceremos de novo, um para o outro.

Pode ser que um dia tudo acabe...

Mas, com a amizade construiremos tudo novamente,

Cada vez de forma diferente,

Sendo único e inesquecível cada momento

Que juntos viveremos nos lembraremos pra sempre

Há duas formas para viver sua vida:

Uma é acreditar que não existe milagre.

A outra é acreditar que todas as coisas são um milagre."

Albert Einstein – 1879-1955



Meus Agradecimentos

A Doutora Fernanda Canduri, pela orientação, pelo companheirismo, e, sobretudo pela amizade e confiança conquistados durante estes anos e o mais importante de tudo, seus conselhos para as minhas futuras decisões, sendo elas profissional ou pessoal,

Aos colegas do laboratório de Sistemas Biomoleculares pela compreensão e atenção dados quando requisitados,

Aos colegas da pós-graduação que de alguma maneira contribuíram para a realização deste trabalho em especial ao Márcio pela paciência e disponibilidade, quando foi pedido à ele ajuda para manuseio de alguns programas de visualização e ao Hugo pela compreensão na troca de informações à respeito de modelagem que resultou em um acréscimo e potencialidade do Parmodel,

Aos colegas de sala do curso de pós-graduação, Doutorando Nelson José Freitas da Silveira e ao Mestrando José Renato Panzarini Junior pela compreensão, companheirismo, confiança, amizade e prestatividade ao serem requisitados para a realização deste trabalho,

A secretária Ilva e aos técnicos Barbosa e Paulinho do departamento de Física, pela atenção dada quando requisitados,

A Capes pelo auxílio da bolsa concedida e à Fapesp, pelos projetos aprovados para equipar o laboratório de bioinformática,

Aos meus pais pelo grande apoio para a busca de uma profissão, esforço e amor para que eu conseguisse finalizar um sonho desejado e pela compreensão oferecida para minha maternidade,

Ao amor de meu esposo André Luis que me alimenta e me conforta, pela felicidade de tê-lo ao meu lado na luta pela sobrevivência e principalmente pelo apoio, compreensão e conselhos oferecidos nas minhas decisões pessoais e profissionais, sobretudo por ter compreendido a minha ausência algumas vezes devido à realização deste projeto,

A minha querida filha Jackeline Rany pela bondade, paciência e compreensão pelos dias de ausência para então realização do projeto,

As novas amizades que fiz durante estes anos de dedicação à pesquisa, como Giovanni Cesar que me orientou de maneira sábia qual o melhor caminho a seguir para minha vida profissional e que partir de agora fará parte da minha carreira profissional,

Índice Ao Professor Doutor Elói que sempre me guiou e orientou para seguir as melhores oportunidades, desde da época de graduação até aos momentos de hoje sempre com sua bondade, confiança e compreensão e que fará parte da minha carreira profissional daqui por diante, 1

ABSTRACT Ao meu e sempre amigo Professor Aledir, que desde o segundo ano de faculdade sempre me orientou e sabiamente me ofereceu oportunidades imperdíveis de aprendizagem que me fizeram agregar valores à minha vida pessoal e ao meu profissionalismo, 2

1.1 A todos meus amigos desde à época de graduação: Élio, Régis, Serginho e Júlio que mesmo distantes não nos afastamos e sempre preservando esta amizade verdadeira, 3

A família de meu esposo que tomei como minha segunda família, desde quando o conheci, e que sempre estiveram presentes nos momentos felizes e complicados, 7

Em especial a minha sogra Lázara que sempre considerei como minha mãe, amiga, conselheira e essencialmente uma super mãe, pois cuida de todos com muita atenção e carinho sem fazer distinção, eu devo muito a ela pela realização deste trabalho, 11

Ao meu Irmão Renato César e cunhada Adriana e sobrinho Bruno Henrique pelos conselhos e principalmente pelo auxílio à minha maternidade, 16

A Deus, pela oportunidade de possibilitar a realização de todos os meus desejos, pelo sepre de vida que me alimenta todos os dias e o mais importante de todos, a enorme força fornecida durante os problemas que passei na minha vida, me levando a permanecer viva, determinada e perspicaz, 21

1.5. DESCRIÇÕES DOS INIBIDORES 26

1.5.1. Flavopiridol 27

1.5.2. H717 28

1.5.3. Himenialdisina 29

1.5.4. Indirubina-5-Sulfato 29

1.5.5. Isopenteniladenina 30

1.5.6. Olomoucina 31

1.5.7. Oxidol 91 32

1.5.8. Purvanalol B 32

1.5.9. Roscovitina 34

1.5.10. Estaursporina 34

2. OBJETIVOS 36

Índice Geral	Páginas
RESUMO	x
ABSTRACT	xi
1. INTRODUÇÃO	
1.1. BIOINFORMÁTICA	1
1.2. APLICAÇÕES	3
1.3. DIVISÃO CELULAR	
1.3.1. Ciclo Celular	5
1.3.2. Controle do Ciclo Celular	7
1.3.3. Os gerentes do sistema controle são formados por quinases dependentes de ciclinas	8
1.4. DESCRIÇÃO DAS QUINASES	
1.4.1. Quinases Dependentes de Ciclinas (CDKs)	16
1.4.1.1. Quinase Dependente de Ciclina 2 (CDK2)	18
1.4.1.2. Quinase Dependente de Ciclina 4 (CDK4)	20
1.4.2. Proteína Quinase 6 do <i>Plasmodium falciparum</i> (PfPK6)	21
1.5. DESCRIÇÕES DOS INIBIDORES	26
1.5.1. Flavopiridol	27
1.5.2. H717	28
1.5.3. Himenialdisina	29
1.5.4. Indirubina-5-Sulfonato	29
1.5.5. Isopenteniladenina	30
1.5.6. Olomoucina	31
1.5.7. Oxindol 91	32
1.5.8. Purvanalol B	32
1.5.9. Roscovitina	34
1.5.10. Estaurosporina	34
2. OBJETIVOS	36



3. MÉTODOS E FERRAMENTAS	
3.1. Modelagem Comparativa por Homologia	37
3.2. Ferramentas Utilizadas Durante o Processo da Modelagem Molecular	47
4. RESULTADOS E DISCUSSÕES	
4.1. Modelagem da CDK4/Ciclina D complexada com inibidores	55
4.1.1. Construção dos Modelos	55
4.1.2. Qualidade dos Modelos	58
4.1.3. Descrição dos Modelos	65
4.1.4. Interação CDK e Ciclina	66
4.1.5. Interação CDK_Inibidores	76
4.2. Modelagem da PfPK6 Complexada com Roscovitina e Olomoucina	87
4.2.1 Construção dos Modelos	87
4.2.2 Qualidade dos Modelos	88
4.2.3. Descrição dos Modelos	91
4.2.4. Interação dos Inibidores com a PfPK6	93
5. CONCLUSÃO E PROPOSTAS FUTURAS	
5.1. Conclusões	100
5.1.1. CDK4/Ciclina D	100
5.1.2. PfPK6	101
5.2. Propostas Futuras	101
6. REFERÊNCIAS BIBLIOGRÁFICAS	103
7. APÊNDICE: Informações sobre o programa MODELLER	111
7.1. Derivação de uma característica pdf através de pdfs básicas	116
7.2. Derivação de uma pdf molecular a partir das pdfs características	117
7.3. Otimização da pdf molecular	118



ANEXO A: ARTIGOS:

A) Molecular model of protein kinase 6 from Plasmodium falciparum 122

B) Molecular models of the Cyclin-Dependent Kinase 4 associated the Cyclin D and Complexed with Inhibitors 122

Tabela 1: Programas e seus respectivos resultados para a modelagem molecular comparativa 27

Tabela 2: Inibidores de CDKs 1, 2, 4 e 5 e os respectivos valores de IC50 em μM 27

Tabela 3: Inibidores de CDK4 e os respectivos valores de IC50 56

Tabela 4: Análises da qualidade estereoquímica e a média dos valores do Fator G para os moldes e modelos 64

Tabela 5: Ligações de hidrogênio intermoleculares entre CDK4 e inibidores 77

Tabela 6: Valores de RMSD calculados pela sobreposição dos bolsões de ligação dos moldes e modelos 80

Tabela 7: Resumo das análises estruturais 82

Tabela 8: Análise da qualidade estereoquímica e a média dos valores do Fator G para modelos e moldes 90

Tabela 9: Valores obtidos pelo método Perfil 3D para os moldes e modelos 90

Tabela 10: Valores de r.m.s.d. para a geometria ideal 91

Tabela 11: Ligações de hidrogênio intermoleculares 93

Tabela 12: Valores de RMSD calculados pela sobreposição dos bolsões de ligação dos moldes e modelos 97

Tabela 13: Análise estrutural dos moldes e modelos comparada com valores funcionais de IC50 97

INDICE DE TABELAS

Figura 1: Fases do ciclo celular20

Tabela 1: Programas e serviços da internet utilizados em modelagem molecular comparativa2

Tabela 2: Inibidores de CDKs 1, 2, 4 e 5 e os respectivos valores de IC50 em μM 27

Figura 4: Proteínas quinases conectadas pelo sequenciamento do Genoma

Tabela 3: Inibidores de CDK4 e os respectivos valores de IC5056

Tabela 4: Análises da qualidade estereoquímica e a média dos valores do Fator G para os moldes e modelos64

Figura 5: Ciclinas se associam ao protossoma através de moléculas de

Tabela 5: Ligações de hidrogênio intermoleculares entre CDK4 e inibidores77

Tabela 6: Valores de RMSD calculados pela sobreposição dos bolsões de ligação dos moldes e modelos.....80

Tabela 7: Resumo das análises estruturais.....82

Tabela 8: Análise da qualidade estereoquímica e a média dos valores do Fator G para modelos e moldes.....90

Figura 6: Representação esquemática do ciclo celular mostrando a atuação

Tabela 9: Valores obtidos pelo método Perfil 3D para os moldes e modelos.....90

Tabela 10: Valores de r.m.s.d. para a geometria ideal.....91

Figura 7: Mecanismo de fosforilação e desfosforilação18

Tabela 11: Ligações de hidrogênio intermoleculares93

Figura 11: Comparação das estruturas cristalográficas da CDK2;

Tabela 12: Valores de RMSD calculados pela sobreposição dos bolsões de ligação dos moldes e modelos.....97

Tabela 13: Análise estrutural dos moldes e modelos comparada com valores funcionais de IC50.....97

Figura 13: Diagrama esquemático do Garfo Molecular20

Figura 14: Mapa de incidência da Malária23

Figura 15: Esquema do ciclo de vida e transmissão da Malária pelo Mosquito transmissor *Anopheles gambiae*24

Figura 16: Elementos de estrutura secundária da estrutura monomérica da PIPK524

ÍNDICE DE FIGURAS

Figura 1: Fases do ciclo celular	6
Figura 2: Comparação entre meiose e mitose	6
Figura 3: Regulação das CDKs através da ciclina	9
Figura 4: Proteínas quinases conhecidas pelo seqüenciamento do Genoma Humano	10
Figura 5: Ativação do fator promotor de mitose	12
Figura 6: Ciclinas são degradadas no proteossoma através de moléculas de Ubiquitina	12
Figura 7: Mudança conformacional no sítio catalítico da CDK2, provocado pela interação com a ciclina A	13
Figura 8: Processo de inibição pelas CKIs. A) CKI liga-se a CDK impedindo a associação à ciclina; B) CKI liga-se ao complexo CDK-ciclina impedindo a atividade quinásica.....	14
Figura 9: Representação esquemática do ciclo celular mostrando a atuação das CDKs nas respectivas fases do ciclo celular.....	15
Figura 10: a) Fórmula estrutural da Trifosfato de Adenosina (ATP) e b) Mecanismo da fosforilação e desfosforilação	16
Figura 11: Comparação das estruturas cristalográficas da CDK2; a) Estrutura monomérica da CDK2; b) Estrutura dimérica da CDK2/ciclinaA, b) mostrando a mudança conformacional do sítio de ligação do ATP	19
Figura 12: a) Complexo CDK2:ATP mostrando o sítio de ligação do ATP e b) Detalhes das ligações de hidrogênio entre ATP e CDK2	19
Figura 13: Diagrama esquemático do Garfo Molecular	20
Figura 14: Mapa de incidência da Malária	23
Figura 15: Esquema do ciclo de vida e transmissão da Malária pelo Mosquito transmissor <i>Anopheles gambiae</i>	24
Figura 16: Elementos de estrutura secundária da estrutura monomérica da PfPK5	24



Figura 17: Fórmulas estruturais: a) alcalóide flavonóide originado de uma planta (*Dysoxylum binectariferum*); b) Flavopiridol28

Figura 18: Fórmula estrutural do H71728

Figura 19: Fórmula estrutural da himenialdisina29

Figura 20: Fórmula estrutural da Indirubina-5-Sulfonato.....30

Figura 21: Fórmula estrutural da Isopenteniladenina31

Figura 22: Fórmula estrutural da Olomoucina.....32

Figura 23: Fórmula estrutural da Oxindol9132

Figura 24: Fórmula estrutural do Purvanalol B.....33

Figura 25: ATP (amarelo) e Purvanalol (preto). As comparações são baseadas na superposição de C α da CDK2. A posição da purina é conservada ..33

Figura 26: Fórmula estrutural da roscovitina34

Figura 27: Fórmula estrutural da estaurosporina35

Figura 28: Esquema dos passos da modelagem molecular40

Figura 29: Alinhamento de seqüências primárias da proteína-molde com a proteína- alvo exibindo "gaps"41

Figura 30: Página de entrada para submissão dos moldes e seqüências alvos...49

Figura 31: Página para submissão dos modelos para validação no Parmodel.....51

Figura 32: Esquema da cadeia polipeptídica indicando os ângulos Φ e ψ 58

Figura 33: Alinhamentos de seqüências primárias: a) alinhamento da CDK2 (1HQ1) com a CDK4; b) alinhamento da Ciclina A (1HQ1) com a Ciclina D5653

Figura 34: Esquema mostrando os passos para obter os moldes para CDK4/ciclinaD57

Figura 35: Protocolo da Modelagem molecular comparativa utilizado para modelar a CDK4/ciclina D em complexo com inibidores58



Figura 36: Gráficos de Ramachandran das 10 estruturas utilizadas como moldes.	59-60
Figura 37: Gráficos de Ramachandran dos 10 melhores modelos de CDK4/ciclinaD complexados com os inibidores.	61-62
Figura 38: a) Diagrama de fitas para CDK2/Ciclina A complexada com Flavopiridol mostrando a região do sítio de ligação.....	67
Figura 38: b) Diagrama de fitas do sítio de ligação do complexo CDK2/ciclina A e seus inibidores.....	68
Figura 39: a) Diagrama de fitas para CDK4/ciclina D complexada com Flavopiridol mostrando a região do sítio de ligação.....	69
Figura 39: b) Diagrama de fitas do sítio de ligação do complexo CDK4/ciclina D e seus inibidores	70
Figura 40: a) Alinhamento do molde (1H1Q) com o modelo (CDK4)	71
Figura 40: b) Alinhamento da ciclina A (1H1Q) com a ciclina D	72
Figura 41: Região do bolsão de ligação da CDK4 mostrando a interface com a ciclinaD.	73
Figura 42: Sobreposição dos Carbonos α do modelo da CDK4 (linha vermelha) com a CDK2 hibridizada (linha verde), mostrando a posição da Thr fosforilável.....	75
Figura 43: Imagens das regiões de interface entre CDK e Ciclina; a) CDK2 e Ciclina A; b) CDK4 e Ciclina D	76
Figura 44: Sobreposições de C α dos sítios catalíticos.....	79
Figura 45: a) Potencial eletrostático da CDK2/ciclina A complexada com o inibidor H717	83
Figura 45: b) Potencial eletrostático do sítio de ligação da CDK2.....	84
Figura 46: a) Potencial eletrostático da CDK4/ciclina D complexada com inibidor H717.....	85
Figura 46: b) Potencial eletrostático do sítio de ligação da CDK4.....	86



Figura 47: Alinhamento de seqüências primárias da CDK2 com a PfPK6.....87

Figura 48: Protocolo para construção dos modelos de PfPK6.....88

Figura 49: Gráficos de Ramachandran: a) Modelos de PfPK6 complexados com roscovitina e olomoucina, respectivamente; b) Estruturas de CDK2 complexadas com roscovitina e olomoucina, respectivamente89

Figura 50: Elementos de estrutura secundária: a) PfPK6:roscovitina; b) CDK2:roscovitina; c) PfPK6:olomoucina e d) CDK2:olomoucina95

Figura 51: Sobreposições dos Carbonos α do bolsão de ligação.....96

Figura 52: Sobreposições dos carbonos α do bolsão de ligação 102

Figura 53: Par de Tirosinas na entrada do bolsão de ligação da PfPK6 a) na presença da roscovitina; b) na presença da olomoucina.....98

Figura 54: Superfície mostrando o potencial eletrostático de a) PfPK6 com 1-roscovitina e 2-olomoucina; b) CDK2 com 3-roscovitina e 4-olomoucina99

Figura 55: Principais passos do processo de modelagem 118

Figura 56: Mudança de conformação pela otimização da função objetivo..... 119

NPP - Fator promotor de mitose

CKI -Proteínas inibidoras de CDKs

CAK -Proteínas ativadoras de CDKs

PKC - Proteína quinase C

ÍNDICE DE ABREVIACES

3D	- Tridimensional
ATP	- Trifosfato de Adenosina
CDK	- Quinase dependente de ciclina
DNA	- Ácido desoxirribonucléico
IC50	- Concentraco inibitria mnima 50%
NMR	- Ressonncia Magnética Nuclear
PDB	- Banco de dados de protenas
PDF	- funço de densidade de probabilidade
Rb	- Retinoblastoma
RMSD	-Desvio quadrático médio
RNA	- Ácido ribonucléico
PfPK	- Protena quinase de <i>Plasmodium falciparum</i>
MPF	- Fator promotor de mitose
CKI	-Protenas inibidoras de CDKs
CAK	-Protenas ativadoras de CDKs
PKC	- Protena quinase C

RESUMO

O presente projeto de mestrado visa a modelagem molecular de quinases, a quinase dependente de ciclina 4 (CDK4/ciclina D) complexada com inibidores e a proteína quinase 6 do *Plasmodium falciparum* (PfPK6) complexada com Roscovitina e Olomoucina que foram comparadas com outras estruturas modeladas pelo grupo do Laboratório de Sistemas Biomoleculares, tais como, CDK1 e CDK5 complexadas com roscovitina e com a estrutura CDK2 complexada com inibidores e também associada com ciclina A. CDKs desempenham uma função essencial na regulação do ciclo de divisão celular. Consistem de uma subunidade catalítica (CDK1-13) e uma subunidade regulatória (ciclina A-H). Cada complexo CDK/ciclina participa de uma fase específica do ciclo celular ou do processo de transcrição. Sendo assim, as CDKs estão diretamente ligadas a processos tumores-genes. Desta forma, CDKs e PfPK6 são alvos para o desenho de drogas. PfPK6 é uma proteína presente no *Plasmodium falciparum*, parasita responsável pela malária, uma doença negligenciada pelo mundo, apesar de matar 1,5 a 2,7 milhões de pessoas no mundo. O estudo estrutural destas quinases tem o objetivo de propor inibidores específicos para cada enzima, gerando drogas eficazes que irão controlar a ação desta importante família de proteínas.

Este projeto de mestrado está vinculado ao projeto Genoma Estrutural (FAPESP) que tem como objetivo fornecer a estrutura tridimensional determinada experimentalmente ou por modelagem, para as proteínas passíveis de estudo, que são codificadas pelo genoma. Espera-se que as estruturas das proteínas determinadas experimentalmente e os modelos computacionais produzam avanços no entendimento da função molecular e do mecanismo de muitas proteínas.



ABSTRACT

The present work of aims at the molecular modeling of Kinases, among them cyclin dependent Kinase 4 (CDK4/cyclin D) complexed with inhibitors and protein Kinase 6 isolated from *Plasmodium falciparum* (PfPK6) complexed with Roscovitine and Olomoucine. These models had been compared with other structural models performed by Laboratory of Biomolecular Systems group, such as, CDK1 and CDK5 complexed with roscovitine and CDK2 structure complexed with inhibitors and associated with cyclin A. CDKs plays an essential function in the regulation of the cycle of cellular division.

They consist of a catalytic subunit (CDK1-13) and a regulatory subunit (cyclin A-H). Each complex CDK/cyclin has a function in a specific fase of the cellular cycle. CDKs 9 and 10 are involved in the transcription process. Thus, the CDKs are related with carcinogenic processes. Then Kinases are target for drugs design. PfPK6 is a protein Kinase from *Plasmodium falciparum*, parasite responsible for the malaria, an disease neglected by world, although to kill from 1,5 to 2,7 million of people on the world. The structural study of these Kinases in presence of specific inhibitors is the purpose, generating efficient drugs that will go to control the action of this important protein family.

This work is part of Structural Genoma project that has the objective to supply the three-dimensional structure solved experimentally or for modelling to the proteins passive of study, that they are codified in the genoma. One expects the computational structures of proteins solved experimentally and modeled produce advances in the understanding of the molecular function and in the mechanism of many proteins.



1. INTRODUÇÃO

1.1 Bioinformática (BAYAT, 2002)

A partir do estudo estrutural foi possível determinar as estruturas tridimensionais de muitas proteínas, cujas funções moleculares não poderiam ser compreendidas usando métodos convencionais. A importância da criação de bancos de dados, as descrições de seqüências de aminoácidos e a resolução de estruturas de proteínas vêm sendo valorizadas e enfatizadas (KINOSHITA e NAKAMURA, 2003).

As análises dos dados de seqüências genômicas emergindo a partir do projeto genoma humano são referências para realizações na área de Bioinformática.

O projeto Genoma Estrutural tem como objetivo a determinação da estrutura tridimensional de proteínas no seu estado nativo e na presença de ligantes.

A bioinformática é essencial para o direcionamento de dados em biologia moderna e para a medicina. Trata-se da aplicação de ferramentas da computação para uma melhor compreensão e interpretação dos dados biológicos.

As principais ferramentas da bioinformática são programas disponíveis na internet, (tabela1). Exemplos que podem ser ressaltados: transferência de informação estrutural por homologia através da modelagem comparativa e técnicas de "threading". O primeiro exemplo trata-se do uso de estruturas conhecidas, resolvidas por cristalografia ou ressonância magnética nuclear como molde para gerar modelos parecidos ou tão próximos da estrutura real, fazendo-se inicialmente, um alinhamento de suas seqüências primárias. O segundo exemplo referido acima, trata-se de uma técnica que faz procura por moldes através da estrutura secundária, isso quando não for possível encontrar a identidade pela seqüência primária (MARTI-RENOM et al., 2004).

Tabela 1: Programas e serviços da internet utilizados em modelagem molecular comparativa (MARTI-RENOM et al., 2004)

Nome	Endereços de sites da internet
BANCOS DE DADOS	
CATH	www.biochem.ucl.ac.uk/bsm/cath/
GenBank	www.ncbi.nlm.nih.gov/Genbank/
GeneCensus	bioinfo.mbb.yale.edu/genome/
MODBASE	guitar.rockefeller.edu/modbase/
PDB	www.rcsb.org/pdb/
PRESAGE	presage.berkeley.edu
SCOP	scop.mrc-lmb.cam.ac.uk/scop/
TrEMBL	srs.ebi.ac.uk
PROCURA POR MOLDES	
123D	123d.ncifcrf.gov/123D+.html
BLAST	www.ncbi.nlm.nih.gov/BLAST/
DALI	www2.ebi.ac.uk/dali/
FastA	www.ebi.ac.uk/fasta33/
MATCHMAKER	bioinformatics.burnham-inst.org
PHD, TOPITS	eubic.bioc.columbia.edu/predictprotein/
PROFIT	www.came.sbg.ac.at
THREADER	insulin.brunel.ac.uk/~jones/threader.html
FRSVR	fold.doe-mbi.ucla.edu
ALINHAMENTO DE SEQUENCIA	
BCM SERVER	searchlauncher.bcm.tmc.edu
BLAST2	www.ncbi.nlm.nih.gov/gor/bl2.html
BLOCK MAKER	blocks.fhcre.org/blocks/blockmkr/make_blocks.html
CLUSTAL	www2.ebi.ac.uk/clustalw/
FASTA3	www2.ebi.ac.uk/fasta3/
MULTALIN	pbl.ibcp.fr
MODELAGENS	
COMPOSER	www.tripos.com/software/composer.html
CONGEN	www.congenomics.com/congen/congen.html
ICM	www.molsoft.com
InsightII	www.accelrys.com
MODELLER	guitar.rockefeller.edu/modeller
QUANTA	www.accelrys.com
SYBYL	www.tripos.com
SCWRL	www.fccc.edu/research/labs/dunbrack/scwrl/
SWISS-MOD	www.expasy.org/swissmod/SWISS-MODEL.html
WHAT IF	www.cmbi.kun.nl/whatif/
VALIDAÇÃO DE MODELO	
ANOLEA	www.fundp.ac.be/sciences/biologie/bms/CGI/anolea.html
AQUA	urchin.bmrp.wisc.edu/~jorgen/aqua/
BIOTECH ^a	biotech.embl-heidelberg.de:8400
ERRAT	www.doe-mbi.ucla.edu/Services/ERRAT/
PROCHECK	www.biochem.ucl.ac.uk/~roman/procheck/procheck.html
ProsaII	www.came.sbg.ac.at
PROVE	www.ucmb.ulb.ac.be/UCMB/PROVE
SQUID	www.ysbl.york.ac.uk/~oldfield/squid/
VERIFY3D	www.doe-mbi.ucla.edu/Services/Verify_3D/
WHATCHECK	www.sander.embl-heidelberg.de/whatcheck/
VALIDAÇÃO DE METODOS	
CASP	predictioncenter.llnl.gov
CAFASP	cafasp.bioinfo.pl
EVA	eubic.bioc.columbia.edu/eva/
LiveBench	bioinfo.pl/LiveBench/

1.2 1.2. Aplicações (BAYAT, 2002)

O crescimento da bioinformática tem sido uma aventura global, com a criação de redes de computadores que permitem acesso fácil a dados biológicos possibilitando o desenvolvimento de programas para análises complexas. Múltiplos projetos internacionais direcionam-se para banco de dados de genes e de proteínas, que são utilizados livremente pela comunidade científica via internet.

Biologistas estruturais também usam bioinformática como forma de manusear os vastos e complexos dados da cristalografia por difração de raios X, ressonância magnética nuclear e investigação por microscopia eletrônica para a criação de modelos tridimensionais de moléculas.

A habilidade para a identificação de alvos específicos marca o uso facilitado das ferramentas de bioinformática para a descoberta de novos candidatos a drogas.

À curto prazo, como resultado das análises de bioinformática emergindo do projeto genoma humano, mais genes causadores de doenças poderão ser identificados e novos alvos de drogas poderão ser simultaneamente descobertos. A bioinformática poderá servir para identificar genes e eliminar as vias patogênicas, portanto, providenciará uma oportunidade para o desenvolvimento de alvos terapêuticos.

A longo prazo, a bioinformática integrará análises genômicas, patológicas e dados clínicos, revelando diversas reações potenciais de drogas em indivíduos, pelo simples uso de testes genéticos.

A bioinformática é um ingrediente fundamental para o sucesso do projeto genoma humano. Este projeto visa o uso de alguns métodos e ferramentas da bioinformática, como o uso do programa MODELLER para a Modelagem Molecular por Homologia, do banco de dados, como o PDB (Protein Data Bank) e o ncbi para a identificação de moldes, programas para alinhamento de seqüências primárias como o CLUSTAL e o



MultAlin; e para validação de modelos tridimensionais como: Procheck e Verify-3D, citados na Tabela 1.

1.3.1. Ciclo Celular

Todos os organismos vivos, das bactérias unicelulares aos membros multicelulares, são produtos de repetidos ciclos de crescimento e divisão celular, desde o início da vida há mais de três bilhões de anos (ALBERTS et al., 1999).

Uma célula se reproduz por uma sequência ordenada de eventos que duplicam seus componentes e depois a dividem em duas. Este ciclo de duplicação e divisão é conhecido como ciclo celular (ALBERTS et al., 1999).

A divisão celular consiste de dois processos consecutivos, caracterizados principalmente pela replicação do DNA e segregação dos cromossomos replicados em duas células separadas. Originalmente, a divisão celular está dividida em dois estágios: Mitose (M). Isto é, a divisão celular propriamente dita; e interfase, que trata do período entre uma fase M e a próxima (VERMEULEN et al., 2003). Entretanto, a interfase é um período de intensa atividade para a célula e é dividida nas três fases remanescentes do ciclo celular. Durante a fase S (S= síntese), a célula replica o DNA nuclear, um pré-requisito essencial para a divisão celular. A fase G é ladeada por duas fases onde a célula continua a crescer. A fase G₁ (G = "gap", lacuna, intervalo) é o intervalo entre o término da fase M e o início da fase S. A fase G₂ é o intervalo entre o término da fase S e o início da fase M. Durante toda a interfase, a célula continua a transcrição dos genes, sintetiza proteínas e cresce em massa, (figura 1) (ALBERTS et al., 1999).

Os primeiros sinais visíveis de que uma célula está entrando na fase M é a progressiva condensação de seus cromossomos, previamente replicados durante a fase S. A condensação dos cromossomos marca o final da fase G₂ (ALBERTS et al., 1999).

1.3. DIVISÃO CELULAR

1.3.1. Ciclo Celular

Todos os organismos vivos, das bactérias unicelulares aos mamíferos multicelulares, são produtos de repetidos ciclos de crescimento e divisão celular, desde o início da vida há mais de três bilhões de anos (ALBERTS et al., 1999).

Uma célula se reproduz por uma sequência ordenada de eventos que duplicam seus componentes e depois a dividem em duas. Este ciclo de duplicação e divisão é conhecido como **ciclo celular** (ALBERTS et al., 1999).

A divisão celular consiste de dois processos consecutivos, caracterizados principalmente pela replicação do DNA e segregação dos cromossomos replicados em duas células separadas. Originalmente, a divisão celular está dividida em dois estágios: Mitose (M), isto é, a divisão celular propriamente dita; e interfase, que trata do período entre uma fase M e a próxima (VERMEULEN et al., 2003). Entretanto, a interfase é um período de intensa atividade para a célula e é dividida nas três fases remanescentes do ciclo celular. Durante a **fase S** (S= síntese), a célula replica o DNA nuclear, um pré-requisito essencial para a divisão celular. A fase S é ladeada por duas fases onde a célula continua a crescer. A **fase G₁** (G = "gap", lacuna, intervalo) é o intervalo entre o término da fase M e o início da fase S. A **fase G₂** é o intervalo entre o término da fase S e o início da fase M. Durante toda a interfase, a célula continua a transcrição dos genes, sintetiza proteínas e cresce em massa, (figura1) (ALBERTS et al., 1999).

Os primeiros sinais visíveis de que uma célula está entrando na fase M é a progressiva condensação de seus cromossomos, previamente replicados durante a fase S. A condensação dos cromossomos marca o final da fase G₂ (ALBERTS et al., 1999).

Figura2: Comparação entre mitose e meiose (ALBERTS et al., 1999).

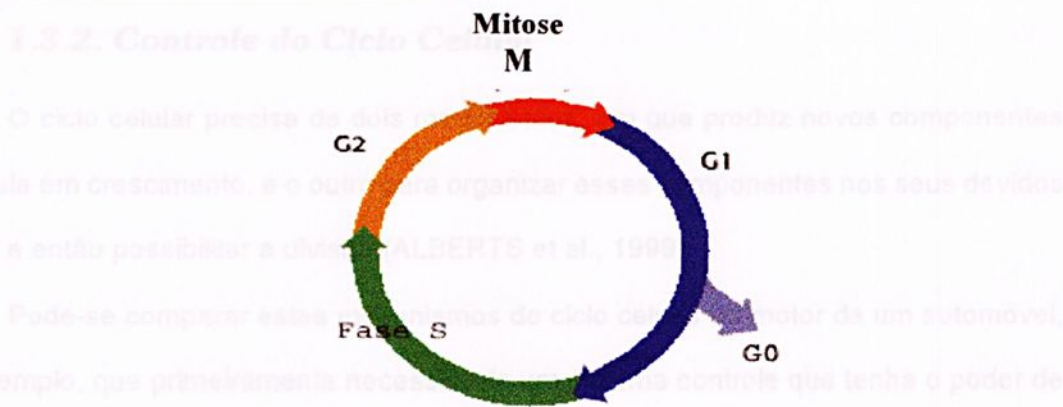


Figura1: Fases do ciclo celular (www.biologia.arizona.edu/.../cellcycle.gif)

Apesar da maioria das características mecânicas da meiose ser semelhante à da mitose, o comportamento dos cromossomos é diferente; a meiose produz quatro células haplóides geneticamente não-similares por duas divisões celulares consecutivas, enquanto que a mitose produz duas células diplóides geneticamente idênticas por uma única divisão celular (figura2) (ALBERTS et al., 1999).

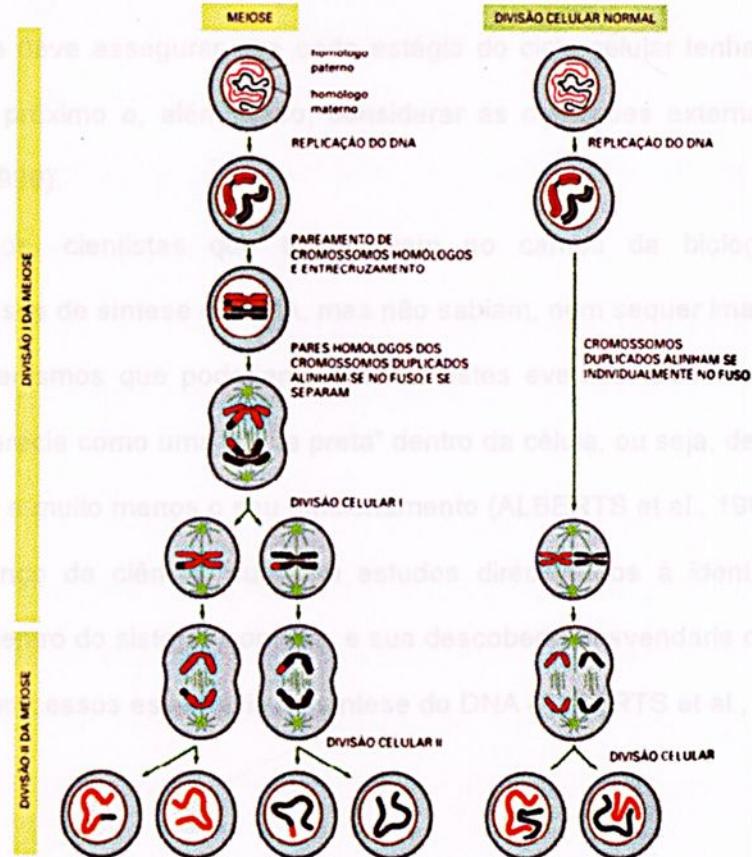


Figura2: Comparação entre meiose e mitose (ALBERTS et al., 1999).

1.3.2. Controle do Ciclo Celular

O ciclo celular precisa de dois mecanismos, um que produz novos componentes da célula em crescimento, e o outro para organizar esses componentes nos seus devidos locais, e então possibilitar a divisão (ALBERTS et al., 1999).

Pode-se comparar estes mecanismos do ciclo celular ao motor de um automóvel, por exemplo, que primeiramente necessita de um sistema controle que tenha o poder de ligar e desligar. Para que este motor trabalhe é preciso que todas as sub partes, ou seja, as peças, estejam interligadas uma a uma e suas funções ativadas de maneira ordenada e bem organizada, sendo elas internas ou externas ao motor. Como pode ser observado, em todo sistema há a necessidade da existência de uma ordem para que as etapas, ou os processos se realizem.

O sistema controle do ciclo celular deve ativar enzimas e outras proteínas responsáveis por um processo no tempo correto e desativá-las logo após o término de tal processo. Também deve assegurar que cada estágio do ciclo celular tenha terminado antes de iniciar o próximo e, além disso, considerar as condições externas à célula (ALBERTS et al., 1999).

Durante anos, cientistas que trabalhavam no campo da biologia celular, observaram processos de síntese de DNA, mas não sabiam, nem sequer imaginavam, a existência de mecanismos que poderiam controlar estes eventos. Sendo assim, este sistema controle parecia como uma "caixa preta" dentro da célula, ou seja, desconhecia-se a sua existência e muito menos o seu funcionamento (ALBERTS et al., 1999).

Com o avanço da ciência, surgiram estudos direcionados à identificação de proteínas chaves dentro do sistema controle, e sua descoberta desvendaria os mistérios que percorrem os processos essenciais na síntese do DNA (ALBERTS et al., 1999).



O sistema controle do ciclo celular consegue controlar tudo por meio de “freios moleculares” que podem parar o ciclo em vários pontos de checagem. Sendo assim, o sistema controle não desencadeia o próximo passo do ciclo antes do término ou estágio precedente (ALBERTS et al., 1999).

A qualidade no controle do ciclo celular é devido aos pontos de restrição (R) e aos pontos de checagem ou controles adicionais que asseguram uma seqüência ordenada de eventos no ciclo (VERMEULEN et al., 2003). Os pontos de checagem asseguram a replicação e a separação cromossômica fiel, por isso a estabilidade genética é conservada. O fracasso destes pontos de checagem ao deter a célula depois do estímulo apropriado é uma das características de câncer. O Ponto de checagem, e o ponto de restrição ocorrem no meio da fase G₁, depois deste ponto, células tornam-se independentes dos fatores de crescimento que comprometem a divisão celular.

Os estudos sobre o sistema controle do ciclo celular têm sido direcionados ao desenvolvimento de um número de compostos com o potencial de restaurar o controle da divisão celular em cânceres humanos (SWANTON, 2004).

1.3.3. Os gerentes do sistema controle são formados por quinases dependentes de ciclinas

O sistema controle governa a maquinaria do ciclo celular pela fosforilação das proteínas chaves que iniciam e regulam a replicação, a mitose e a citocinese (ALBERTS et al., 1999).

A transição entre as fases M, G₁, S, e G₂, é controlada por duas famílias de proteínas que são chaves para o seu funcionamento (BARVIAN et al., 2000). Uma delas é a família das quinases dependentes de ciclinas (CDKs) (DE AZEVEDO et al., 1996), que regulam a progressão do ciclo celular principalmente em eucariotos. As CDKs são subunidades catalíticas de complexos heterodiméricos (GOULD, 2000) que recebem este



nome devido ao fato de não possuir atividade, a menos que estejam associadas a um membro da família das ciclinas (DE AZEVEDO et al., 1996; GOULD, 2000), denominadas proteínas ativadoras (Figuras 3 e 4).

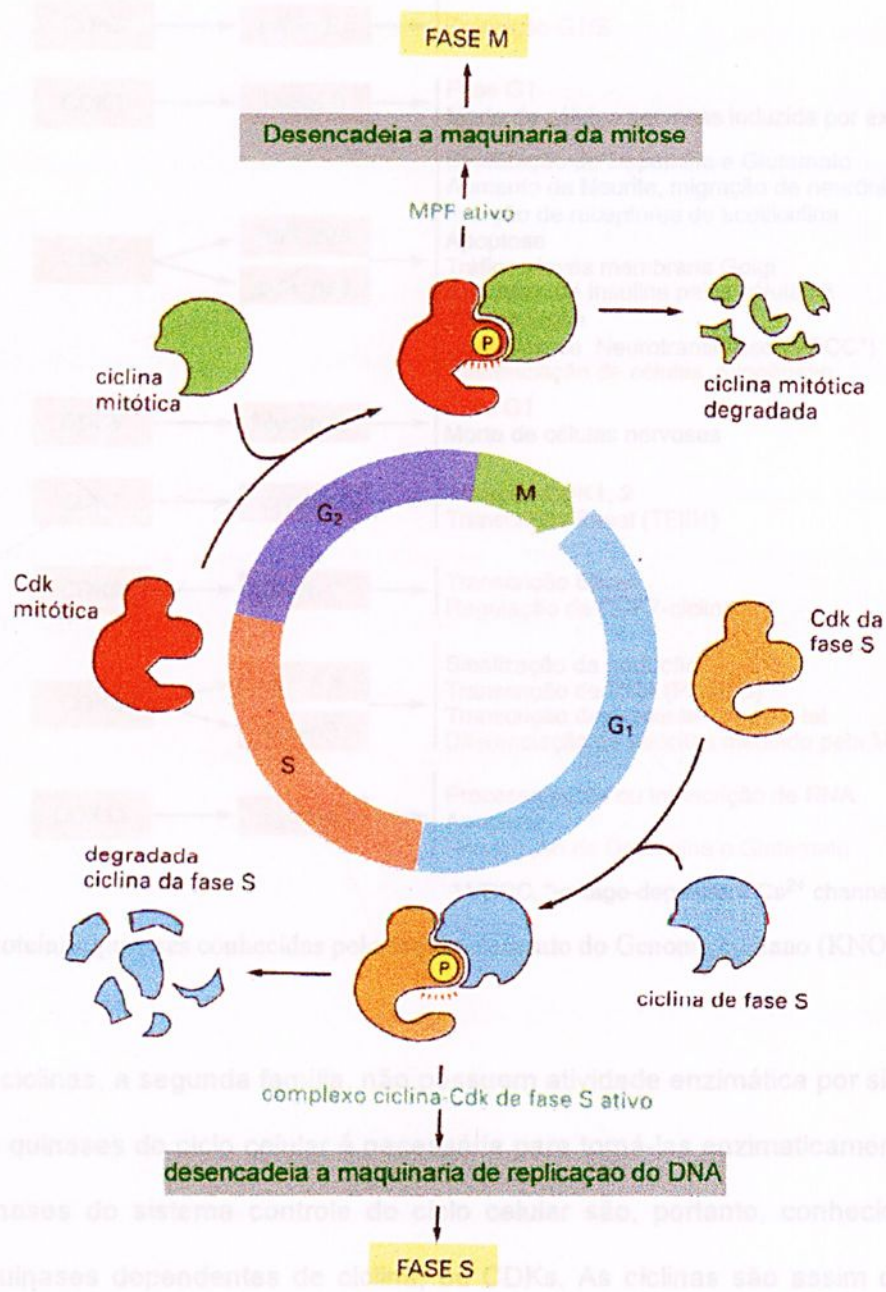


Figura3: Regulação das CDKs através da ciclina (ALBERTS et al., 1999).

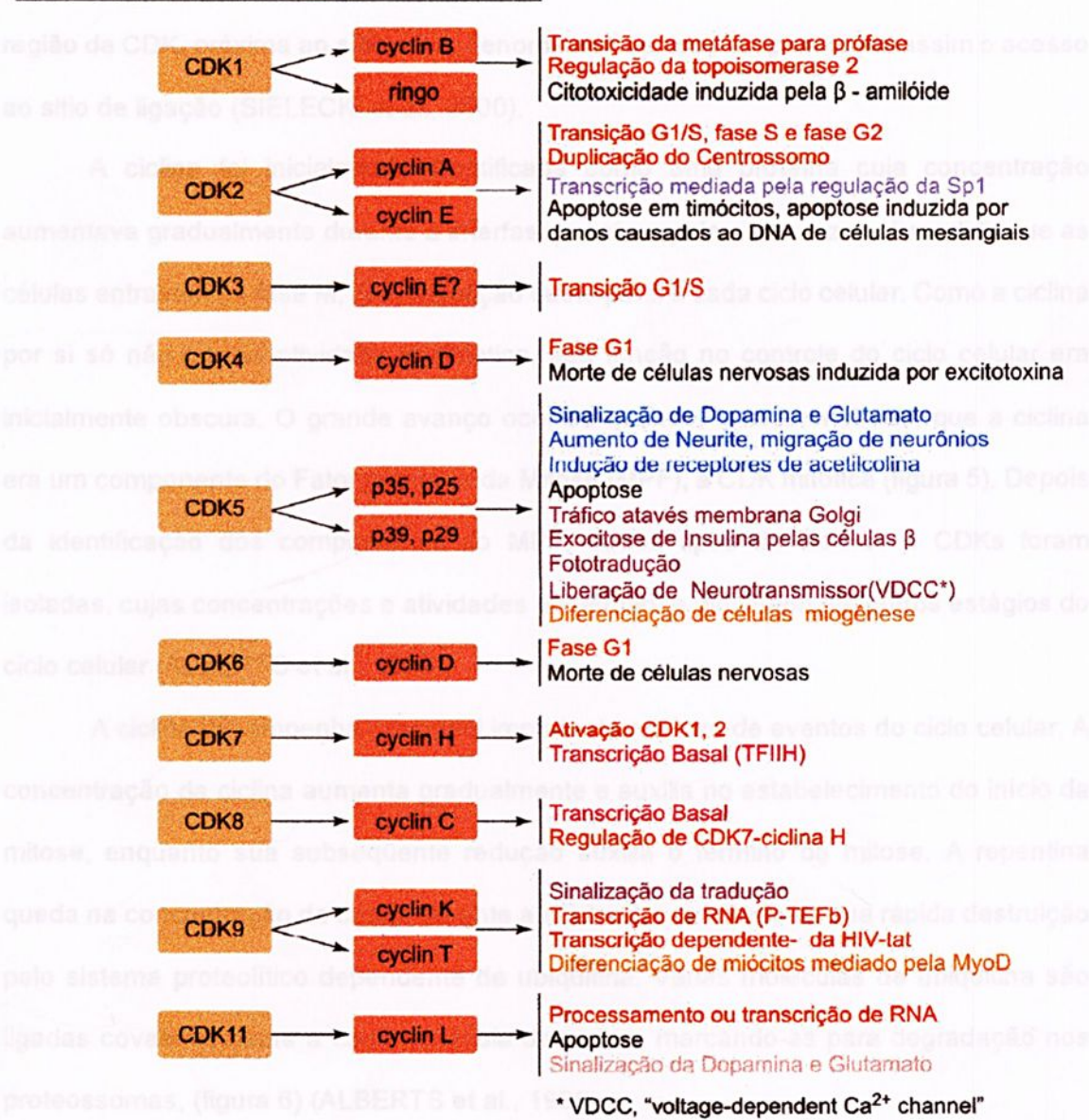


Figura 4: Proteínas quinases conhecidas pelo seqüenciamento do Genoma humano (KNOCKAERT et al., 2002).

As ciclinas, a segunda família, não possuem atividade enzimática por si, mas sua ligação às quinases do ciclo celular é necessária para torná-las enzimaticamente ativas. Estas quinases do sistema controle do ciclo celular são, portanto, conhecidas como proteíno-quinases dependentes de ciclina, ou CDKs. As ciclinas são assim chamadas porque, ao contrário das CDKs, sua concentração varia de maneira cíclica durante o ciclo celular (ALBERTS et al., 1999). Esta associação resulta na mudança estrutural de uma

região da CDK, próxima ao sítio ativo, denominada de "loop" T, permitindo assim o acesso ao sítio de ligação (SIELECKI et al., 2000).

A ciclina foi inicialmente identificada como uma proteína cuja concentração aumentava gradualmente durante a interfase e caía rapidamente a zero à medida que as células entravam na fase M, com repetição deste perfil a cada ciclo celular. Como a ciclina por si só não possui atividade enzimática, sua função no controle do ciclo celular era inicialmente obscura. O grande avanço ocorreu quando, foi demonstrado, que a ciclina era um componente do Fator promotor da Mitose (MPF), a CDK mitótica (figura 5). Depois da identificação dos componentes do MPF, outros tipos de ciclinas e CDKs foram isoladas, cujas concentrações e atividades aumentam e diminuem em outros estágios do ciclo celular (ALBERTS et al., 1999).

A ciclina desempenha um papel importante no fluxo de eventos do ciclo celular. A concentração da ciclina aumenta gradualmente e auxilia no estabelecimento do início da mitose, enquanto sua subsequente redução auxilia o término da mitose. A repentina queda na concentração da ciclina durante a mitose é o resultado da sua rápida destruição pelo sistema proteolítico dependente de ubiquitina. Várias moléculas de ubiquitina são ligadas covalentemente a cada molécula de ciclina, marcando-as para degradação nos proteossomas, (figura 6) (ALBERTS et al., 1999).

Um complexo CDK- ciclina atua na fase G_2 e promove a passagem para a fase M, enquanto as ciclinas *de fase S*, ligam-se às moléculas de CDK ao final da fase G_1 para promover a progressão para a fase S. Outras ainda, chamadas ciclinas *de fase G_1* , atuam no início de G_1 ; ligam-se à molécula de CDK e auxiliam na formação e ativação dos complexos CDK- ciclina de fase S, promovendo a passagem para a fase S (ALBERTS et al., 1999).

Figura 6: Ciclinas são degradadas no proteossoma através de moléculas de ubiquitina (www.unesp.br/img/figuras/1461/737202.gif)

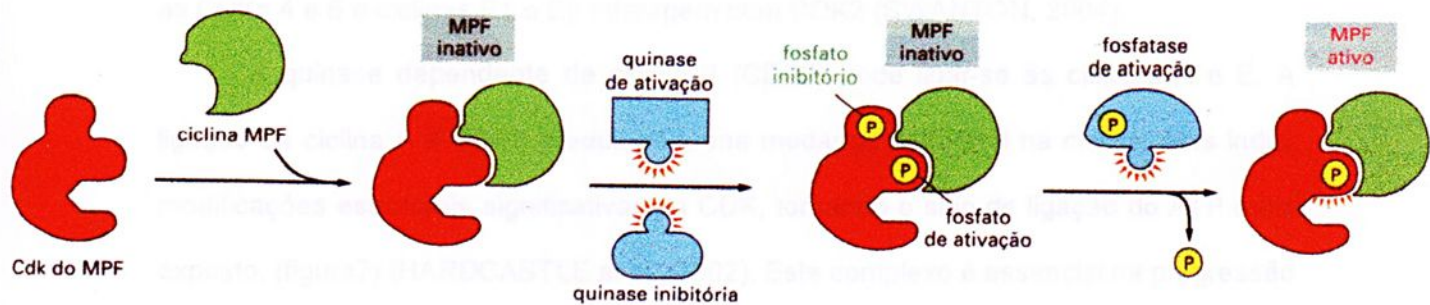


Figura5: Ativação do fator promotor de mitose (ALBERTS et al., 1999).

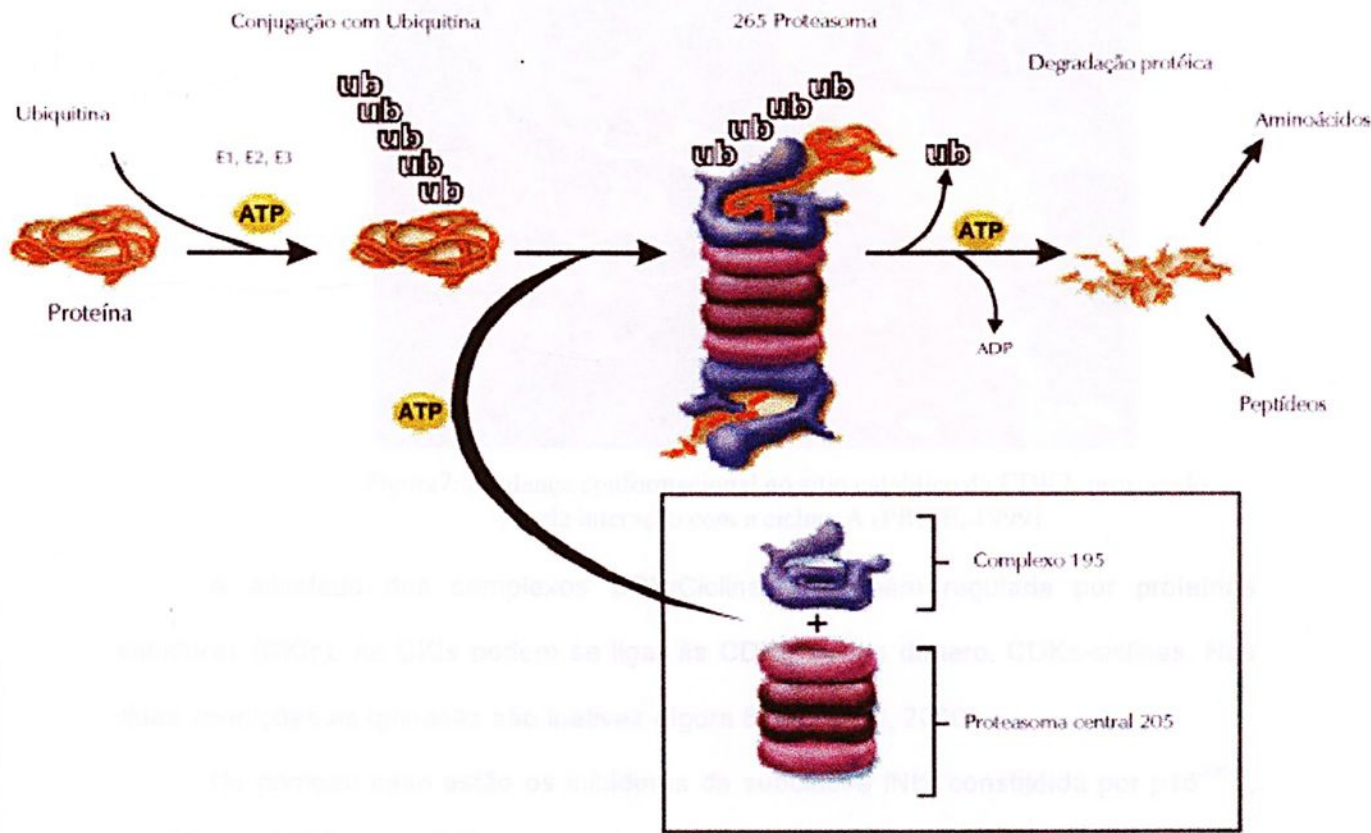


Figura6: Ciclinas são degradadas no proteossoma através de moléculas de ubiquitina (www.scielo.br/img/fbpe/m/v14n1/7572f2.gif).

A passagem do ponto de restrição é coordenado por duas famílias de ciclinas, a família da ciclina D (D1, D2 e D3) e a família da ciclina E (E1 e E2). As ciclinas D ligam-se as CDKs 4 e 6 e ciclinas E1 e E2 interagem com CDK2 (SWANTON, 2004).

A quinase dependente de ciclina 2 (CDK2) pode ligar-se às ciclinas A e E. A ligação da ciclina A à CDK2 produz pequena mudança estrutural na ciclina, mas induz modificações estruturais significativas na CDK, tornando o sítio de ligação do ATP mais exposto, (figura7) (HARDCASTLE et al., 2002). Este complexo é essencial na progressão do ciclo celular, embora estudos realizados por ALEEM et al., (2004) e BERTHET et al., (2003) com modelos de ratos com CDK2 nocauteadas revelaram que, nem CDK2, nem ciclina A são essenciais no ciclo celular. Estudos adicionais poderão confirmar esta proposta.

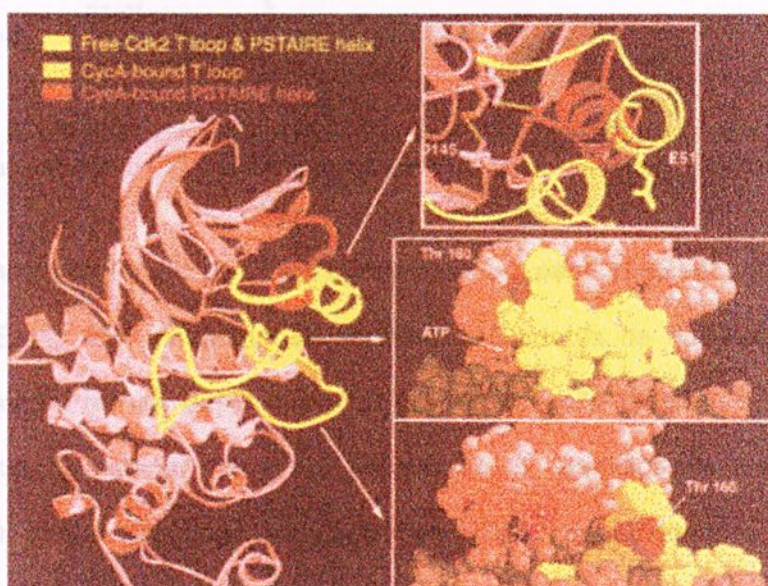


Figura7: Mudança conformacional no sítio catalítico da CDK2, provocado pela interação com a ciclina A (PRIZE, 1999).

A atividade dos complexos CDK/Ciclina é também regulada por proteínas inibidoras (CKIs). As CKIs podem se ligar às CDKs, ou ao dímero, CDKs-ciclinas. Nas duas condições as quinases são inativas (figura 8) (GOULD, 2000).

No primeiro caso estão os inibidores da subclasse INK, constituída por $p16^{INK4a}$, $p15^{INK4b}$, $p18^{INK4c}$ e $p19^{INK4d}$, que se ligam a CDK4/6 impedindo a formação de complexos

com as ciclinas. No segundo caso estão os inibidores da subclasse KIP, constituída por $p21^{Waf1/Cip1}$, $p27^{Kip1}$, e $p57^{Kip2}$, que se ligam aos complexos Ciclina/CDK já formados impedindo sua atividade quinásica (figura 9) (HALL et al., 1995; MORGAN, 1995).

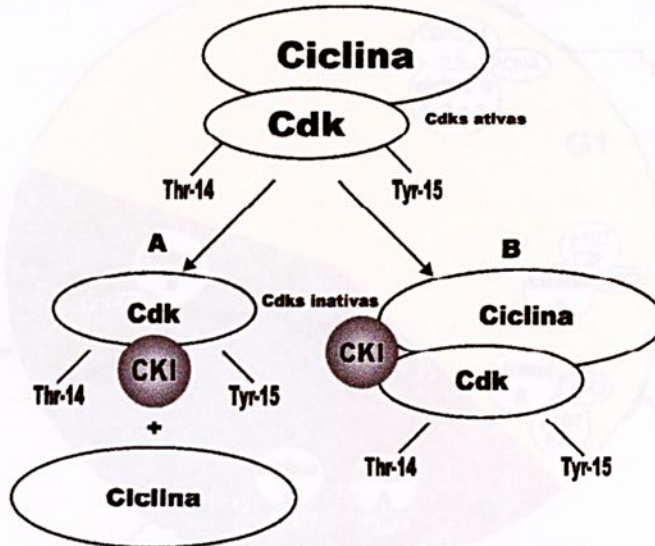


Figura 8: Processo de inibição pelas CKIs A) CKI liga-se a CDK impedindo a associação à ciclina; B) CKI liga-se ao complexo CDK-ciclina impedindo a atividade quinásica (GOULD, 2000).

Alguns membros da família de CDKs apresentam um papel regulador da atividade de outras CDKs. Estas são denominadas proteínas ativadoras de quinases (CAKs); algumas delas parecem ter um papel duplo no que diz respeito também à função de fosforilar o domínio C-terminal da RNA polimerase II, participando do controle transcricional. Uma destas proteínas ativadoras de CDKs conhecidas é o complexo CDK7/Ciclina H, (Figura 9) (SIELECKI et al., 2000).

As CDKs e as ciclinas são alvos para intervenções em diferentes tipos de câncer humano, estimulando a procura por inibidores químicos de CDKs. Diferentes estratégias para intervenções terapêuticas podem atuar nos moduladores de CDKs: quando o alvo são as proteínas reguladoras da atividade das CDKs (estratégia indireta), ou quando o alvo é a inibição da atividade catalítica das CDKs (estratégia direta) (VERMEULEN et al., 2003). Danos ocorridos na p53, por exemplo, podem levar ao câncer.

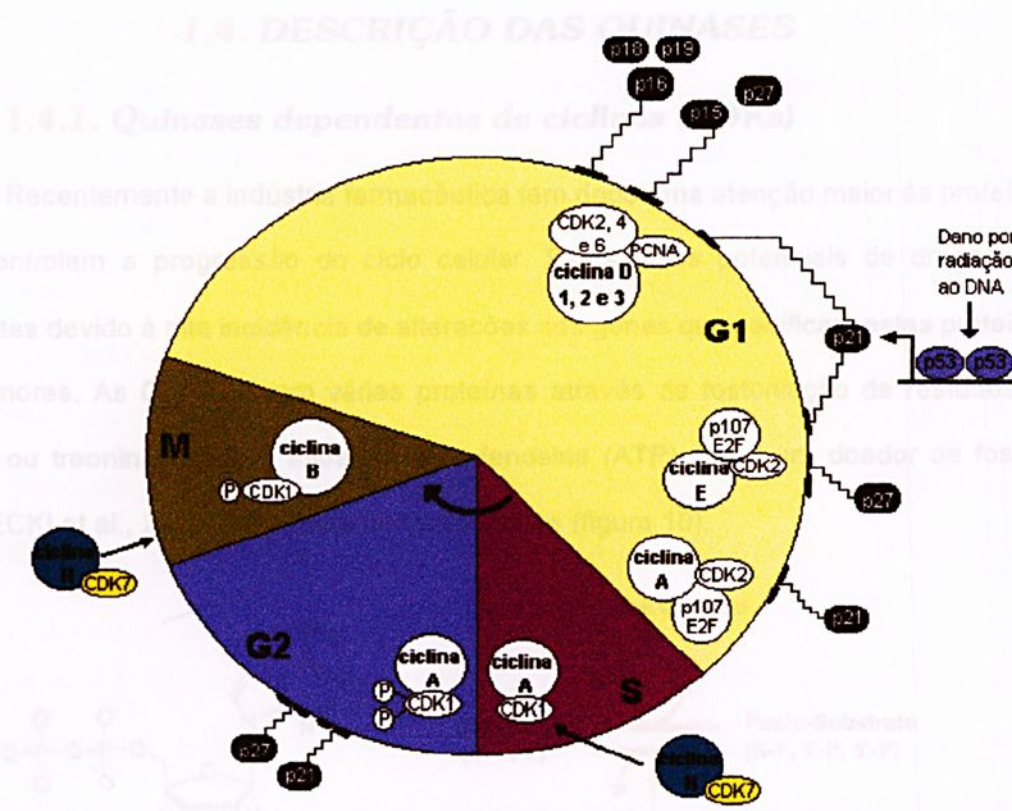


Figura 9: Representação esquemática do ciclo celular mostrando a atuação das CDKs nas respectivas fases do ciclo celular

1.4. DESCRIÇÃO DAS QUINASES

1.4.1. Quinases dependentes de ciclinas (CDKs)

Recentemente a indústria farmacêutica tem dado uma atenção maior às proteínas que controlam a progressão do ciclo celular. Estes alvos potenciais de drogas são evidentes devido à alta incidência de alterações nos genes que codificam estas proteínas em tumores. As CDKs ativam várias proteínas através da fosforilação de resíduos de serina ou treonina, usando trifosfato de adenosina (ATP) como um doador de fosfato (SIELECKI et al., 2000), são portanto transferases (figura 10).

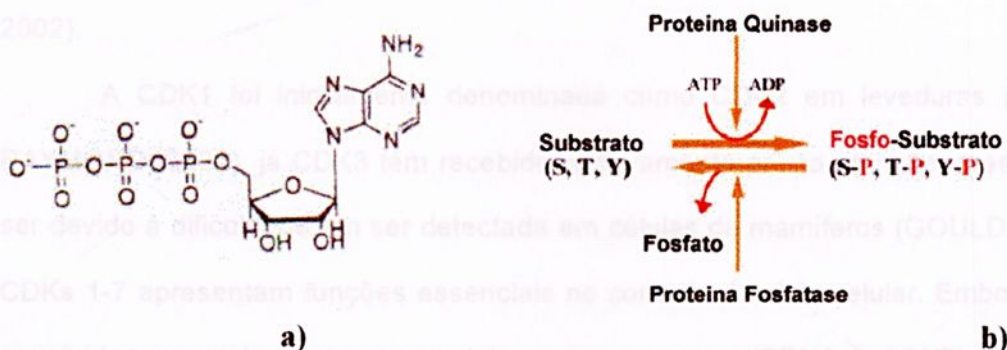


Figura 10: a) Fórmula estrutural da Trifosfato de Adenosina (ATP) e b) Mecanismo da fosforilação e desfosforilação (SIELECKI et al., 2000; MEIJER e RAYMOND, 2002).

As CDKs foram pela primeira vez identificadas em células de leveduras em crescimento (*Saccharomyces cerevisiae* e *Schizosaccharomyces pombe*) (GOULD, 2000). Leveduras são fungos unicelulares e eucariotos relativamente simples que se dividem quase tão rapidamente como as bactérias. Um grande número de mutantes de leveduras foram isolados a partir de diferentes fases do ciclo celular, e, utilizando estes mutantes, foi possível identificar vários genes responsáveis pelo controle do ciclo celular (GOULD, 2000).

Muitos genes que controlam o ciclo celular sofreram pouca alteração durante a evolução, tanto que o gene humano funciona perfeitamente na célula de levedura. Por

exemplo, uma célula de levedura cuja cópia de um gene de CDK ou ciclina particular seja defeituoso, não conseguirá dividir-se, mas a divisão ocorrerá normalmente se uma cópia do gene humano correspondente for artificialmente introduzido na célula defeituosa. Apesar dos bilhões de anos de evolução divergente, todas as células eucarióticas (leveduras, animais ou vegetais) utilizam, essencialmente, as mesmas moléculas para controlar os eventos do seu ciclo celular (ALBERTS et al., 1999).

No Genoma humano aparecem codificadas cerca de 13 CDKs e 25 ciclinas, muitas das quais ainda não foram caracterizadas. Logo, espera-se uma ampla variedade de aplicações e, portanto, inibidores mais seletivos de CDKs (MEIJER e RAYMOND, 2002).

A CDK1 foi inicialmente denominada como CDC2 em leveduras (MEIJER e RAYMOND, 2002), já CDK3 tem recebido relativamente pouca atenção, mas isso deve ser devido à dificuldade em ser detectada em células de mamíferos (GOULD, 2000). As CDKs 1-7 apresentam funções essenciais no controle do ciclo celular. Embora estejam envolvidas em outros processos, tais como apoptose (CDK1 e CDK5), em funções neuronais (CDK5 e CDK11) e como reguladoras da transcrição (CDK7, CDK8 e CDK9) figura 5 (MEIJER e RAYMOND, 2002).

Análises das seqüências primárias de proteínas juntamente com estudos de cristalografia têm revelado que tais quinases possuem estruturas similares (HARDCASTLE et al., 2002). As mesmas características vêm sendo observadas em 12 regiões de subdomínios, que são importantes estrutural e funcionalmente. Esta arquitetura comum, particularmente aparece no domínio de ligação do ATP, as quais possuem três resíduos (IRE) altamente conservados na região catalítica (HARDCASTLE et al., 2002). Portanto, sua aplicação potencial tem sido revisada incluindo câncer, doenças neuro-degenerativas, doenças cardiovasculares (restenoses), infecções virais (HIV, Herpes, citomegalovirus, papilomavirus), vários parasitas unicelulares dos gêneros

(*Plasmodium*, *Trypanossoma*, *Leishmania*), e também são usadas em reprodução *in vitro* e para técnicas de clonagem (MEIJER e RAYMOND, 2002).

1.4.1.1. Quinase dependente de ciclina 2 (CDK2)

A base estrutural de ativação da CDK2 está bem compreendida. A estrutura monomérica inativa da CDK2, foi determinada em 1993 por DE BONDT et al (figura 11).

Extensivos estudos por cristalografia de difração de raios X foram realizados com a CDK2. Os dados ofereceram uma explicação estrutural detalhada para a ativação da CDK2 por ciclinas e CAKs que pode ser usada para todas as quinases dependentes de ciclinas (PRIZE, 1999).

No processo de ativação, a ciclina interage com ambos os lóbulos da CDK, formando uma interface proteína-proteína. Os contatos da ciclina A com a hélice C (PSTAIRE) tem importância central tornando o sítio catalítico da CDK2 exposto para ligar-se com seu substrato (PRIZE, 1999). Estas mudanças estruturais indicam que as CDKs possuem uma flexibilidade estrutural intrínseca, especialmente na e ao redor da chave catalítica (RUSSO et al., 1996; PRIZE, 1999).

Na associação da CDK2 com a ciclina A, a hélice C rotaciona e muda de posição para realizar um maior contato com o lóbulo N_terminal da CDK2. O segmento de ativação é pouco enovelado e os dois lóbulos mudam de posição para uma orientação mais fechada. CDK2 é flexível, mas a ciclina A é rígida. O componente final do processo de ativação da CDK2 é a combinação da fosforilação da Thr160 e associação com a ciclina A (figura 11). (JOHNSON et al., 2002). A Thr160 fosforilada age como uma organizadora central para realizar interações com os resíduos de Arg50 na hélice C, e Arg126, que é adjacente ao Asp127 (HARDCASTLE et al., 2002).

Figura 12: a) Complexo CDK2-ATP mostrando o sítio de ligação do ATP e b) Detalhes das ligações de hidrogênio entre ATP e CDK2 (DE BONDT et al., 1993; JOHNSON et al., 2002)

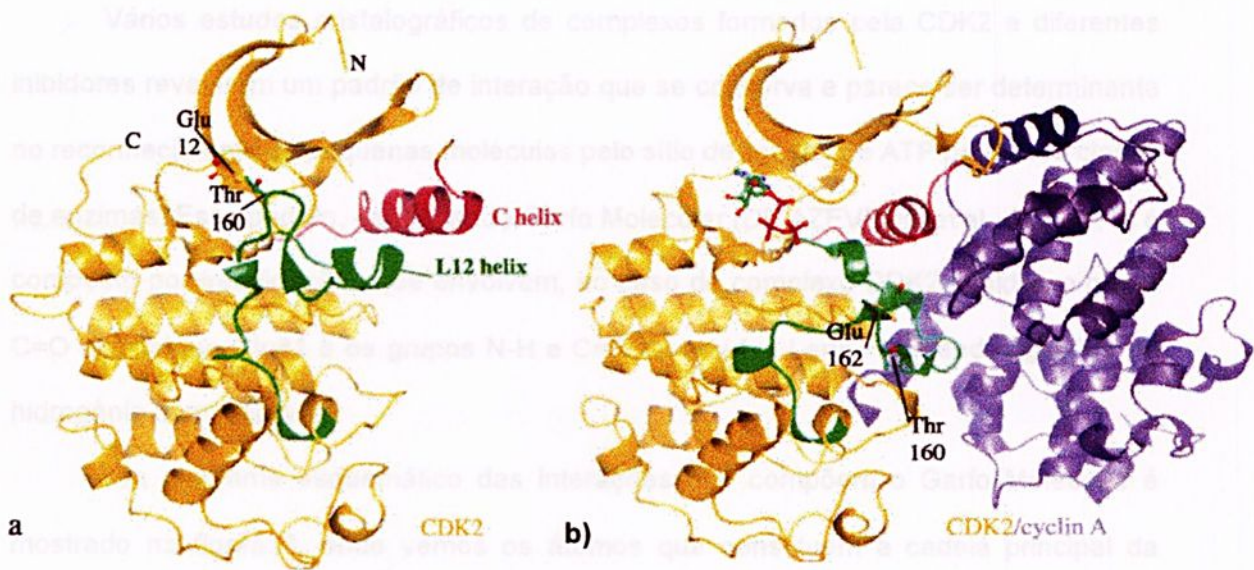


Figura 11: Comparação das estruturas cristalográficas da CDK2; a) Estrutura monomérica da CDK2; b) Estrutura dimérica da CDK2/ciclinaA, mostrando a mudança conformacional do sítio de ligação do ATP (DE BONDY et al., 1993; JOHNSON et al., 2002).

A estrutura monomérica da CDK2 com ATP revela importantes interações proteína-ligante. O ATP faz ligações de hidrogênio com a cadeia principal, principalmente com os resíduos Glu81 e Leu83. Além das interações por ligações de hidrogênio, contatos hidrofóbicos e de Van der Waals serem formados entre a enzima e o anel de adenina do ATP (figuras 12 a e b) (HARDCASTLE et al., 2002).

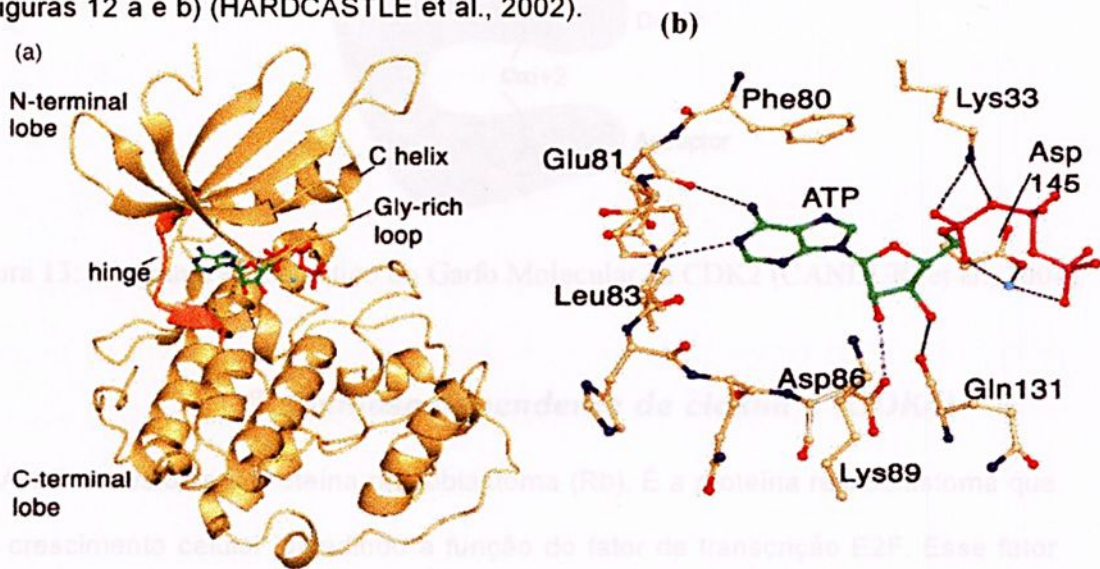


Figura 12: a) Complexo CDK2:ATP mostrando o sítio de ligação do ATP e b) Detalhes das ligações de hidrogênio entre ATP e CDK2 (DE BONDY et al., 1993; JOHNSON et al., 2002).

Vários estudos cristalográficos de complexos formados pela CDK2 e diferentes inibidores revelaram um padrão de interação que se conserva e parece ser determinante no reconhecimento de pequenas moléculas pelo sítio de ligação de ATP para essa classe de enzimas. Esse padrão, denominado Garfo Molecular (DE AZEVEDO et al., 2002 a, b) é composto por três ligações que envolvem, no caso do complexo CDK2-inibidor, o grupo C=O do resíduo Glu81 e os grupos N-H e C=O do resíduo Leu83 formando ligações de hidrogênio com o ligante.

Um diagrama esquemático das interações que compõem o Garfo Molecular é mostrado na figura 13, onde vemos os átomos que constituem a cadeia principal da proteína, na região do Garfo Molecular e os grupos doadores e aceptores de ligações de hidrogênio.

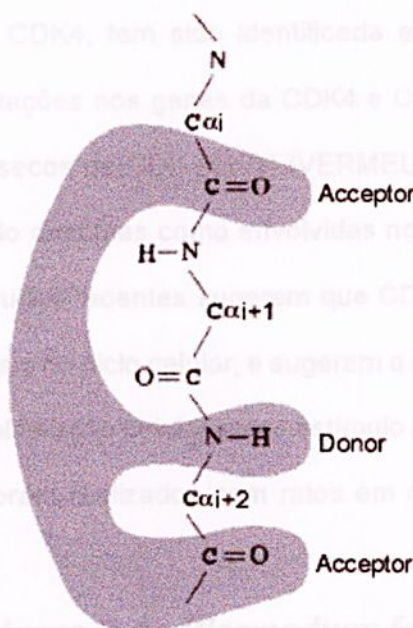


Figura 13: Diagrama esquemático do Garfo Molecular da CDK2 (CANDURI et al., 2004)

1.4.1.2. Quinase dependente de ciclina 4 (CDK4)

A CDK4 fosforila a proteína retinoblastoma (Rb). É a proteína retinoblastoma que inibe o crescimento celular impedindo a função do fator de transcrição E2F. Esse fator

ativa genes que promovem a proliferação celular. A fosforilação da Rb pela CDK4 reverte a capacidade da Rb de reprimir o crescimento celular. A Rb é conhecida como um supressor tumoral porque a sua perda pode levar ao desenvolvimento de retinoblastomas (cânceres da retina) e de outros tumores (VOET et al., 2002).

Os fatores de transcrição Rb e E2F são chaves regulatórias das fases G₁ e S do ciclo celular. Está bem estabelecido que a função da proteína Rb é regulada pela fosforilação, sendo que, as quinases dependentes de ciclina D e ciclina E estão envolvidas nesta fosforilação, sendo elas, CDKs 4, 6, 1 e 2 (XIN, et al., 2002).

A CDK1, 4 e 6 desempenham funções essenciais na morte celular de neurônios, e a inibição destas quinases resulta em neuroproteção (KNOCKAERT, et al., 2002).

A super expressão da CDK4, tem sido identificada em aglomerados celulares como melanoma e glioma. Mutações nos genes da CDK4 e CDK6 resultam numa baixa interação com inibidores intrínsecos de CDK (CKIs) (VERMEULEN et al., 2003). Apesar da CDK4 e da CDK6 terem sido descritas como envolvidas no controle e progressão do ciclo celular de eucariotos, estudos recentes sugerem que CDK4 e CDK6 em complexo com ciclina D, não são essenciais no ciclo celular, e sugerem a existência de mecanismos alternativos para o início da proliferação celular sobre estímulo mitogênico (MALUMBRES et al., 2004). Estes estudos foram realizados com ratos em que estas quinases foram nocauteadas.

1.4.2. Proteína Quinase 6 do *Plasmodium falciparum* (PfPK6)

Algumas doenças infecciosas são causadas por protistas, tais como parasitas causadores da malária e tripanossomatídeos. As proteínas quinases representam alvos potenciais e atrativos para o desenho de drogas contra tais agentes infecciosos (WARD et al., 2004).

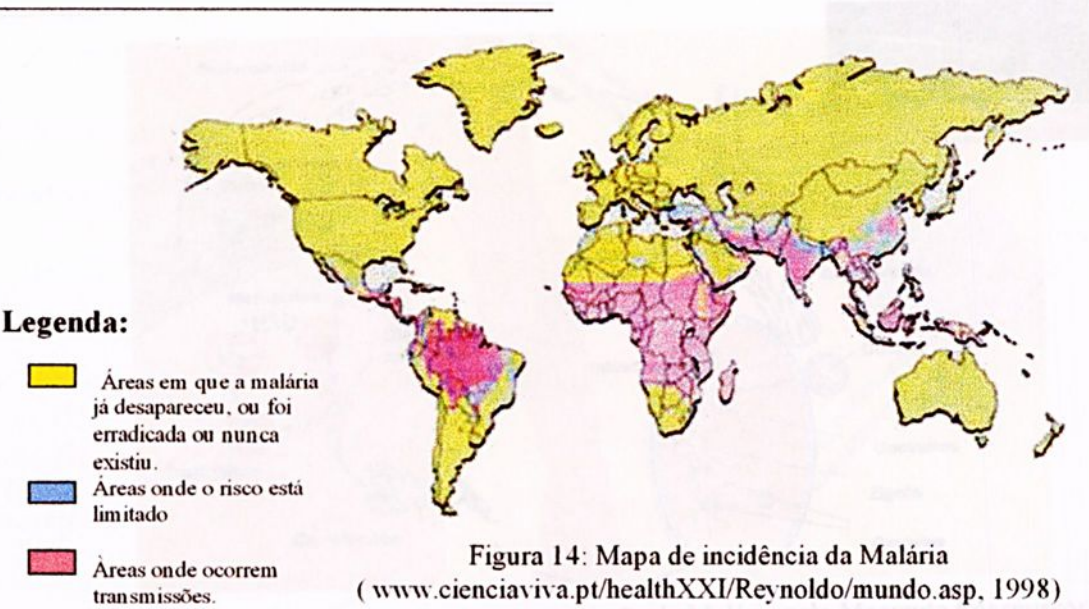


O genoma do *Plasmodium* contém vários genes que codificam proteínas relacionadas as CDKs, a proteína quinase 5 do *Plasmodium falciparum* (PfPK5) foi a primeira identificada. É uma quinase homóloga a CDK1 (LE ROCH et al., 2000). Recentemente foi identificada a proteína quinase 6 (PfPK6), uma enzima altamente homóloga as CDKs (DOOLITTLE, 2002).

Plasmodium falciparum é o agente causador de diversas formas de malária humana, tem um crescimento distinto e um processo de desenvolvimento diferente de muitas espécies de eucariotos. O parasita sofre muitos ciclos de replicação do DNA e divisões nucleares na ausência de citocineses que resultam numa simples célula contendo de 8 a 32 núcleos. Quanto ao mecanismo regulatório do ciclo celular do *Plasmodium* é altamente desconhecido. É evidente que as quinases representam um importante papel no desenvolvimento do parasita. Várias quinases plasmodiais tem sido caracterizadas e mostram possuir mecanismos regulatórios, tais como ligação às ciclinas e fosforilação (WOODARD et al., 2003).

O *Plasmodium* continua sendo um dos maiores causadores de doenças humanas em áreas tropicais e subtropicais. Na África, morrem mais de um milhão de crianças menores de cinco anos a cada ano (BOWNMAN et al., 1999). Aproximadamente 40% da população mundial vivem em áreas onde a malária é transmitida. A estimativa é em torno de 300 à 500 milhões de casos e mais de 2,7 milhões de mortes por malária a cada ano (figura 14) (GARDNER et al., 2002).





O parasita da malária tem um ciclo de vida complexo. A infecção inicia-se pela picada de um mosquito infectado chamado *Anopheles gambiae*, que deposita os esporozóitos dentro da corrente sangüínea. Células sanguíneas estabelecem uma infecção no interior dos hepatócitos, onde eles realizam uma intensa multiplicação gerando milhares de merozóitos, um processo chamado de esquizogonia exo-eritrocítica. Os merozóitos invadem os eritrócitos, onde estabelecem a esquizogonia, processo responsável pela patogênese da malária. Alguns merozóitos, contudo, detém o ciclo celular e diferenciam-se dentro dos gametócitos do macho ou da fêmea, que são infecciosos para o mosquito. Uma vez ingerido pelo inseto, o desenvolvimento dos gametócitos ocorre dentro dos gametas (que para as células do macho envolvem três rápidos caminhos da divisão celular) e eclodem dentro de um zigoto. No mosquito ocorre a esporogonia, produzindo esporozóitos que acumulam nas glândulas salivares, onde estão prontos para infectar um novo hospedeiro humano (figura 15) (WARD et al., 2004).

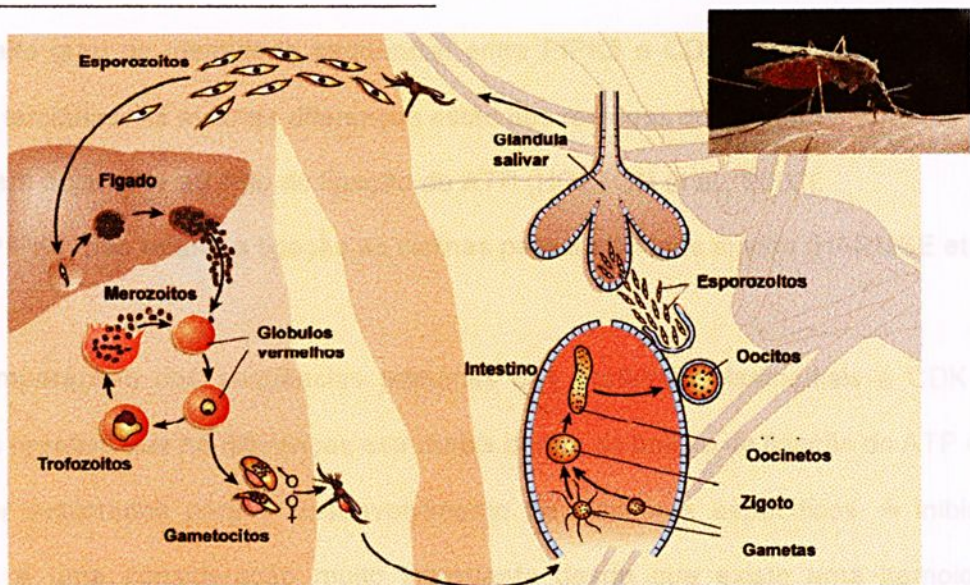


Figura 15: Esquema do ciclo de vida e transmissão da Malária pelo Mosquito transmissor *Anopheles gambiae* (DOOLITTLE, 2002)

A análise de sequência primária entre a PfPK5 e CDKs, mostra que a PfPK5 compartilha 60% de identidade com a CDK1 e CDK5 humana. A estrutura da PfPK5 foi recentemente determinada (código do PDB 1OB3 (HOLTON et al., 2003)), e estão depositadas as estruturas da apoenzima e de três inibidores competitivos com trifosfato de adenosina (ATP), (figuras 16) (HARMSE et al., 2001; HOLTON et al., 2003).

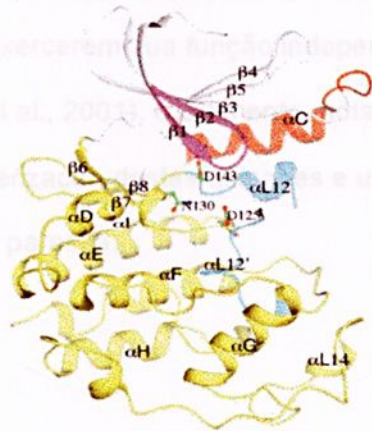


Figure 16: Elementos de estrutura secundária: A estrutura monomérica da PfPK5; (HOLTON et al., 2003).

O alto grau de identidade sequencial entre PfPK5 e CDK2 é refletido nas suas estruturas terciárias. As maiores diferenças estão nos resíduos do domínio C-terminal que se estendem adjacente ao sítio de ligação do ATP (HOLTON et al., 2003).

A PfPK6, não requer a ligação as ciclinas para a atividade *in vitro* (HARMSE et al., 2001).

Comparações das seqüências primárias das proteínas plasmodiais e CDK de mamíferos revelam que há diferenças estruturais dentro do bolsão de ligação do ATP que podem ser explorados para o desenvolvimento de inibidores específicos. A inibição específica é uma consideração muito importante desde que existe uma homologia seqüencial significativa dentro da família das proteínas quinases. A identificação de compostos que demonstram a seletiva inibição do crescimento do parasita e a subsequente investigação de seus alvos moleculares devem providenciar informações sobre a função das enzimas relacionadas à CDK no ciclo de vida dos parasitas (HARMSE et al., 2001).

Estudos indicam que as PfPKs necessitam associar-se a ciclinas para se tomarem ativas, já outros estudos defendem a idéia de que as PfPKs não precisam da associação de ciclinas para serem ativas e exercerem sua função independente delas (BRACCHI-RICARD et al., 2000; HARMSE et al., 2001). Claramente, nota-se a necessidade de estudos adicionais para a caracterização destas quinases e um melhor conhecimento de suas funções no ciclo de vida do parasita.

1.5. DESCRIÇÃO DOS INIBIDORES

A procura por inibidores estende-se por áreas da medicina, para casos como doenças cardiovasculares, doenças neurodegenerativas e até mesmo na luta contra parasitas unicelulares e doenças virais (MALUMBRES e BARBACID, 2001).

Quando o assunto é sobre inibidores de CDKs, fica demonstrado que a seletividade é uma questão chave. Considerando que há conservação dos aminoácidos que formam o bolsão de ligação do ATP, e elevada homologia estrutural na família das CDKs, o desenho de inibidores específicos para as quinases pode ser difícil. Modelos estruturais de CDKs podem providenciar informações sobre o desenho de inibidores seletivos para CDKs. Mais de 50 inibidores farmacológicos de CDKs foram descritos, alguns com atividade potencial antitumor (DAÍ & GRANT, 2003). Eles compartilham algumas propriedades comuns: 1) possuem baixo peso molecular (<600); 2) São planos, hidrofóbicos e heterocíclicos; 3) São competitivos com ATP; 4) se ligam principalmente por interações hidrofóbicas e ligações de hidrogênio; e 5) a cadeia principal e cadeia lateral do aminoácido Leu83 (CDK2) age, respectivamente, como um acceptor e um doador de hidrogênio para os inibidores, enquanto, a cadeia principal Glu81 (CDK2), freqüentemente age como um acceptor de hidrogênio (KNOCKAERT et al., 2002). Vários inibidores de CDKs estão em testes pré-clínicos e clínicos (p.ex. Flavopiridol), e a procura por novos inibidores de CDKs continua (DAÍ & GRANT, 2003).

A seguir serão descritos todos os inibidores usados para a construção dos modelos neste estudo. Os inibidores foram escolhidos segundo os valores de IC50 para cada quinase em estudo (tabela 2).



Tabela 2: Inibidores de CDKs 1, 2, 4 e 5 e os respectivos valores de IC50 em μM (KNOCKAERT et al., 2002).

Inibidor	Classe	CDK2-Ciclina A	CDK4- Ciclina D
Isopenteniladenina	Purina (adenina)	50	>100
Olomoucina	Purina (adenina)	7	>1000
Roscovitina	Purina (adenina)	0,7	>100
Purvanalol B	Purina (adenina)	0,006	>10
H717	Purina (adenina)	-	2
Himentialdisina	Pirrolloazepina	0,07	0,6
Flavopiridol	Flavona	0,1	0,4
Estaurosporina	Indololocarbazola	0,007	<10
Oxindol 91	Oxindol (indolinona)	0,010	0,13
Indirubina-5-sulfonato	Oxindol (indolinona) (bis-indole)	0,04	0,3

1.5.1. Flavopiridol

Flavopiridol, [cis-5,7-diidroxi-2-(2-chlorofenil)-8-[4-(3-hidroxi-1-metil)-piperidinil]-1-benzopirano-4-1] é uma molécula pequena semi-sintética estruturalmente relacionada a um alcalóide flavonóide derivado de rohitukina, que foi extraído da casca da árvore da planta *Dysoxylum binectariferum*, (figuras 17 a e b). O Flavopiridol exibe atividade citotóxica *in vitro* (MURTHI et al., 2000; DAÍ e GRANT, 2003; HARDCASTLE et al., 2002; SAUSVILLE, 2002).

Estas informações estão sendo usadas na procura de novos candidatos a drogas, e para ajudar na identificação de alvos moleculares adequados para intervenções farmacológicas. Seus avanços vêm de sua avaliação pré-clínica e clínica, baseados em sua atividade *in vitro* e *in vivo* e propriedades farmacêuticas e toxicológicas. Estes efeitos biológicos do flavopiridol correlacionam-se com sua habilidade para inibir potencialmente a atividade das CDK1, 2 e 4. Testes clínicos de fase 1 foram concluídos recentemente em

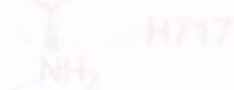


Figura 18: Fórmula estrutural do H717 (KNOCKAERT et al., 2002)

pacientes e está agora entrando na fase 2 (PATEL et al., 1998). Este inibidor induz a interrupção de ambas fases G1 e G2 do ciclo celular (MURTHI et al., 2000).

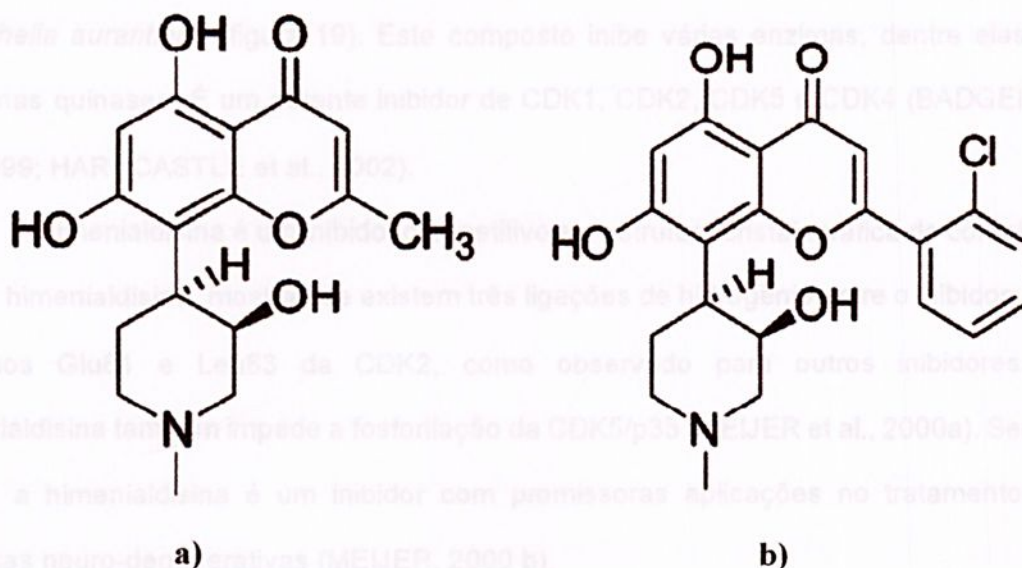


Figura 17: Fórmulas estruturais: a) alcalóide flavonóide originado de uma planta (*Dysoxylum binectariferum*); b) Flavopiridol (HARDCASTLE et al., 2002).

1.5.2. H717

H717 trata-se de 2,6,9-Purina substituída análoga ao ATP e foi sintetizada numa série de derivados substituídos 6-2[*trans*-(4-aminociclohexil)amino]-9-ciclopentenilpurina (figura 18). A estrutura do cristal da CDK2 humana em complexo com H717 amplia a compreensão das interações inibitórias com a proteína (DREYER et al., 2001).

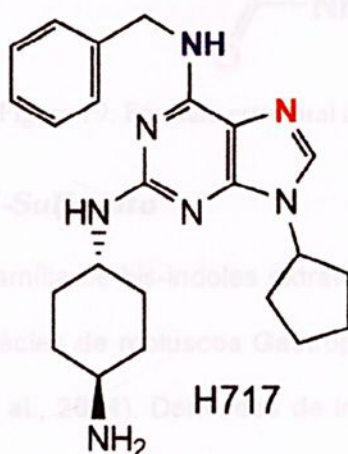


Figura 18: Fórmula estrutural do H717 (KNOCKAERT et al., 2002)

5.1.3. Himenialdisina

Himenialdisina foi originalmente isolada de esponjas marinhas *Axinella verrucosa* e *Acanthella aurantiaca* (figura 19). Este composto inibe várias enzimas, dentre elas as proteínas quinases. É um potente inibidor de CDK1, CDK2, CDK5 e CDK4 (BADGER et al., 1999; HARDCASTLE et al., 2002).

A himenialdisina é um inibidor competitivo e a estrutura cristalográfica do complexo CDK2 himenialdisina, mostra que existem três ligações de hidrogênio entre o inibidor e os resíduos Glu81 e Leu83 da CDK2, como observado para outros inibidores. A himenialdisina também impede a fosforilação da CDK5/p35 (MEIJER et al., 2000a). Sendo assim a himenialdisina é um inibidor com promissoras aplicações no tratamento de doenças neuro-degenerativas (MEIJER, 2000 b).

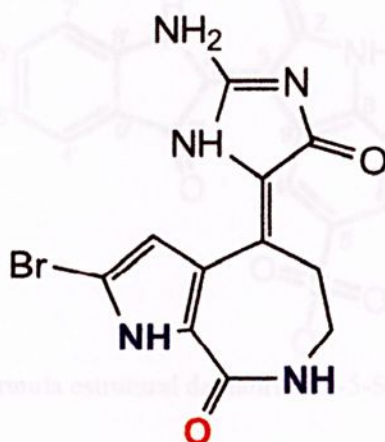


Figura 19: Fórmula estrutural da himenialdisina

5.1.4. Indirubina-5-Sulfonato

As indirubinas, uma família de bis-indoles extraídas de várias fontes naturais, tais como, plantas, diversas espécies de moluscos Gastrópodes e bactérias recombinantes (POLYCHRONOPOULOS et al., 2004). Derivados de indirubina são inibidores potentes

da CDK1, CDK2, CDK4 e CDK5. Foi identificada como o componente ativo de um tradicional medicamento chinês para tratamento de leucemia mielocítica crônica, ingrediente chave de uma mistura de plantas que são usadas para o tratamento deste tipo de leucemia (HARDCASTLE et al., 2002; HOESSEL et al., 1999; POLYCHRONOPOULOS et al., 2004).

A indirubina foi co-cristalizada com CDK2 e CDK2/Ciclina A, (DAVIES et al., 2001), e recentemente com CDK5/p25 e a PfPK5 do *Plasmodium falciparum* (DAVIES et al., 2002).

O mais potente dos derivados de Indirubina é a Indirubina-5-Sulfonato (figura 20). O arranjo das interações é dependente da reestruturação do Loop DFG para uma conformação mais similar do que aquela observada no complexo CDK2/Ciclina A (HOESSEL et al., 1999).

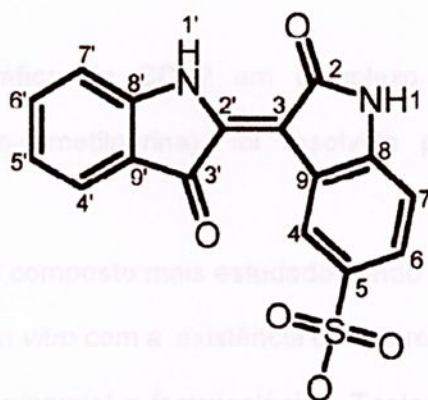


Figura 20: Fórmula estrutural da Indirubina-5-Sulfonato (DAVIES et al., 2001).

5.1.5. Isopenteniladenina

A estrutura do cristal da CDK2 complexada com o inibidor isopenteniladenina (2-(2-hidroxi-etilamino)-6-benzilamino-7-metilpurina) (figura 21), foi resolvida por SCHULZE-GAHMEN et al., (1995), e mostra claramente a densidade eletrônica para todos os átomos do inibidor. A molécula do inibidor é bastante hidrofóbica e, portanto, sua ligação

à CDK2 é caracterizada por interações predominantemente hidrofóbicas e de Van der Waals.

Estruturas tridimensionais de outros complexos com Isopenteniladenina poderão ajudar no entendimento destas interações.

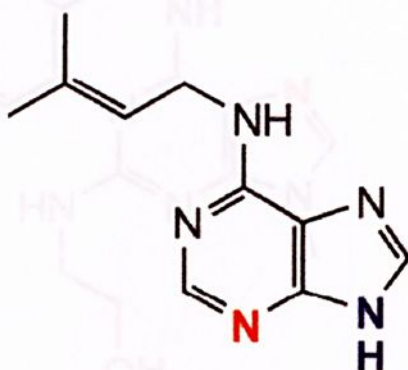


Figura 21: Fórmula estrutural da Isopenteniladenina (KNOCKAERT et al., 2002).

5.1.6. Olomoucina

A estrutura cristalográfica da CDK2 em complexo com Olomoucina (2-(2-hidroxiethylamino)-6-benzilamino-9-metilpurina), foi resolvida por SCHULZE-GAHMEN, (1995), (figura 22).

Olomoucina tem sido o composto mais estudado, tendo diversas vantagens. É um composto bem caracterizado *in vitro* com a existência de estereoisômeros com diferentes atividades e disponibilidade comercial e farmacológica. Testes *in vitro* indicaram que a olomoucina é seletiva para CDKs 1, 2 e 5 (valores de $IC_{50} = 7,7 \mu M$; $0,65 \mu M$ e $0,16 \mu M$ respectivamente). Subseqüentemente, duas análises *in vitro* demonstraram seletividade da olomoucina para CDK7 (MONACO et al., 2004).

Para o complexo CDK2-olomoucina, observa-se uma orientação diferente do anel de purina sendo que o grupo benzil é a única parte da olomoucina que se liga do outro lado do bolsão de ligação, são indícios da especificidade deste inibidor pelas CDKs (SCHUTTE et al., 1997).

Estudos mostraram que altas doses de olomoucina são capazes de bloquear completamente a progressão do ciclo celular e induzir a apoptose (SCHUTTE et al., 1997).

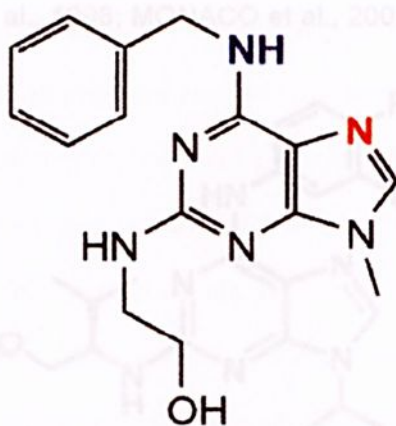


Figura 22: Fórmula estrutural da Olomoucina (KNOCKAERT et al., 2000)

1.5.7. Oxindol 91

O complexo CDK2-Oxindol 91 (1-H-indole-2,3-dione-3-fenilhidrazones), teve sua estrutura cristalográfica resolvida por BRAMSON et al., 2001 (figura 23).

Diversos compostos baseados em oxindol foram identificados como seletivos para quinases plasmodiais.

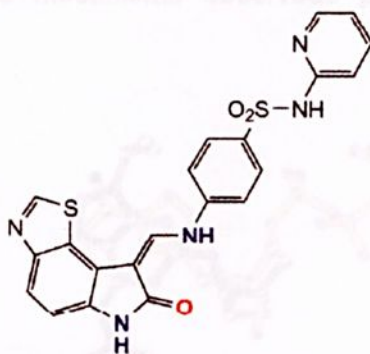


Figura 23: Fórmula estrutural da Oxindol91 (KNOCKAERT et al., 2002).

1.5.8. Purvanalol B

Purvanalol B, 2-(1R-isopropil-2-hidroxietilamino)-6-(3-cloro-4-carboxianilino)-9-isopropilpurina, (figura 24). Foi descoberto por GRAY et al., em 1998. Este inibidor mostra

um alto grau de seletividade, cerca de 22 quinases humana foram testadas, e somente um subgrupo das CDKs (CDK1-ciclina B, CDK2-ciclina A, CDK2-ciclina E e CDK5-p35) foram significativamente inibidas com valores de IC_{50} = 4-6, 6-70 e 6-75 η M, respectivamente (GRAY et al., 1998; MONACO et al., 2004).

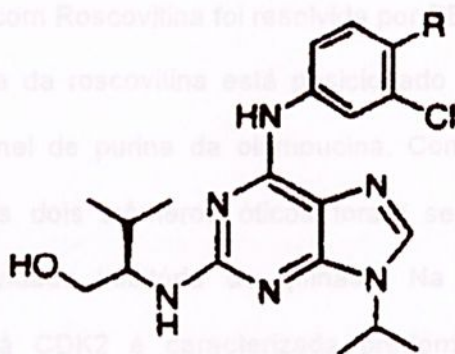


Figura 24: Fórmula estrutural do Purvanalol B (GRAY et al., 1998).

A estrutura cristalográfica da CDK2 em complexo com purvanalol B foi resolvida a 2,05 Å de resolução (figura 25) (GRAY et al., 1998).

Para determinar os efeitos deste inibidor no ciclo celular, purvanalol foi testado em 60 linhas de células tumorais humana. Estudos mostraram que purvanalol potencialmente induz apoptose similar ao mecanismo observado para Olomoucina e Roscovitina (MONACO et al., 2004).

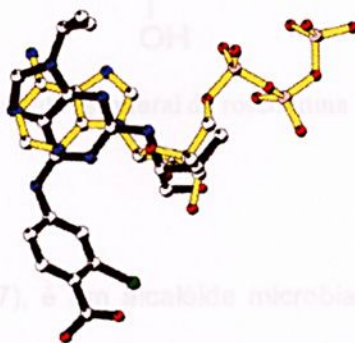


Figura 25: ATP (amarelo) e Purvanalol (preto). As comparações são baseadas na superposição de $C\alpha$ da CDK2. A posição da purina é conservada (GRAY et al., 1998).

1.5.9. Roscovitina

A Roscovitina [2-(1-etil-2-hidroxi-etilamino)-6-benzilamino-9-isopropilpurina] (figura 26), foi identificada através de uma série de substituições de C2, N⁶, N9-adeninas em CDK1/ciclinaB. Roscovitina mostrou ser seletivo para CDKs, e a estrutura cristalográfica da CDK2 em complexo com Roscovitina foi resolvida por DE AZEVEDO et al., (1997).

O anel de purina da roscovitina está posicionado no bolsão de ligação numa orientação similar ao anel de purina da olomoucina. Como a roscovitina contém um carbono assimétrico, os dois isômeros óticos foram separadamente sintetizados e testados para sua atividade inibitória de quinase. Na estrutura complexada com roscovitina, a ligação à CDK2 é caracterizada predominantemente por interações hidrofóbicas e de van der Waals, com os mesmos resíduos hidrofóbicos da enzima que formam o bolsão da adenina no complexo estrutural com ATP (DE AZEVEDO et al., 1997).

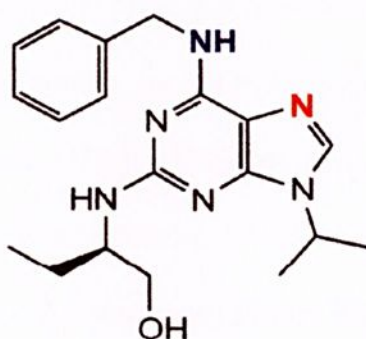


Figura 26: Fórmula estrutural da roscovitina (KNOCKAERT et al., 2002)

1.5.10. Estaurosporina

Estaurosporina (figura 27), é um alcalóide microbial isolado de uma espécie de *Streptomyces*, que têm atraído considerável atenção desde sua primeira caracterização como inibidora da proteína quinase C. Subseqüentemente, ela foi demonstrada como um

potente inibidor de CDKs. A estaurosporina induz apoptose e impede a progressão do ciclo celular (LAWRIE et al., 1997).

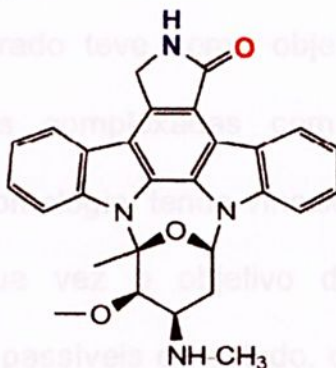


Figura 27: Fórmula estrutural da estaurosporina (KNOCKAERT et al., 2002).

A estrutura molecular da CDK2 complexada com estaurosporina foi resolvida a 2,0 Å. A comparação da estrutura do complexo da CDK2-estaurosporina com outras estruturas de proteínas quinases cristalograficamente determinadas sugerem possíveis explicações para as diferenças observadas na seletividade da estaurosporina em relação as CDKs (LAWRIE et al., 1997).

3. MÉTODOS & FERRAMENTAS

2. OBJETIVOS

Este projeto de mestrado teve como objetivo obter as estruturas tridimensionais de quinases complexadas com inibidores através de Modelagem Molecular por Homologia, tendo vínculo com o projeto Genoma Estrutural que tem por sua vez o objetivo de fornecer a estrutura tridimensional das proteínas passíveis de estudo, que estão codificadas no genoma Humano, sendo elas determinadas por técnicas experimentais tais como Cristalografia e ou Ressonância Magnética Nuclear, ou até mesmo modeladas.

Levando em consideração este objetivo geral, a realização deste trabalho consiste em modelarmos proteínas quinases com estruturas ainda não conhecidas, compreendendo a CDK4 associada com ciclina D complexada com: Flavopiridol, H717, Himenialdisina, Indirubina-5-Sulfonato, Isopenteniladenina, Olomoucina, Oxindol 91, Purvanalol B, Roscovitina e Estaurosporina, além da PfPK6 complexada com Roscovitina e Olomoucina. Ambas proteínas alvos terão como molde a estrutura cristalográfica da CDK2. Tendo assim a finalidade de propor a compreensão da interação proteína/ligante, proteína/proteína e possibilitar o desenho de drogas.



3. MÉTODOS & FERRAMENTAS

3.1. Modelagem Comparativa por Homologia

O genoma humano foi seqüenciado recentemente e constitui a base para a elucidação do genoma Humano. Grande parte dos bioinformatas modernos trabalha com dados de projetos genoma ou transcriptoma. Procurando associar proteínas a suas funções, a bioinformática pode e deverá trazer, nas próximas décadas, suas maiores contribuições à biologia. Desafio maior, entretanto, será a elucidação estrutural de novos alvos moleculares, principalmente proteínas, e receptores, e novas drogas que certamente emergirão dos dados provenientes do genoma humano e de outros genomas em estudo (FILHO SANTOS e DE ALENCASTRO, 2003).

O conhecimento da estrutura tridimensionais de uma proteína constitui uma informação valiosa para a determinação de sua função, pois pode permitir a identificação de domínios, como sítios catalíticos, sítios de regulação alostérica e outros. Apesar das consideráveis inovações técnicas, sobretudo nas áreas de cristalografia de raios X e difração de nêutrons e de ressonância magnética nuclear (RMN), muitos problemas básicos persistem. A obtenção de amostras em quantidade suficiente para os ensaios necessários são, em muitos casos, difíceis e os cristais obtidos nem sempre têm a qualidade necessária para a obtenção de dados de difração de raios X (somente uma em cada vinte proteínas, aproximadamente, produz cristais adequados). Além disso, em certas classes de proteínas, como por exemplo, as proteínas de membrana celular, a determinação estrutural é um desafio. Essas proteínas raramente cristalizam, e dificilmente podem ser tratadas de modo satisfatório por RMN (FILHO SANTOS e DE ALENCASTRO, 2003).

A ferramenta mais bem sucedida de predição de estruturas tridimensionais de proteínas é a modelagem por homologia, também conhecida como modelagem comparativa. A modelagem molecular é um método alternativo, não experimental, que permite, com base nos conhecimentos da estereoquímica dos aminoácidos e nas informações adquiridas das estruturas tridimensionais já resolvidas, prever a conformação de proteínas a partir da sequência primária dos aminoácidos (FILHO SANTOS e DE ALENCASTRO, 2003).

Uma das formas de se realizar a modelagem de proteínas é utilizar como referência uma ou mais proteínas homólogas e de estruturas tridimensionais conhecidas. Esta abordagem baseia-se em alguns padrões gerais que têm sido observados, em nível molecular, no processo de evolução biológica:

- (a) homologia entre sequências de aminoácidos implica em semelhança estrutural e funcional;
- (b) proteínas homólogas apresentam regiões internas conservadas (principalmente constituídas de elementos de estrutura secundária: hélices- α e fitas- β);
- (c) as principais diferenças estruturais entre proteínas homólogas ocorrem nas regiões externas, constituídas principalmente por alças ("loops"), que ligam os elementos de estruturas secundárias.

Outro fator importante é que as proteínas agrupam-se em um número limitado de famílias tridimensionais. Estima-se que existam cerca de 5.000 famílias protéicas. Conseqüentemente, quando se conhece a estrutura de pelo menos um representante de uma família, é geralmente possível modelar, por homologia, os demais membros da família. É bom lembrar que similaridade e homologia são conceitos diferentes. O alinhamento indica o grau de similaridade entre sequências, já a homologia é uma hipótese de cunho evolutivo, e não possui gradação (SALI e BLUNDELL, 1993). O mecanismo evolucionário de duplicação de genes, associado às mutações, leva a



divergências moleculares e, conseqüentemente, à formação de famílias de proteínas estruturalmente relacionadas. Proteínas derivadas de um ancestral comum são ditas **homólogas**. Em função do número de mutações envolvidas, as seqüências de aminoácidos de proteínas homólogas podem ser, idênticas ou semelhantes. Conseqüentemente, a semelhança entre as seqüências de aminoácidos em proteínas homólogas, expressa pelo grau (percentual) de identidade, é menos preservada do que a semelhança de estruturas tridimensionais. Em outras palavras, as estruturas tridimensionais de proteínas homólogas tendem a se conservar porque a estrutura ancestral comum é crucial para a manutenção da função destas proteínas (FILHO SANTOS e DE ALENCASTRO, 2003). É importante enfatizar que a conservação da estrutura terciária de proteínas homólogas não é uma propriedade intrínseca das proteínas, mas uma conseqüência da evolução, regida por restrições funcionais. A modelagem de uma proteína (proteína-alvo) pelo método da homologia baseia-se no conceito de evolução molecular. Isto é, parte-se do princípio de que a semelhança entre as estruturas primárias desta proteína e de proteínas homólogas de estruturas tridimensionais conhecidas (proteínas-molde) implica em similaridade estrutural entre elas (FILHO SANTOS e DE ALENCASTRO, 2003). Os métodos correntes de modelagem de proteínas por homologia implicam basicamente em quatro passos sucessivos: identificação e seleção de proteínas-molde; alinhamento das seqüências primárias; construção das coordenadas do modelo e validação. A seguir é exibido um esquema geral do processo de modelagem de proteínas por homologia (Figura 28) (SALI e OVERINGTON, 1994).

se conhecida, que servirá de molde para a determinação da estrutura da proteína-alvo. Duas situações são, em geral, possíveis: conhece-se a família protéica a que pertence a proteína-alvo, para este caso é suficiente selecionar a proteína-molde diretamente do PDB (BERNSTEIN et al., 1977; BERMAN et al., 2000) (<http://www.rcsb.org/pdb/>); não se sabe a que família a proteína-alvo pertence, para este

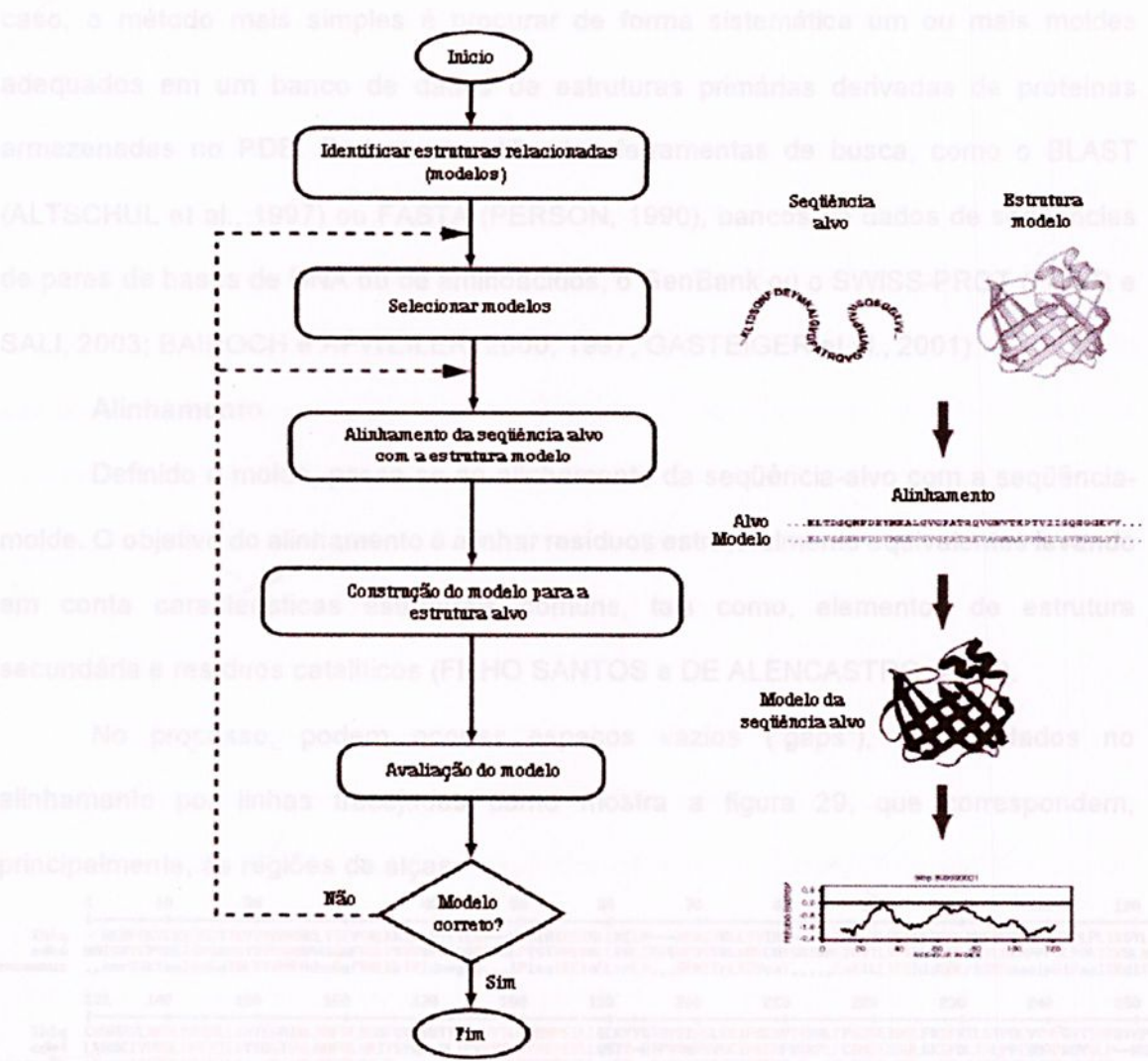


Figura 28: Esquema dos passos da modelagem molecular (SALI e BLUNDELL, 1993)

Identificação e seleção de proteínas molde

A primeira etapa do método é a identificação de pelo menos uma proteína de estrutura tridimensional conhecida, que servirá de molde para a determinação da estrutura da proteína-alvo. Duas situações são, em geral, possíveis: conhece-se a família protéica a que pertence a proteína-alvo, para este caso é suficiente selecionar a proteína-molde diretamente do PDB (BERNSTEIN et al., 1977; BERMAN et al., 2000) (<http://www.rcsb.org.pdb>); não se sabe a que família a proteína-alvo pertence, para este

caso, o método mais simples é procurar de forma sistemática um ou mais moldes adequados em um banco de dados de estruturas primárias derivadas de proteínas armazenadas no PDB. Podem ser utilizadas ferramentas de busca, como o BLAST (ALTSCHUL et al., 1997) ou FASTA (PERSON, 1990), bancos de dados de seqüências de pares de bases de DNA ou de aminoácidos, o GenBank ou o SWISS-PROT (FISER e SALI, 2003; BAIROCH e APWEILER, 2000; 1997; GASTEIGER et al., 2001).

Alinhamento

Definido o molde, passa-se ao alinhamento da seqüência-alvo com a seqüência-molde. O objetivo do alinhamento é alinhar resíduos estruturalmente equivalentes levando em conta características estruturais comuns, tais como, elementos de estrutura secundária e resíduos catalíticos (FILHO SANTOS e DE ALENCASTRO, 2003).

No processo, podem ocorrer espaços vazios ("gaps"), representados no alinhamento por linhas tracejadas como mostra a figura 29, que correspondem, principalmente, às regiões de alças.

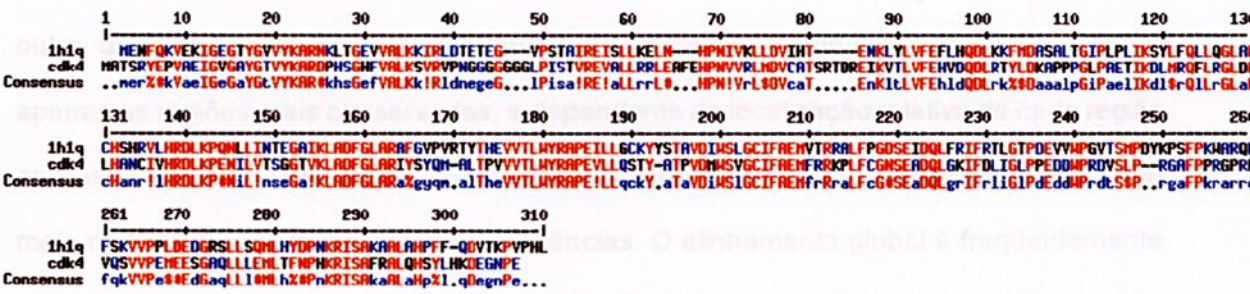


Figura 29: Alinhamento de seqüências primárias da proteína-molde com a proteína-alvo exibindo os "gaps".

Se, na etapa de identificação de moldes, mais de uma estrutura for selecionada, procura-se alinhar a seqüência-alvo com as seqüências de todos os moldes disponíveis. Neste caso, obtém-se um alinhamento múltiplo.

A maior parte dos métodos de alinhamento baseia-se em técnicas de programação dinâmica. Os programas mais utilizados nos processos de alinhamentos são o BLAST (ALTSCHUL et al., 1997), o FASTA (PERSON, 1990), o CLUSTAL (THOMPSON, et al.,

1994), e o MultAlin (CORPET, 1988), todos disponíveis como servidores na Internet. Após o alinhamento, é possível reconhecer regiões estruturalmente conservadas e regiões variáveis. As primeiras correspondem às regiões de máxima similaridade, isto é, em que as conformações são muito semelhantes. Nas regiões variáveis não há, em geral, correspondência estrutural sendo constituídas de alças (FISER e SALI, 2003).

Para o presente trabalho foi utilizado o programa MultAlin. O processo consiste em introduzir espaços (gaps) entre os monômeros de uma ou mais seqüências a fim de obter o melhor alinhamento possível. A qualidade de um alinhamento é determinada pela soma dos pontos obtidos para cada unidade pareada ("match"), menos as penalidades pela introdução de gaps e posições não pareadas ("mismatch") (FILHO SANTOS e DE ALENCASTRO, 2003).

Quanto à região analisada, o alinhamento de seqüências pode ser grosseiramente classificado em dois tipos, o alinhamento global e o alinhamento local. No alinhamento global, as seqüências envolvidas devem ser alinhadas de um extremo ao outro, dando origem a apenas um resultado. Já no alinhamento local, procura-se alinhar apenas as regiões mais conservadas, independente da localização relativa de cada região em sua seqüência. Conseqüentemente, este alinhamento tem como resultado uma ou mais regiões conservadas entre as seqüências. O alinhamento global é freqüentemente utilizado para determinar regiões mais conservadas de seqüências homólogas e o programa MultAlin faz uso deste alinhamento (CORPET, 1988).

Construção do Modelo

Feito o alinhamento, uma variedade de métodos pode ser usada para construirmos um modelo tridimensional da proteína alvo. O primeiro método executa a modelagem através da montagem de corpo rígido. Este método constrói o modelo a partir de algumas regiões centrais, alças e cadeias laterais, que são obtidas ao se dissecar as estruturas relacionadas. O segundo método modela os segmentos emparelhando-os e confiando-se



nas posições aproximadas dos átomos conservados nos modelos e calculando as coordenadas de outros átomos. O terceiro método faz a modelagem satisfazendo as restrições espaciais e usa distância geométrica ou técnicas de otimização para satisfazer as restrições espaciais obtidas do alinhamento da sequência designada com o modelo estrutural. As precisões dos métodos são relativamente semelhantes quando adequadamente usados (FISER e SALI, 2000). E para o presente projeto foi utilizado o terceiro método.

Otimização do Modelo gerado

Após a construção do modelo da proteína-alvo é necessário otimizá-lo. As interações desfavoráveis entre átomos não-ligados ("nonbonded contacts"), bem como as energias de ângulos torcionais e de ligação são otimizadas por mecânica molecular. Como regra geral, esta otimização não pode ser extensiva para não desviar o sistema de seu molde original. Conseqüentemente, as etapas de otimização e validação do modelo devem ser efetuadas simultaneamente (FISER, et al., 2000).

Avaliação dos Modelos

A validação é uma etapa essencial que pode ser executada em diferentes níveis de organização estrutural. Deve-se avaliar a qualidade do empacotamento global da proteína, os possíveis erros estruturais em regiões localizadas e os parâmetros estereoquímicos. Como a qualidade do modelo depende muito da estrutura da proteína-molde utilizada, é preciso que esta última tenha sido obtida em alta resolução e com um fator-R satisfatório. É preferível usar como moldes, estruturas cuja resolução seja igual ou superior a 2 Å e fator-R inferior a 20% (FILHO SANTOS e DE ALENCASTRO, 2003). Um primeiro requisito, que uma proteína adequadamente modelada deve atender é ter uma estrutura tridimensional satisfatória. Sua qualidade depende da proteína escolhida como molde e do alinhamento calculado. É importante verificar se existem grandes diferenças conformacionais não explicadas entre os elementos de estrutura secundária (regiões

conservadas) das estruturas-molde e da estrutura-modelada (FILHO SANTOS e DE ALENCASTRO, 2003).

A qualidade estereoquímica do modelo é de importância fundamental. O programa mais utilizado na avaliação dos parâmetros estereoquímicos, o PROCHECK (LASKOWSKI, et al., 1993), avalia os comprimentos de ligação, os ângulos planos, a planaridade dos anéis de cadeias laterais, a quiralidade, as conformações das cadeias laterais, a planaridade das ligações peptídicas, os ângulos torcionais da cadeia principal e das cadeias laterais, o impedimento estérico entre pares de átomos não-ligados e a qualidade do gráfico de Ramachandran (RAMAKRISHNAN e RAMACHANDRAN, 1965). O gráfico de Ramachandran é particularmente útil porque ele define os resíduos que se encontram nas regiões mais favoráveis e desfavoráveis e orienta a avaliação da qualidade de modelos teóricos ou experimentais de proteínas (RAMAKRISHNAN e RAMACHANDRAN, 1965).

Possíveis Limitações e Erros do Método

A probabilidade de ocorrência de erros em um modelo de proteína construído pelo método da homologia é inversamente proporcional ao grau de identidade entre as seqüências primárias utilizadas. No entanto, CHOTHIA e LESK (1987), demonstraram que relações mais favoráveis entre a qualidade de modelos gerados e o grau de identidade com o molde são favoráveis. Em consequência, modelar uma proteína com baixo grau de identidade com o molde utilizado é geralmente muito mais difícil.

Em geral, os erros não se distribuem igualmente por toda a estrutura. Como um primeiro critério, os r.m.s.d (desvio da raiz quadrada média) das distâncias entre os átomos da cadeia principal de estruturas conservadas de pares de proteínas dependem do grau de identidade das seqüências utilizadas. As regiões cuja estrutura é conservada são mais confiáveis que as regiões de alças. Quando o grau de identidade é superior a cerca de 30%, os resíduos formados por hélices- α e fitas- β têm, em geral, bom



alinhamento com os resíduos correspondentes do molde. Abaixo deste grau de identidade, as fitas aparentemente se alinham melhor que as hélices (FISER e SALI, 2003).

As etapas mais problemáticas da modelagem de proteínas por homologia são o alinhamento das seqüências e a modelagem das regiões de alças. Cada modelagem é um problema particular e deve-se ter sempre em mente que a construção de um modelo de proteína é uma tarefa que deve ser fundamentada em princípios de evolução molecular para cada grupo específico de proteínas homólogas. Quando as proteínas têm vários domínios, a construção do modelo é um processo ainda mais complexo (FILHO SANTOS e DE ALENCASTRO, 2003).

Em resumo, os erros normalmente encontrados em modelos de proteínas gerados por homologia podem ser divididos em cinco categorias (FISER e SALI, 2003):

1. *Erros de modelagem das cadeias laterais.* Quando o grau de identidade entre a estrutura-alvo e a estrutura-molde é baixo podem surgir problemas de encaixe de algumas cadeias laterais no corpo principal do modelo, formando regiões interrompidas.

2. *Deslizamento em regiões do alinhamento.* Divergências entre seqüências podem provocar distorções durante a localização dos resíduos da cadeia principal, mesmo que o enovelamento global do modelo seja satisfatório. A utilização de várias proteínas-molde pode reduzir este tipo de erro.

3. *Erros em regiões sem molde.* Este tipo de erro pode ocorrer em segmentos da estrutura-alvo que não têm equivalentes na estrutura-molde, principalmente em regiões de alças grandes. A avaliação crítica do alinhamento proposto e do ambiente em que a alça está localizada pode levar à superação deste tipo de erro.

4. *Erros de alinhamento.* Ocorrem principalmente quando o grau de identidade entre as seqüências envolvidas é inferior a 30%. Para reduzi-los, deve-se dar preferência



a alinhamentos múltiplos e, se possível, modificar iterativamente as regiões do alinhamento tidas como problemáticas.

5. *Moldes incorretos.* Este tipo de erro é mais provável quando o grau de identidade é inferior a 25%. Nestes casos, deve-se buscar parâmetros experimentais e modificar com cautela o alinhamento.

Comentários

Apesar das limitações inerentes ao método, a modelagem de proteínas por homologia é uma ferramenta adequada para a predição teórica da estrutura de proteínas. A técnica tem sido aplicada com sucesso na indústria farmacêutica e na pesquisa fundamental, sendo uma poderosa alternativa para a superação dos problemas envolvidos na elucidação de estruturas protéicas por técnicas experimentais (FILHO SANTOS e DE ALENCASTRO, 2003).

A modelagem por homologia é um processo iterativo de ajuste de parâmetros e verificação dos resultados. Normalmente é necessário que o processo seja repetido várias vezes até que uma estrutura terciária adequada seja obtida. Além disso, a modelagem de proteínas, como um todo, é uma técnica heurística: mesmo que a estrutura obtida concorde perfeitamente com todas as restrições impostas, não há garantias de que esteja correta. Deve-se lembrar que uma estrutura bastante semelhante à real pode ser o suficiente para formulação de novas hipóteses e atingir as expectativas do usuário desta técnica (FISER e SALI, 2003).

Uma abordagem recente, que possui um crescente número de adeptos e acumula bons resultados, é a modelagem através de "threading" de proteína. Esta técnica é baseada na comparação da proteína em questão com modelos descritivos dos enovelamentos de proteínas homólogas. Nesses modelos são descritas: a distância entre os resíduos de aminoácidos, a estrutura secundária de cada fragmento e as

dinâmica molecular com recozimento simulado (*simulated annealing*). (Apêndice).



características físico-químicos de cada resíduo (FILHO SANTOS e DE ALENCASTRO, 2003).

Um grande desejo dos que trabalham com proteínas é o desenvolvimento de programas realmente eficientes para a modelagem *ab initio*, ou seja, que sejam capazes de prever a estrutura terciária de uma proteína, tendo como informação apenas a sequência dos resíduos de aminoácidos e suas interações físico-químicas, entre si e com o meio. Programas assim existem hoje, mas ainda tem muito a melhorar para que possa se confiar unicamente nos seus resultados (FILHO SANTOS e DE ALENCASTRO, 2003).

A modelagem, propriamente dita, é realizada através de softwares como o MODELLER e SWISS-MODEL (KOPP & SCHWEDE, 2004) dentre outros. Esses programas normalmente procuram encontrar a estrutura tridimensional que melhor se aproxime da disposição dos átomos das proteínas utilizadas como modelo, e ao mesmo tempo atenda às restrições estereoquímicas. Para a construção deste projeto foi utilizado o programa MODELLER (Apêndice).

3.2. Ferramentas utilizadas durante o processo da Modelagem Molecular

Para a realização do presente projeto de mestrado foi usado o programa MODELLER (SALI e BLUNDELL, 1993) o qual implementa uma abordagem automatizada para a modelagem comparativa de estruturas de proteínas baseada na satisfação de restrições espaciais.

O MODELLER nos fornece a possibilidade de variar a conformação inicial que é em seguida otimizada durante o processo de modelagem que compreende de restrições espaciais que foram automaticamente combinados em uma função objetivo a ser otimizada no espaço cartesiano utilizando-se os métodos de gradientes conjugados e dinâmica molecular com recozimento simulado (*simulated annealing*), (Apêndice).



O ponto inicial para a realização das modelagens comparativas foi a identificação de estruturas de proteínas relacionadas com a seqüência alvo, e então, a seleção das estruturas dos moldes, sendo a estrutura cristalográfica da proteína CDK2.

Para esta etapa do processo de modelagem foi utilizado os bancos de dados de proteínas (PDB) e Swiss-Prot. Tendo em mãos as seqüências primárias dos possíveis moldes relacionados com os alvos, foi feito os alinhamentos destas seqüências utilizando o programa MultAlin (CORPET, 1988).

A porcentagem de identidade obtida a partir desses alinhamentos ficaram em torno dos 29% á 46%. Para realização das modelagens foi utilizado o Programa Parmodel (UCHÔA et al., 2004) que é a paralelização do programa MODELLER. Trata-se de um serviço de internet para automatizar a modelagem por homologia de proteínas e avaliar as estruturas modeladas. O programa foi paralelizado em um cluster Beowulf com 16 nós (BioComp, AMD Athlon XP 2100+). Mil modelos foram gerados para cada seqüência alvo e com isso tendo uma maior otimização e confiabilidade nos modelos.

O Parmodel integra os principais programas usados neste processo, sendo dividido em quatro módulos: Parmodel Modelagem (figura30), Parmodel Avaliação, Parmodel Visualização e Parmodel Otimização (UCHÔA et al., 2004).

Na avaliação dos modelos, o Parmodel integra ferramentas utilizadas para a validação destes modelos: PROCHECK (figura31) e VERIFY-3D (BOWIE et al., 1991).

Após o processo da modelagem foi escolhido, dos 1000 modelos gerados, o que apresentou melhor qualidade estereoquímica.

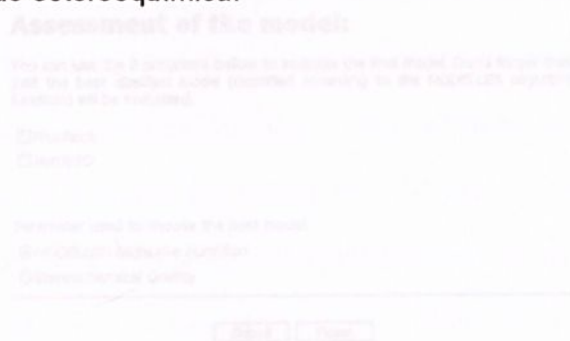


Figura30: Página de entrada para submissão dos moldes e seqüências alvos (UCHÔA et al., 2004)

User's data:

Be aware of putting your correct e-mail because the results are sent through it.

User's Name

Institution

E-mail address*

Necessary parameters for modeling:

At first you must specify the number of models will be built (max number of models = 1 to 1000). Don't forget that the models will be built from a same alignment and the changes in the initial spatial restraints by MODELLER will generate distinct models. You must submit a file with the sequence of the protein (in one letter code) that will be modeled and afterwards you can submit up to 8 files with the templates (in pdb format) for the modeling. At last you must choose the method used to solve the template(s) and the number of ligands (if there is some ligand) present in the template(s).

Number of Models*

From

to

Sequence* (In one letter code)

Procurar...

Templates* (In pdb format)

Procurar...

Procurar...

Procurar...

Procurar...

Procurar...

Procurar...

Procurar...

Experimental Method (Templates)

☒ X-ray Diffraction

☐ NMR

☐ Modeling

Number of Ligands* (Choose the number of ligands in the protein)

☒ 0 ligand

☐ 1 ligand

☐ 2 ligands

☐ 3 ligands

Assessment of the model:

You can use the 2 programs below to evaluate the final model. Don't forget that just the best qualified model (classified according to the MODELLER objective function) will be evaluated.

☐ Procheck

☐ Verify3D

Parameter used to choose the best model:

☒ MODELLER Objective Function

☐ Stereochemical Quality

Submit

Reset

Figura30: Página de entrada para submissão dos moldes e seqüências alvos (UCHÔA et al., 2004)

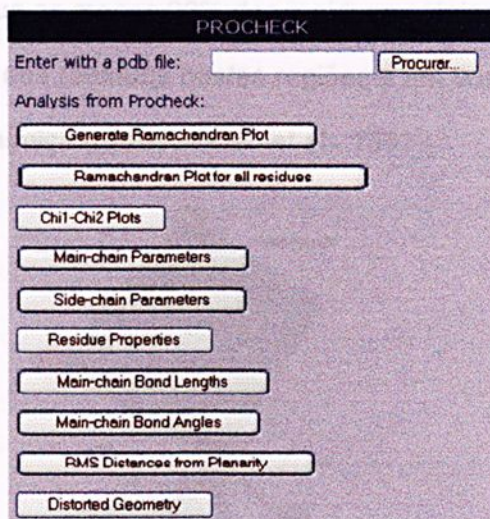


Figura 31: Página para submissão dos modelos para validação no Parmodel (UCHÔA et al., 2004)

O PROCHECK (LASKOWSKI ET AL., 1993) dentre outras funções, avalia a qualidade estereoquímica dos modelos o LSQKAB (CCP4, 1994) realiza a sobreposição das estruturas tridimensionais, o X-PLOR (BRUNGER, 1992) analisa a geometria ideal e o AREAIMOL E RESAREA (CCP4, 1994) que calcula a área de contato proteína-proteína, proteína-ligante.

O princípio do PROCHECK é o acesso à geometria dos resíduos numa dada estrutura de proteína. Esta é comparada com parâmetros estereoquímicos, derivados de estruturas bem refinadas e de alta resolução (LASKOWSKI et al., 1993).

A entrada do PROCHECK é um simples arquivo pdb contendo as coordenadas de sua estrutura de proteína. Os átomos de hidrogênio são ignorados (LASKOWSKI et al., 1993). As saídas deste programa compreendem um número de gráficos com uma detalhada listagem resíduo a resíduo. Dentre estes gráficos está o Gráfico de Ramachandran que exhibe os ângulos de torção Φ e Ψ para todos os resíduos da estrutura. Os resíduos de glicina e prolina são identificados por triângulos. A coloração do gráfico permite delimitar as regiões (MORRIS et al., 1986). Regiões em vermelho

correspondem ao "núcleo" representando as combinações mais favoráveis dos valores de phi-psi (figura 32). O ideal é que ao menos 90% dos resíduos estejam nestas regiões mais favoráveis. As porcentagens dos resíduos nestas regiões é um dos melhores guias para avaliar a qualidade estereoquímica (LASKOWSKI et al., 1993).

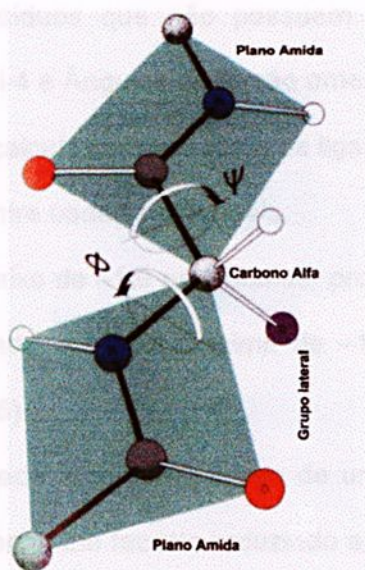


Figura 32: Esquema da cadeia polipeptídica indicando os ângulos Φ e ψ (VOET et al., 2002).

Os Ângulos ϕ (fi) e ψ (psi), são utilizados para a designação de rotações em torno das ligações C-N e C α -C, respectivamente. A conformação do esqueleto de uma proteína pode ser descrita pela especificação dos valores de ϕ e ψ para cada resíduo. Dois tipos de estrutura secundária ocorrem freqüentemente nas proteínas, que são as estruturas de hélices α e de folhas β pregueadas, mantidas por ligações de hidrogênio. A hélice α e as folhas β não são as únicas estruturas secundárias possíveis mas, elas são, sem dúvida, as mais importantes (RAMAKRISHNAN e RAMACHANDRAN, 1965).

Outro resultado obtido pelo programa PROCHECK são os valores do Fator G, que se trata de uma medida de todas as normalidades da estrutura. Os valores gerais são obtidos a partir de uma média de todos os Fatores G diferentes para cada resíduo da estrutura (LASKOWSKI et al., 1993).

O Fator G, oferece uma medida de uma propriedade estereoquímica, se é dada como normal, ou não. Valores do Fator G são computados para as seguintes propriedades (LASKOWSKI et al., 1993):

- a) Ângulos de Torsão: combinações de phi-psi; combinações de chi1-chi2; torsões de chi1 para aqueles resíduos que não possuem um chi2; Ângulos de torsão combinando chi-3 e chi-4 e Ângulos de torsão omega.
- b) Geometria covalente: calcula comprimentos de ligações entre cadeias principais e Ângulos de ligações entre cadeias principais.

Valores do Fator G abaixo de $-1,0$ pode indicar propriedades que necessitam ser investigados com mais cuidado, e valores acima de $-1,0$ indicam boa qualidade da estrutura (MORRIS et al., 1986).

O método Perfil 3D mede a compatibilidade de uma sequência de aminoácidos com uma estrutura tridimensional. Ele faz isto reduzindo a estrutura tridimensional a uma representação unidimensional simplificada chamada cadeia de ambiente, que então é comparado com a sequência de aminoácido unidimensional. Também pode ser usado para verificar a validade de uma estrutura de proteína obtida através de modelagem molecular medindo a compatibilidade daquela estrutura com a própria sequência da proteína. Apesar da aplicação, o método implica três operações básicas:

- Redução da estrutura tridimensional a uma cadeia unidimensional de ambientes de resíduos. Esses ambientes são categorizados segundo a área do resíduo que está na parte interna da proteína, a fração da área da cadeia lateral que está exposta a moléculas polares e a estrutura secundária local.
- Geração de uma matriz de comparação dependente da posição conhecida como perfil 3D. É calculada através da cadeia de ambientes de resíduo, penalidade de "gaps" e por uma matriz de classificação pré-calculada. As penalidades de "gaps" podem variar com a posição no perfil. A matriz de classificação é determinada



através da probabilidade de encontrar cada um dos vinte aminoácidos em cada uma das classes de ambiente como observado em um banco de dados de estruturas conhecidas e seqüências relacionadas.

- Alinhamento de uma seqüência com o perfil 3D. O resultado do escore de alinhamento é uma medida da compatibilidade da seqüência com a estrutura descrita pelo perfil 3D.

O uso do Perfil 3D aliado a outros métodos pode oferecer resultados mais confiáveis na avaliação de modelos tridimensionais de proteínas.

Para verificação da estrutura usando o método do perfil 3D é necessário compensar os efeitos de comprimento e também comparar com o escore obtido usando estruturas conhecidas. LÜTHY et al., (1992) trataram este problema calculando os escores da auto-compatibilidade 3D-1D para todas as estruturas do banco de dados de Brookhaven (BERNSTEIN et al., 1977) determinadas a alta resolução ($\leq 1,5 \text{ \AA}$) e com R-fatores em torno de 20%. Fizeram um gráfico de dispersão em uma escala "log-log", entre o total de escore com o comprimento da seqüência e este gráfico apresentou uma linha quase reta. Um ajuste linear feito nestes dados foram transformados em coordenadas logarítmicas, através da relação abaixo:

$$S_{calc} = \exp(-0,83 + 1,008x \ln(L))$$

Onde S_{calc} é o escore esperado para uma estrutura correta que tem uma seqüência de tamanho L. Se o escore estiver abaixo de $0,45S_{calc}$, ou próximo deste valor indica uma estrutura de qualidade questionável. Um escore abaixo de $0,45S_{calc}$ indica uma estrutura incorreta. Um escore acima de $0,45 S_{calc}$ indica uma estrutura de confiança.

Valores de R.M.S.D. (desvio da raiz quadrada média) para a geometria ideal foram calculados usando X-PLOR os quais compreendem: comprimento de ligação (espera-se obter valores até $1 \text{ \AA}$); ângulos de ligação (espera-se obter valores inferior e/ou até $10,0^\circ$);



MÉTODOS E FERRAMENTAS

ângulos diedrais (valores ideais estão acima de 20°) e ângulos impróprios (valores ideais devem ser igual ou inferior a 90°).

Para o cálculo do potencial eletrostático de superfície, foi usado o programa GRASP (NICHOLLS et al., 1991) e para visualização dos elementos de estrutura secundária das proteínas modeladas foi usado o programa MOLMOL (KORADI et al., 1996). Para a análise das interações proteína-ligantes foi usado o programa XtalView (MCREE, 1999).

Todas estas ferramentas foram utilizadas para auxiliar nas análises quantitativas e qualitativas das proteínas em estudo, possibilitando a validação e avaliação dos modelos gerados através do método de modelagem molecular comparativa.

A busca realizada no PDB forneceu as coordenadas atômicas da CDK2/Ciclina A complexada com inibidor vC094, (código 1H1Q) resolvida a $2.5 \text{ \AA}$. Como era necessário a obtenção das coordenadas da CDK associada à ciclina, e em complexo com ligantes, esta foi a única opção. É importante lembrar que a CDK2 (Código Swiss_Prot: P24941) possui 298 resíduos de aminoácidos com peso molecular de 33935 Da e a ciclina A (Código Swiss_Prot: CCNA ou P78386) possui 256 resíduos com peso molecular de 52358 Da , contra CDK4 que possui 303 resíduos (Código Swiss_Prot: P11802) com peso molecular de 33730 Da e a ciclina D que possui 293 resíduos (Código Swiss_Prot: CCND ou P30281) com peso molecular de 32520 Da .

A busca pelos inibidores da CDK também foi realizada no banco de dados de proteínas, onde foram utilizadas estruturas de CDK2 resolvidas com os mesmos

4. RESULTADOS E DISCUSSÕES

A seguir serão apresentados e discutidos os resultados obtidos para a CDK4 complexada com inibidores e associada com ciclina D, e para a proteína quinase 6 do *Plasmodium falciparum* também na presença de inibidor. Ambas proteínas alvo tiveram como molde a estrutura cristalográfica da CDK2.

4.1. Modelagem da CDK4/Ciclina D complexada com inibidores

4.1.1 CONSTRUÇÃO DOS MODELOS

A construção dos modelos de CDK4/ciclinaD necessitou a elaboração de um protocolo diferente das modelagens realizadas para proteínas e ligantes em geral, pois neste caso, além da estrutura dimérica, também havia a presença de ligantes.

Primeiramente, foi necessário a preparação dos moldes, ou seja, obtenção das coordenadas atômicas de estruturas resolvidas a alta resolução. No caso foi utilizado a estrutura da CDK2 associada à ciclina A em complexo com inibidores.

A busca realizada no PDB forneceu as coordenadas atômicas da CDK2/Ciclina A complexada com inibidor v6094, (código 1H1Q) resolvida a 2,5 Å. Como era necessário a obtenção das coordenadas da CDK associada à ciclina, e em complexo com ligantes, esta foi a única opção. É importante lembrar que a CDK2 (Código Swiss_Prot: P24941) possui 298 resíduos de aminoácidos com peso molecular de 33930 Da e a ciclina A (Código Swiss_Prot: CCNA ou P78396) possui 256 resíduos com peso molecular de 52358Da, contra CDK4 que possui 303 resíduos (Código Swiss_Prot: P11802) com peso molecular de 33730Da e a ciclina D que possui 293 resíduos (Código Swiss_Prot: CCND ou P30281) com peso molecular de 32520Da.

A busca pelos inibidores de CDK também foi realizado no banco de dados de proteínas, onde foram utilizadas estruturas de CDK2 resolvidas com os mesmos

Figura 33: Alinhamentos de seqüências primárias: a) alinhamento da CDK2 (1H1Q) com a CDK4; b) alinhamento da Ciclina A (1H1Q) com a Ciclina D (MultiAlign, CORPET, 1988)

inibidores. A tabela3 relaciona os códigos PDB para os inibidores e os valores de IC50 para a CDK4.

Tabela 3: Inibidores de CDK4 e os respectivos valores de IC50

Inibidores	IC50 (µM)	Código PDB
Flavopiridol	0,4	ND
H717	2	1G5S
Himentialdisina	0,6	1DM2
Indirubina-5-sulfonato	0,3	1E9H
Isopenteniladenina	>100	ND
Olomoucina	>1000	ND
Oxindol 91	0,13	1FVV
Purvanalol B	>10	1CKP
Roscovitina	>100	ND
Estausporina	<10	1AQ1

ND: Não depositado no PDB

Foram realizados alinhamentos entre as seqüências primárias da CDK2 com a CDK4 e entre a ciclina A e a ciclina D e foram obtidos 45,2% de identidade entre as CDKs, 29,2% de identidade entre as ciclinas (figuras 33 a e b). Estes valores sugerem que estas proteínas são passíveis de modelagem.

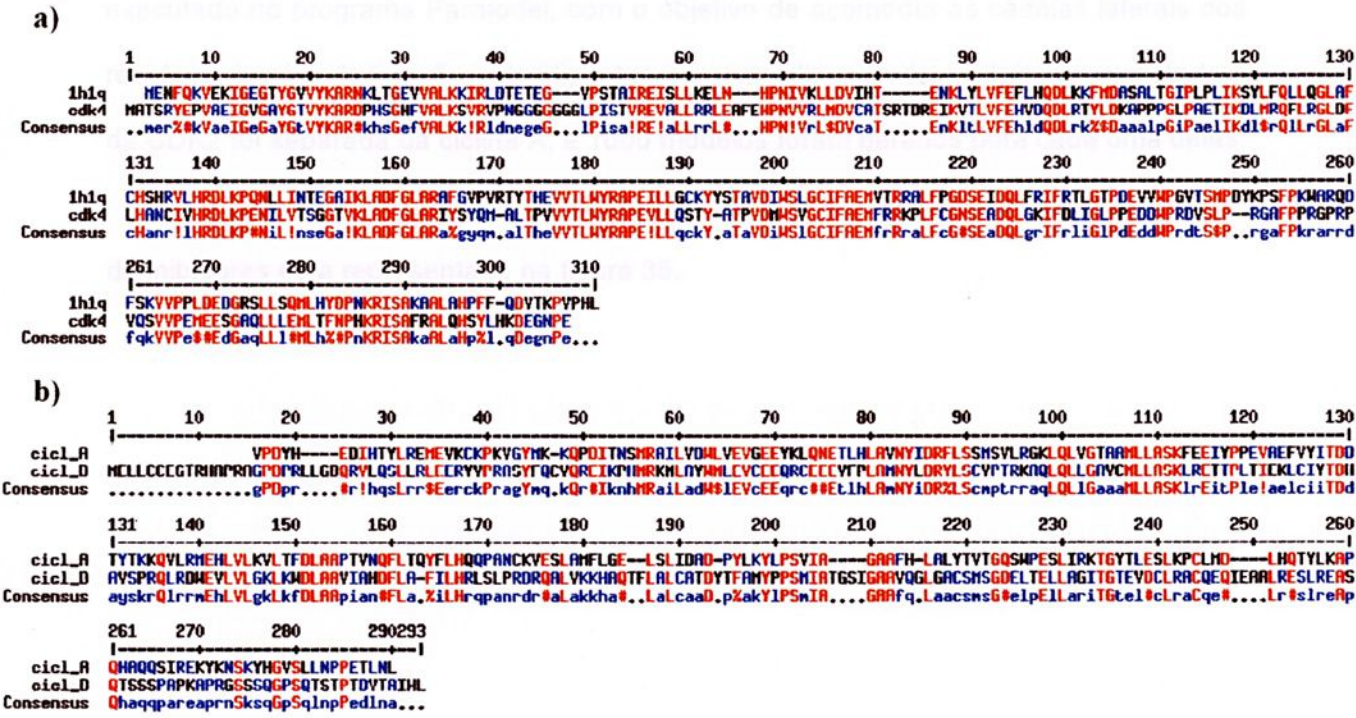


Figura 33: Alinhamentos de seqüências primárias: a) alinhamento da CDK2 (1HQ1) com a CDK4; b) alinhamento da Ciclina A (1HQ1) com a Ciclina D (MultAlin, CORPET, 1988))

Com as informações obtidas pelo alinhamento de seqüências primárias, dá-se seguimento ao processo de modelagem molecular comparativa.

Os moldes necessários para realização das modelagens por homologia foram obtidos após a sobreposição da CDK2/ciclina A com as estruturas complexadas com os inibidores escolhidos (figura 34) (tabela 3).

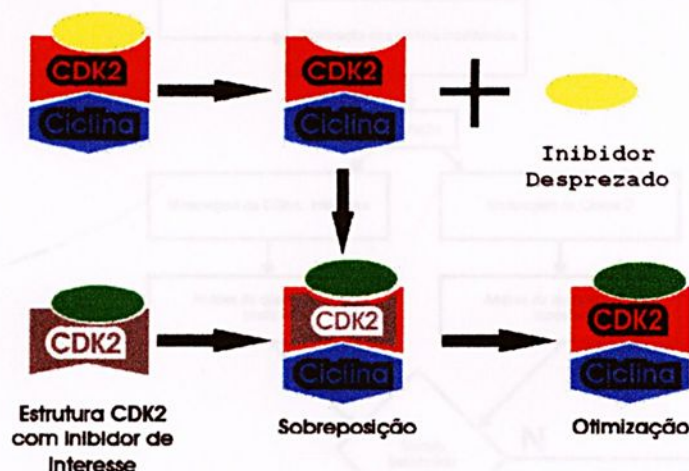


Figura 34: Esquema mostrando os passos para obter os moldes

Após a sobreposição, os moldes foram submetidos a um passo de otimização, executado no programa Parmodel, com o objetivo de acomodar as cadeias laterais dos resíduos do sítio de ligação e dos ligantes. No procedimento de modelagem, a estrutura da CDK2 foi separada da ciclina A, e 1000 modelos foram gerados para cada uma delas.

O protocolo utilizado para a modelagem do complexo CDK4/ciclinaD na presença de inibidores está representado na figura 35.

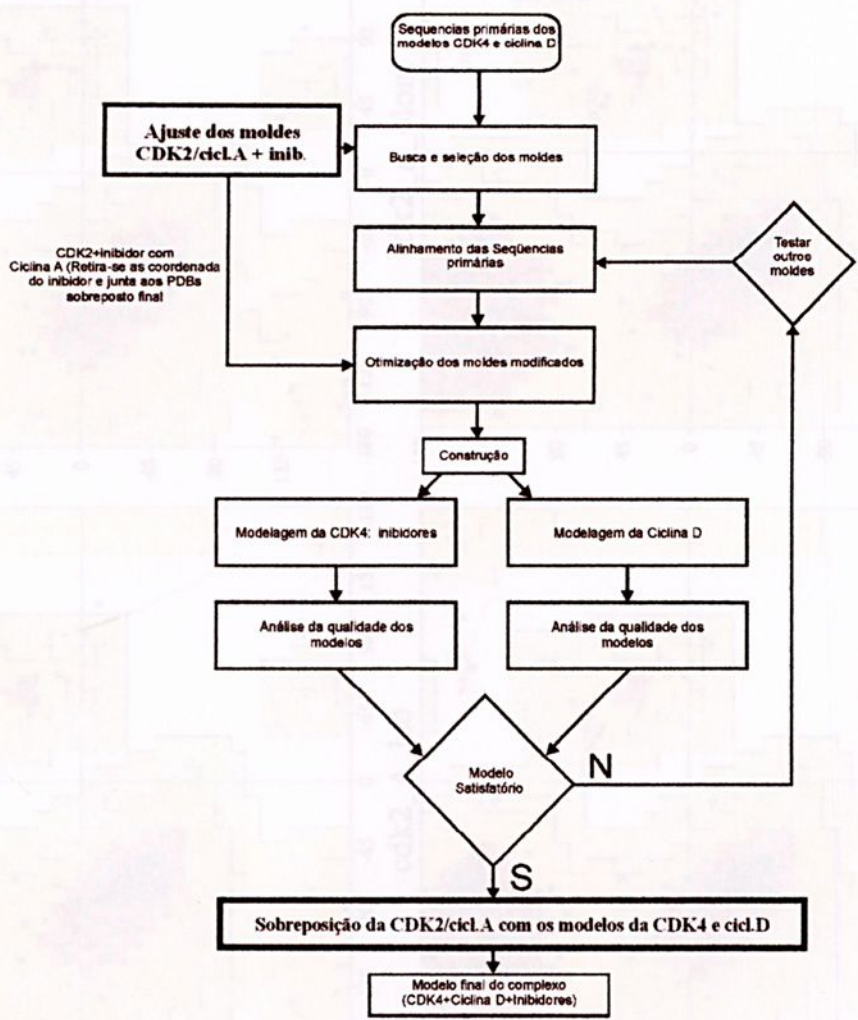


Figura 35: Protocolo da Modelagem molecular comparativa utilizado para modelar a CDK4/ciclina D em complexo com inibidores

4.1.2 QUALIDADE DOS MODELOS

A análise da qualidade estereoquímica dos modelos finais da CDK4/ciclina D complexada com inibidores foram comparados aos valores obtidos para o molde CDK2/ciclina A (Figuras 36 e 37), levando em consideração o gráfico de Ramachandran onde foram observados que a maioria dos resíduos (de preferência acima de 80%) encontravam-se nas regiões permitidas e o mínimo possível dos resíduos (de preferência abaixo de 1%, ou nenhum resíduo) encontravam-se na região proibida.

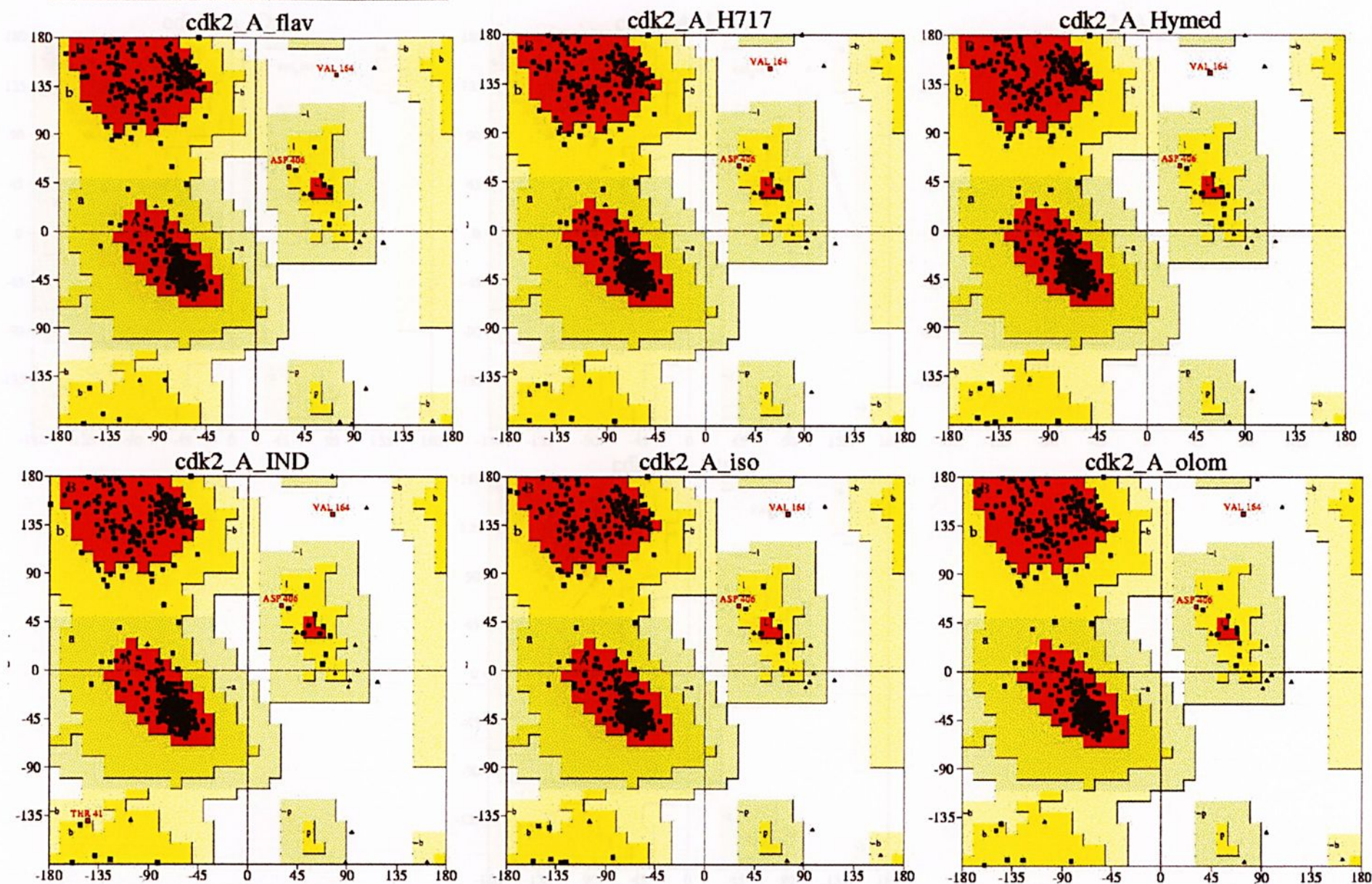


Figura 36: Gráficos de Ramachandran das 10 estruturas utilizadas como moldes. Regiões em vermelho são permitidas, em amarelo são adicionalmente permitidas e em amarelo claro são generosamente permitidas e em branco são regiões proibidas. Os triângulos são resíduos de glicina.

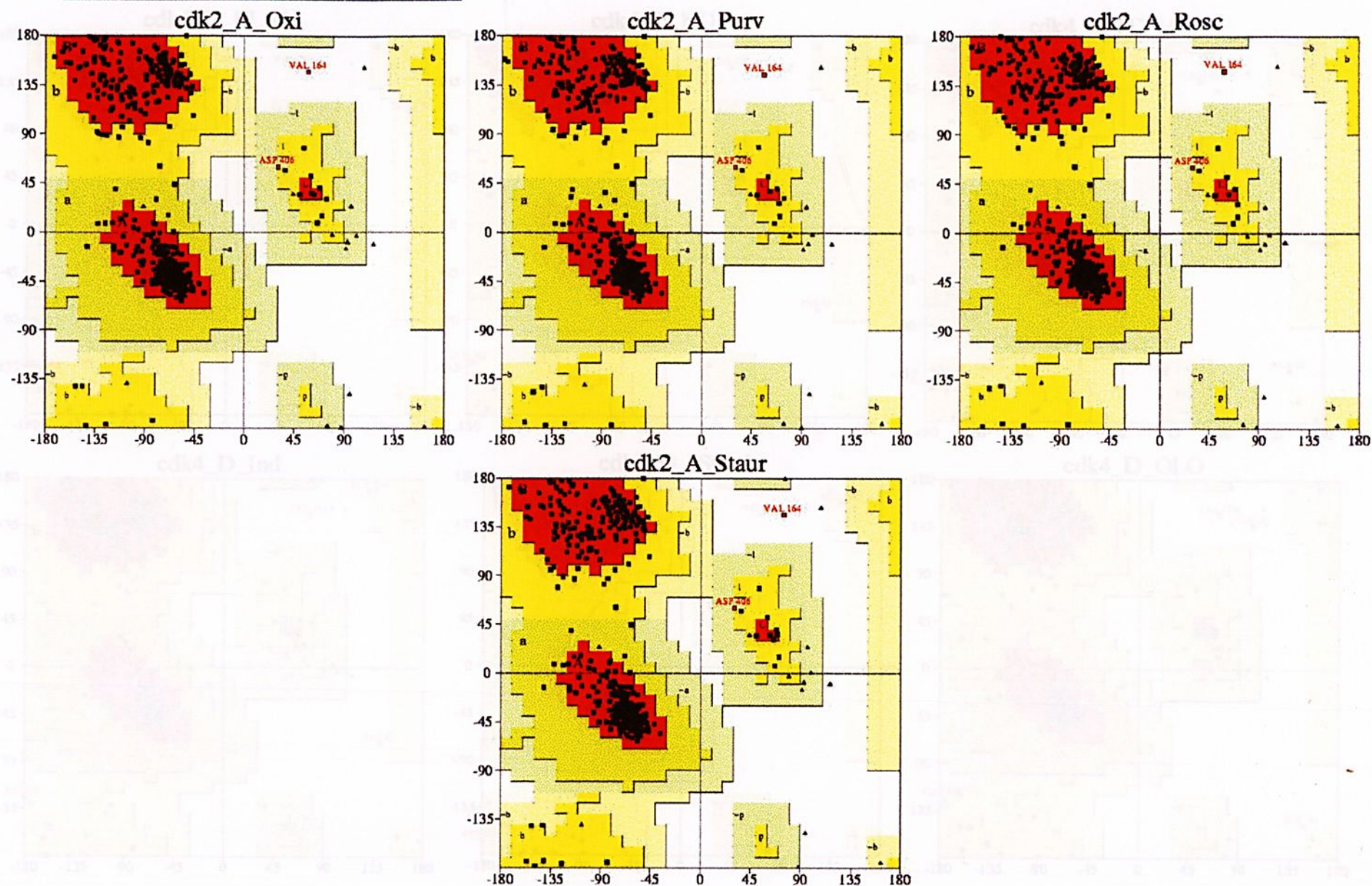


Figura 36: Gráficos de Ramachandran das 10 estruturas utilizadas como moldes. Regiões em vermelho são permitidas, em amarelo são adicionalmente permitidas e em amarelo claro são generosamente permitidas e em branco são regiões proibidas. Os triângulos são resíduos de glicina.

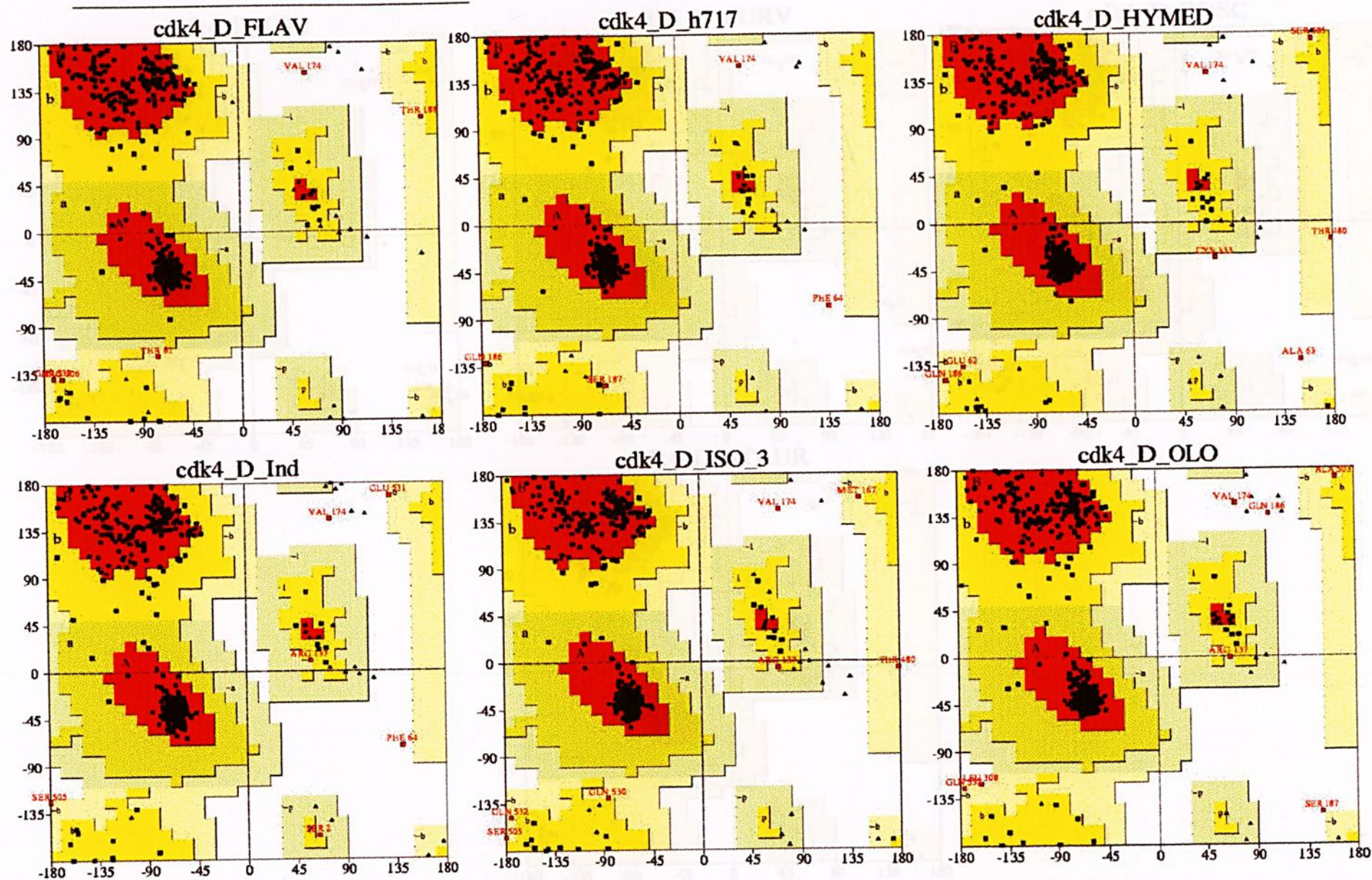


Figura 371: Gráficos de Ramachandran das 10 estruturas utilizadas como moldes. Regiões em vermelho são permitidas, em amarelo são adicionalmente permitidas e em amarelo claro são generosamente permitidas e em branco são regiões proibidas. Os triângulos são resíduos de glicina.

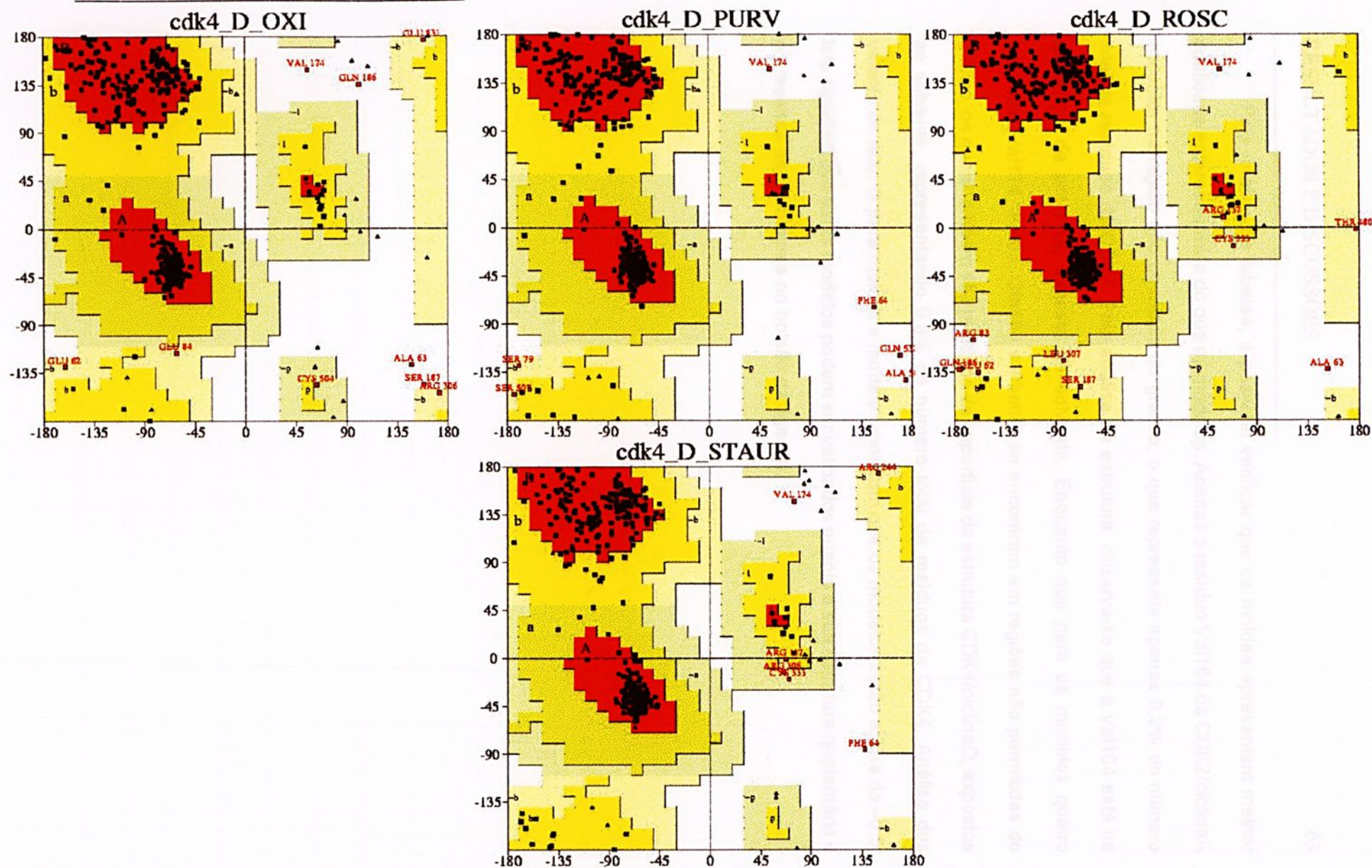


Figura 37: Gráficos de Ramachandran dos 10 melhores modelos de CDK4/cilinaD complexados com os inibidores. Regiões em vermelho são permitidas, em amarelo são adicionalmente permitidas, em amarelo claro são generosamente permitidas e em branco são regiões proibidas. Os triângulos são resíduos de glicina.

Pela análise da tabela 4, é possível verificar que os moldes apresentam melhor qualidade estereoquímica do que os modelos. Apenas o resíduo Val164 da CDK2/ciclinaA encontra na região não permitida do gráfico, o que representa apenas 0,2% do número total de resíduos da CDK2. Pela análise da estrutura, observa-se que a Val164 está na superfície da proteína, exposta ao solvente. Enquanto que para os modelos quatro resíduos (Val174, Ala63, Gln186 e Phe64) se encontram em regiões não permitidas do gráfico, os quais também são resíduos de superfície da estrutura CDK4/ciclinaD, expostos ao solvente, representando 0,8% do número total de resíduos da CDK4. Análise dos valores do fator G para moldes e modelos mostram que os modelos estão acima de $-0,5$. Isto significa que estes modelos podem ser validados quanto a sua estrutura quaternária e a presença dos inibidores no bolsão de ligação.

Observação	91,3 (92,3)	7,3 (7,1)	0,8 (0,2)	0,8 (0,2)	0,04 (-0,02)	-0,29 (-0,75)
Orindolol 91	91,5 (92,7)	6,7 (6,9)	1,0 (0,2)	0,8 (0,2)	-0,01 (-0,06)	-0,36 (-0,90)
Parvosolol B	92,5 (92,1)	6,3 (7,3)	0,4 (0,2)	0,8 (0,2)	0,03 (-0,03)	-0,31 (-0,74)
Besornitina	91,7 (92,7)	6,5 (6,9)	1,1 (0,2)	0,6 (0,2)	0,03 (-0,04)	-0,31 (-0,73)
Tapacresporina	90,5 (92,7)	8,3 (6,9)	0,8 (0,2)	0,4 (0,2)	0,07 (-0,05)	-0,33 (-0,75)

Os valores em parênteses são dos moldes

Os valores de r.m.s.d para o comprimento da ligação, ângulo de ligação, ângulos diédricos e angulários foram $0,61 \text{ \AA}$; $3,52^\circ$, $22,88^\circ$ e $2,67^\circ$, respectivamente, para o molde e $0,80 \text{ \AA}$; $2,71^\circ$; $22,40^\circ$ e $2,63^\circ$, respectivamente, para os modelos. A comparação dos valores entre molde e modelos indica que os complexos CDK4/ciclinaD-inibidores são bons modelos e podem ser considerados para estudos futuros.

Os valores obtidos a partir do programa proSA 3D para os moldes foram: Escore geral = $-250,00$; Escore Local = $-254,09$ e para a qualidade da estrutura = $0,98$, enquanto

Tabela 4: Análises da qualidade estereoquímica e a média dos valores do Fator G para os moldes e modelos

Proteínas	Gráfico de Ramachandran				Fator G	
	Região Permitida %	Região Adicionalmente Permitida %	Região Generosamente Permitida %	Região Proibida %	Ângulos de torção	Forças Covalentes
CDK4_ciclinaD (CDK2_ciclinaA)						
Flavopiridol	91,9 (92,9)	7,1 (6,7)	0,8 (0,2)	0,2 (0,2)	0,05 (-0,04)	-0,29 (-0,74)
H717	91,5 (92,3)	7,7 (7,3)	0,4 (0,2)	0,4 (0,2)	0,04 (-0,04)	-0,29 (-0,73)
Himentialdisina	90,1 (92,1)	8,5 (7,5)	0,8 (0,2)	0,8 (0,2)	0,05 (-0,04)	-0,30 (-0,74)
Indirubina_5_sulfonato	90,5 (92,3)	8,3 (7,2)	0,8 (0,4)	0,4 (0,2)	0,04 (-0,04)	-0,32 (-0,74)
Isopropeniladenina	91,3 (91,9)	7,3 (7,7)	1,2 (0,2)	0,2 (0,2)	0,07 (-0,02)	(-0,25) (-0,71)
Olomoucina	91,3 (92,5)	7,3 (7,1)	0,8 (0,2)	0,8 (0,2)	0,04 (-0,02)	-0,29 (-0,75)
Oxindol 91	91,5 (92,7)	6,7 (6,9)	1,0 (0,2)	0,8 (0,2)	0,01 (-0,06)	-0,36 (-0,90)
Purvanalol B	92,5 (92,1)	6,3 (7,5)	0,4 (0,2)	0,8 (0,2)	0,03 (-0,03)	-0,31 (-0,74)
Roscovitina	91,7 (92,7)	6,3 (6,9)	1,4 (0,2)	0,6 (0,2)	0,03 (-0,04)	-0,31 (-0,73)
Estaurosporina	90,5 (92,7)	8,3 (6,9)	0,8 (0,2)	0,4 (0,2)	0,02 (-0,05)	-0,33 (-0,75)

Os valores em parênteses são dos moldes

Os valores de r.m.s.d para o comprimento de ligação, ângulo de ligação, ângulos diédricos e impróprios foram 0,61 Å; 3,52°; 22,88° e 2,67°, respectivamente, para o molde e 0,60 Å; 2,71°; 22,40° e 2,63°, respectivamente, para os modelos. A comparação dos valores entre moldes e modelos indica que os complexos CDK4/ciclinaD-inibidores são bons modelos e podem ser considerados para estudos futuros.

Os valores obtidos a partir do programa perfil 3D para os moldes foram: Escore total = 250,00; Escore ideal = 254,09 e para a qualidade da estrutura = 0,98, enquanto

que para os modelos foram obtidos os seguintes valores: Escore total = 242,13; Escore ideal = 263,80 e para a qualidade dos modelos = 0,92. O valor da qualidade é obtido através da divisão do Escore total pelo Escore ideal, sendo que seu valor ideal estabelecido pelo método para uma boa qualidade e confiabilidade deve estar acima de 0,45S.

Ambos moldes e modelos apresentam valores dentro do limite estabelecido pelo método.

4.1.3. DESCRIÇÃO DOS MODELOS

O modelo da CDK4/ciclinaD, em complexo com inibidores, apresentou uma estrutura similar à estrutura da CDK2/ciclinaA. Formada por dois lóbulos principais unidos por uma região de ligação, sendo que o sítio onde se liga a molécula do inibidor se encontra em uma fenda entre os dois lóbulos.

A região N-terminal é composta de um conjunto de cinco folhas β dispostas de forma anti-paralela (β 1- β 5) e uma única e longa hélice α (α 1). Já a região C-terminal é composta de um empacotamento pseudo-4-helicoidal (α 2,3,4,6), uma pequena folha β (β 6- β 8) e duas hélices adicionais (α 5,7).

Estas características estruturais são descritas para os monômeros CDK2 e CDK4. As ciclinas são compostas inteiramente de hélices α .

As figuras 38a e 39a mostram o diagrama de fitas do complexo CDK2/ciclina A e CDK4/ciclina D, ambos com Flavopiridol, e as figuras 38b e 39b, enfocam a região do sítio de ligação com os 10 ligantes.

As menores diferenças observadas entre as estruturas da CDK2 e CDK4 estão no N-terminal (resíduos 1 e 2) e C-terminal (resíduos 301 a 303) da CDK4. A seqüência primária da CDK4 é maior do que a CDK2 (303 contra 298 resíduos). As maiores diferenças foram observadas nas alças 46 a 48, 65 a 67 e 81 a 85 da CDK4, que podem ser observadas no alinhamento das seqüências primárias (figura 40a).



4.1.4. INTERAÇÃO CDK E CICLINA

A ativação da CDK2 pela fosforilação e associação com a ciclina A apresentam diferenças conformacionais que podem ser igualmente observadas para os modelos da CDK4/ciclinaD.

No alinhamento de sequência primária entre as ciclinas foram observadas principalmente diferenças no N-terminal (resíduos 1 a 15), que foi excluído para a modelagem por não haver molde disponível. Estudos posteriores poderão contribuir para a obtenção de resultados referentes a esta região e as três alças (21 a 24; 204 a 207 e 247 a 250) da ciclina D (figura 40b), as quais também não possuem moldes disponíveis.

No complexo CDK4/ciclina D foi notado que o lóbulo N-terminal é constituído principalmente de um conjunto de folhas β pregueadas com uma única hélice, e a hélice C ou hélice PISTVRE, localizada entre os resíduos 1 a 90 (na estrutura da CDK2/ciclinaA compreendem os resíduos 1 a 79). Ambas as estruturas apresentam uma alça rica em Glicinas localizado entre as fitas β ($\beta 1$ e $\beta 2$).

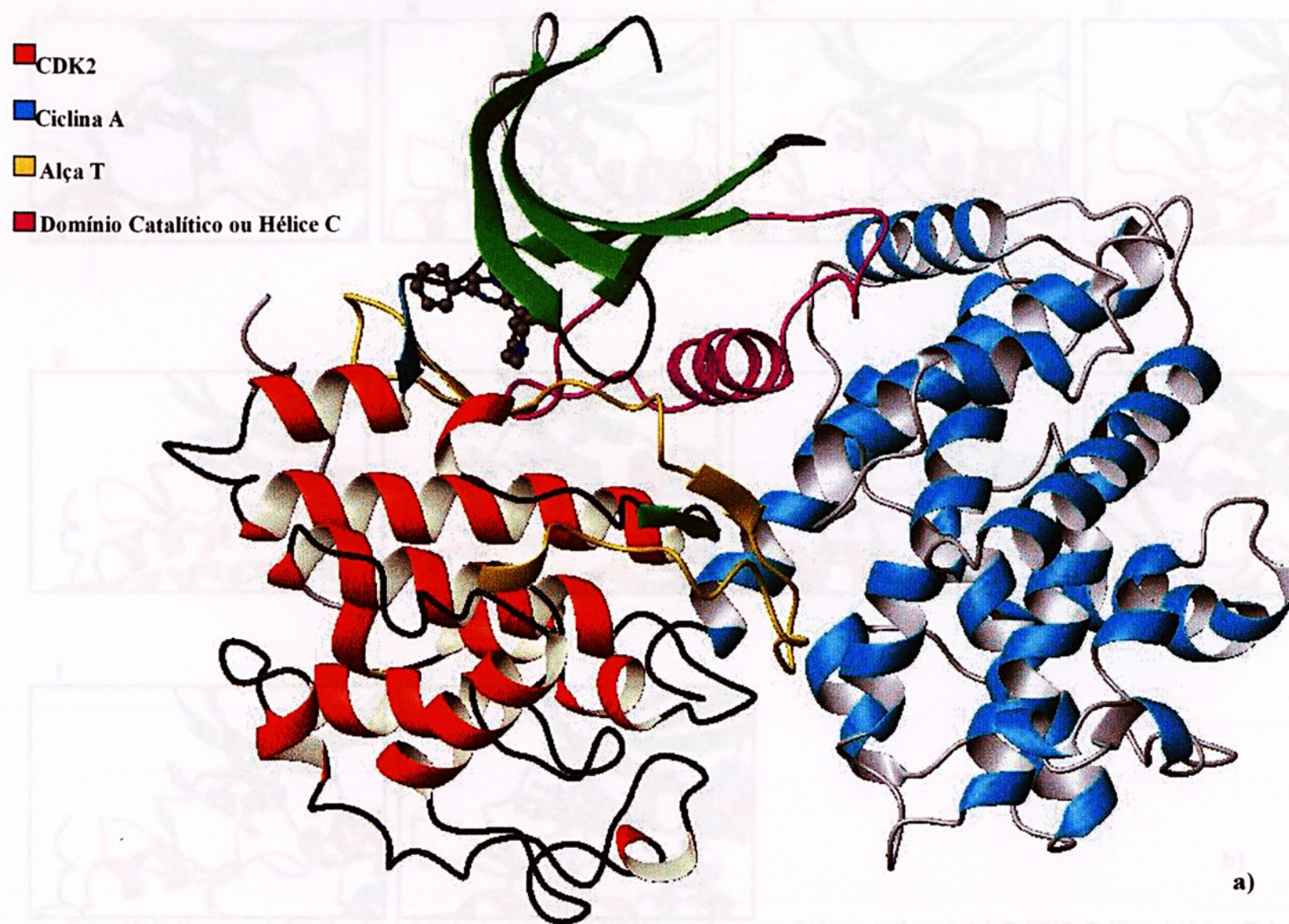


Figura 38: a) Diagrama de fitas para CDK2/Ciclina A complexada com Flavopiridol mostrando a região do sítio de ligação.



b)

Figura38: b) Diagrama de fitas do sítio de ligação do complexo CDK2/ciclina A e seus inibidores: **A.** flavopiridol; **B.** H717; **C.** Himenialdisina; **D.** Indirubina; **E.** Isopenteniladenina; **F.** olomoucina; **G.** oxindol; **H.** purvanalol; **I.** roscovitina e **J.** estaurosporina

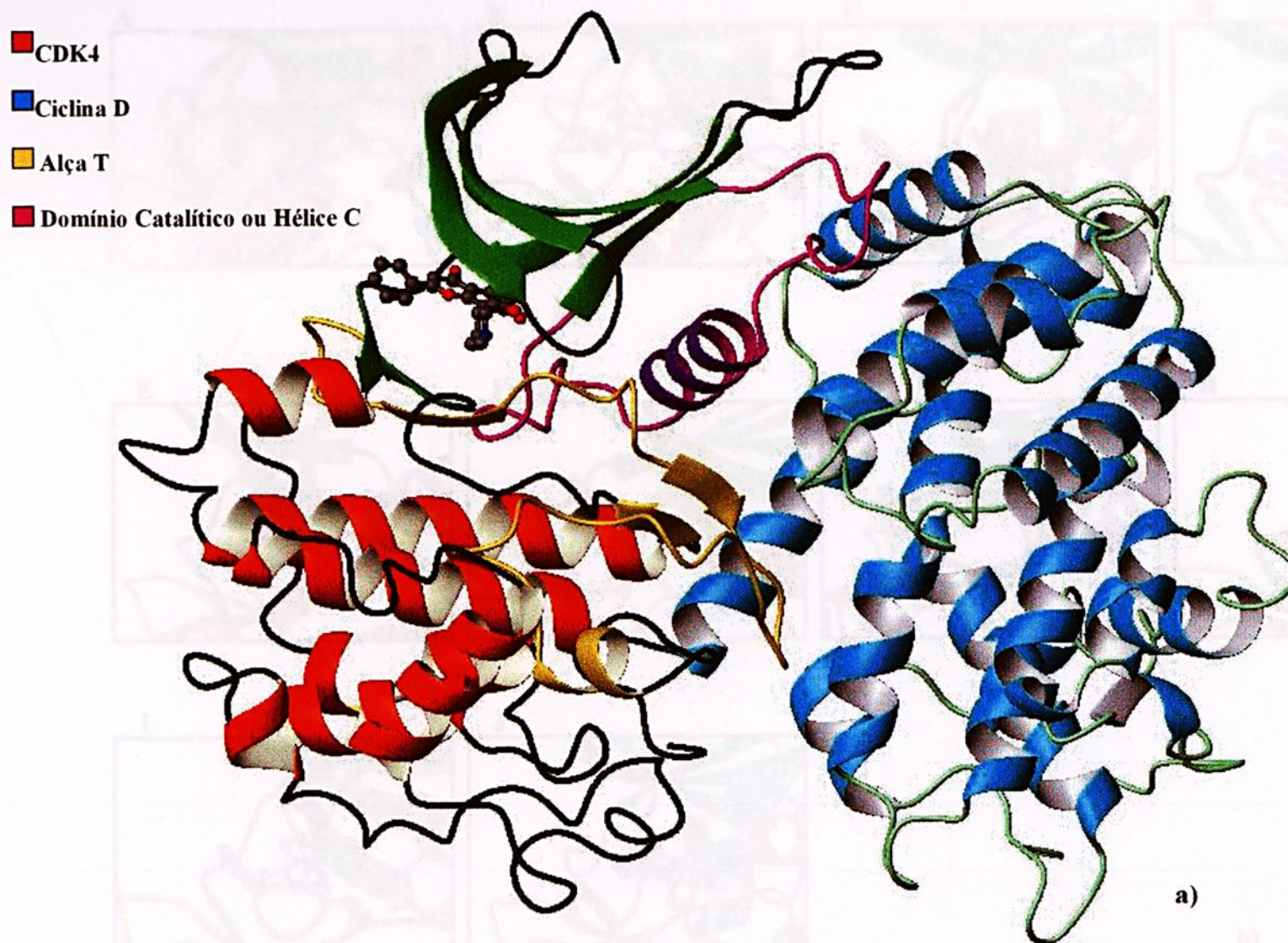


Figura 39: a) Diagrama de fitas para CDK4/ciclina D complexada com Flavopiridol mostrando a região do sítio de ligação.

Figura 39: b) Diagrama de fitas do sítio de ligação do complexo CDK4/ciclina D e seus inibidores: A, Flavopiridol; B, H717; C, Roscovitina; D, Indinavir; E, Ispinesib; F, olaparimetabólito; G, erastol; H, puricinas; I, resorcinóis e J, estrogênios.



Figura39: b) Diagrama de fitas do sitio de ligação do complexo CDK4/ciclina D e seus inibidores: **A.** flavopiridol; **B.** H717; **C.** Himenialdisina; **D.** Indirubina; **E.** Isopenteniladenina; **F.** olomoucina; **G.** oxindol; **H.** purvalanol; **I.** roscovitina e **J.** estaurosporina

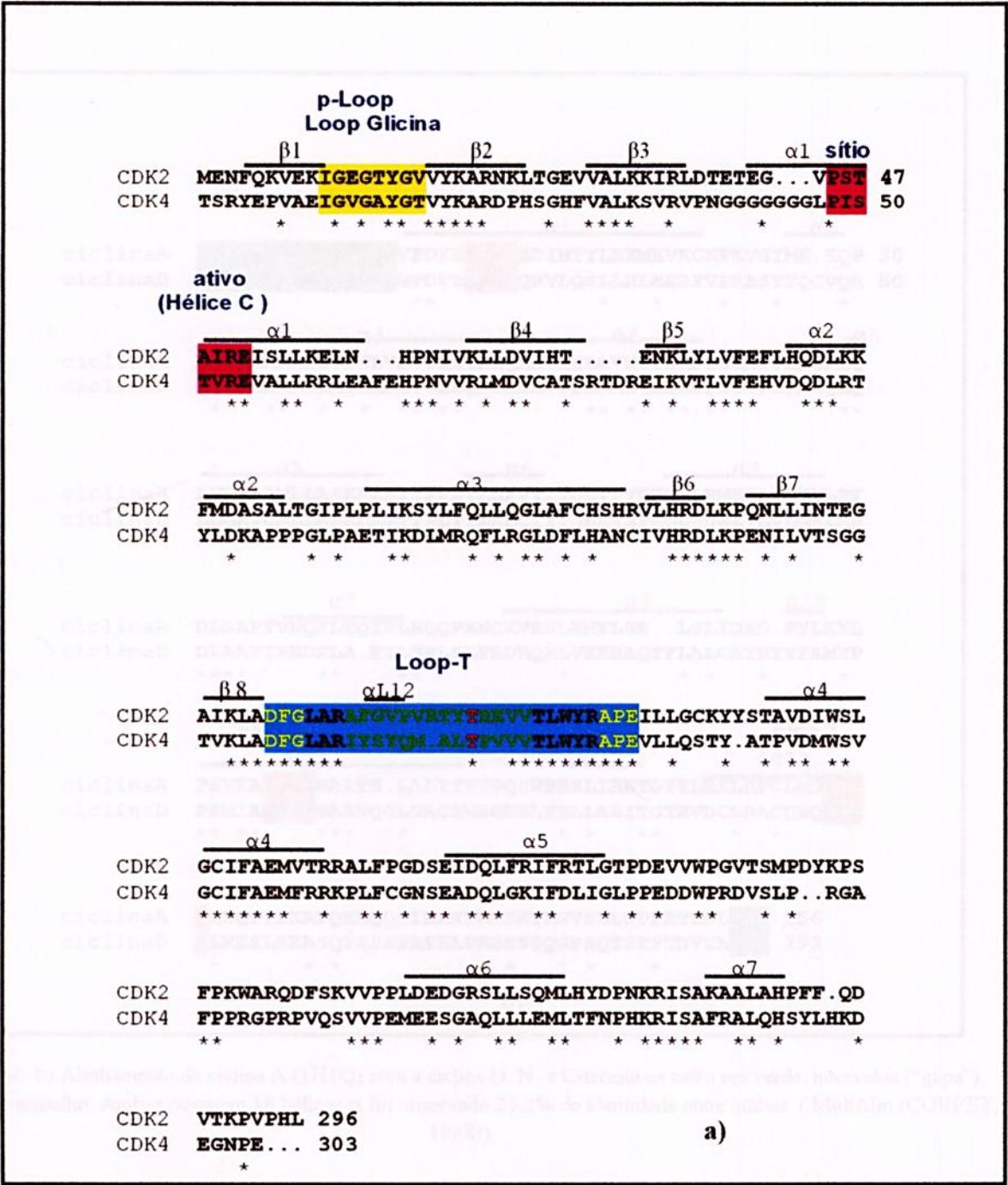
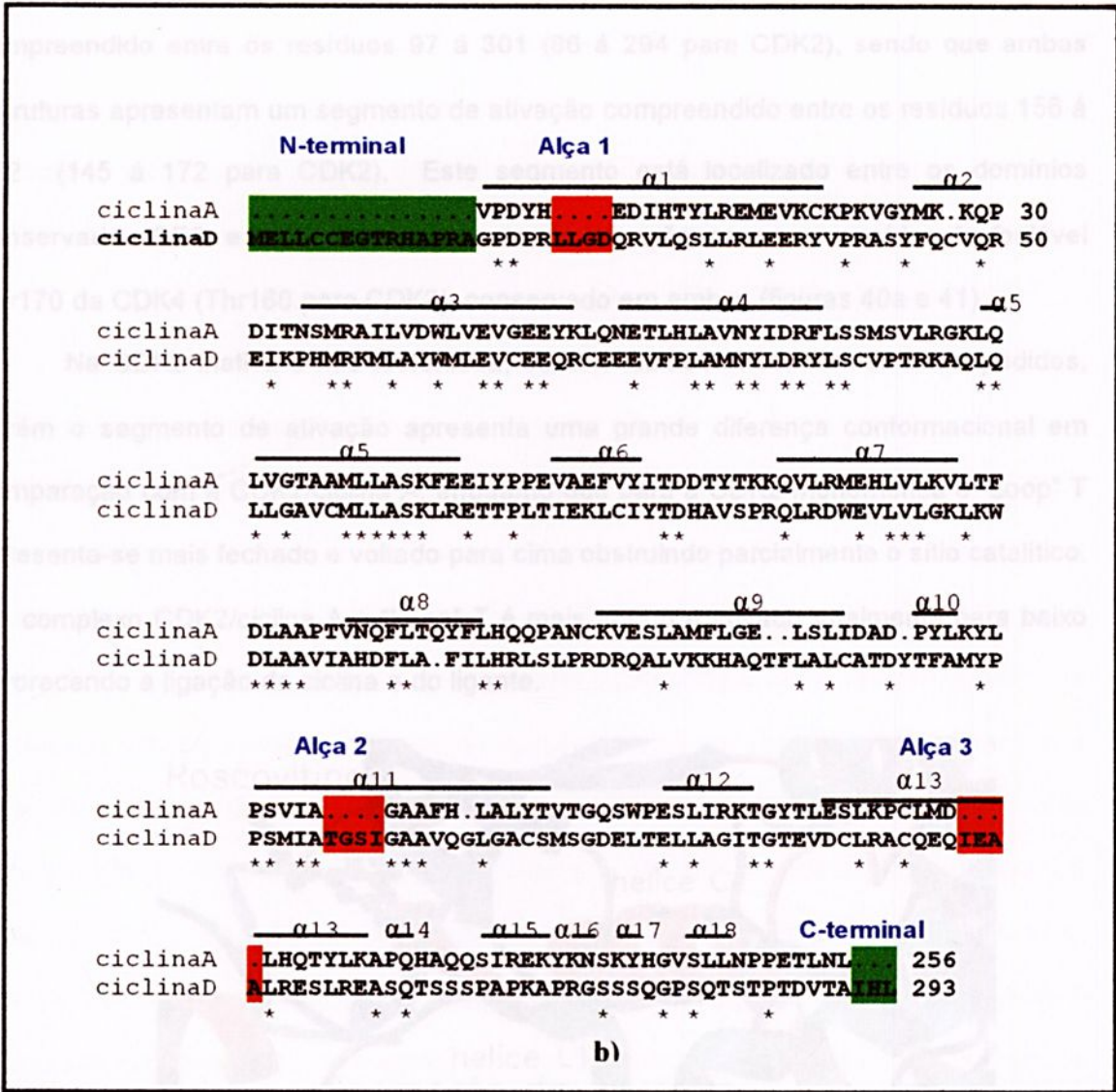


Figura40: a) Alinhamento do molde (1H1Q) com o modelo (CDK4) indicando em cores as regiões: Loop rico em glicinas (amarelo), o segmento de ativação (azul claro), o sítio de ligação, hélice C (rosa), Loop_T (resíduos em verde) indicando a Treonina fosforilável (resíduo em vermelho), pequenos domínios conservados (resíduos em amarelo) e as indicações das regiões onde se encontram as hélices_ α e fitas_ β (Traço preto), é observado 45,3% de identidade entre ambas (MultAlin (CORPET, 1988)).



b)

Figura40: b) Alinhamento da ciclina A (1HIQ) com a ciclina D. N- e C-terminais estão em verde, intervalos ("gaps") estão em vermelho. Ambas possuem 18 hélices α foi observado 29,2% de identidade entre ambas (MultAlin (CORPET, 1988)).

O lóbulo N-terminal é unido ao lóbulo C-terminal pelos resíduos 91 á 96 (80 á 85 para CDK2). O lóbulo C-terminal é predominantemente formado de hélices α compreendido entre os resíduos 97 á 301 (86 á 294 para CDK2), sendo que ambas estruturas apresentam um segmento de ativação compreendido entre os resíduos 156 á 182 (145 á 172 para CDK2). Este segmento está localizado entre os domínios conservados DFG e APE. Este segmento de ativação carrega o resíduo fosforilável Thr170 da CDK4 (Thr160 para CDK2), conservado em ambas (figuras 40a e 41).

Na CDK2 inativa e não fosforilada, os dois lóbulos encontram-se desimpedidos, porém o segmento de ativação apresenta uma grande diferença conformacional em comparação com a CDK2/ciclina A, enquanto que para a CDK2 monomérica o "Loop" T apresenta-se mais fechado e voltado para cima obstruindo parcialmente o sítio catalítico. No complexo CDK2/ciclina A o "Loop" T é mais aberto e voltado totalmente para baixo favorecendo a ligação da ciclina e do ligante.

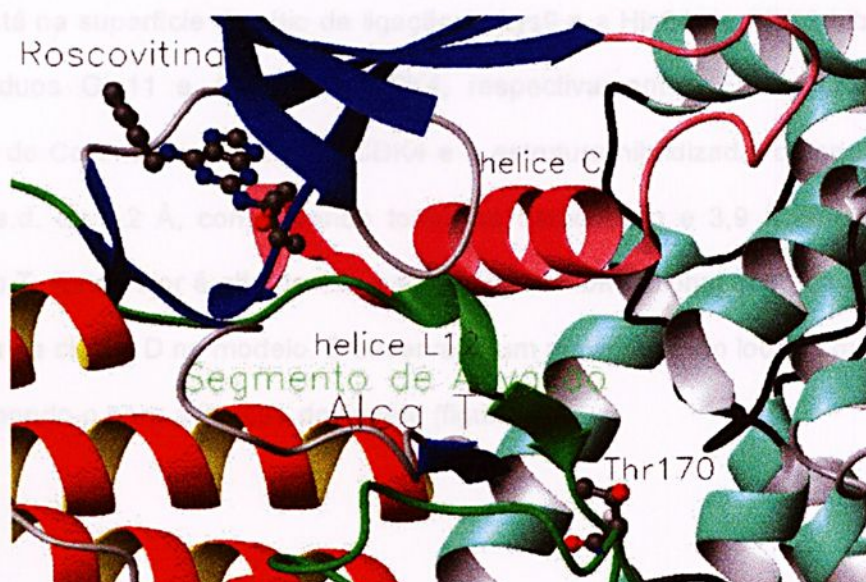


Figura 41: Região do bolsão de ligação da CDK4 mostrando a interface com a ciclina D. Estão em evidência a hélice C (em rosa), a alça (loop) T (em verde), mostrando a Thr170, o segmento de ativação (em verde), a hélice L12 e o inibidor Roscovitina

A fosforilação da Thr160 na CDK2 monomérica resulta no deslocamento e algum desordenamento do segmento de ativação, sem maiores mudanças na estrutura.

Sobre a associação da CDK4 com ciclina D, a hélice C e ciclina D rotacionam e mudam de posição para fazer maior contato com o lóbulo N-terminal e posiciona os resíduos chaves para a ligação do ATP, os resíduos Lys33 e Glu54, conservados em ambas as estruturas (Glu51 em CDK2). As CDKs são flexíveis, porém as ciclinas são rígidas. O componente final do processo de ativação é a Thr170 (Thr160 em CDK2).

ANDERSON, et al. (2003) sugeriram uma estratégia de hibridização entre CDK2 e CDK4. Estes autores substituíram os resíduos do sítio de ligação da CDK2 por resíduos da CDK4, baseado em estudos funcionais com uma classe de inibidores seletivos. Esta estrutura hibridizada foi usada como comparação com o sítio ativo do modelo da CDK4 na presença de ciclina D.

A maior diferença observada no modelo CDK4/ciclina D em relação à estrutura hibridizada está na superfície do sítio de ligação (a Lys9 e a His84 da CDK2 hibridizada para os resíduos Glu11 e Asp97 na CDK4, respectivamente). Foi realizado uma sobreposição de C α entre o modelo da CDK4 e a estrutura hibridizada, obtendo-se um valor de r.m.s.d. de 3,2 Å, considerando todos os carbonos α e 3,9 Å considerando apenas o loop T. Este valor é alto devido à mudança conformacional ocorrida no loop T pela presença da ciclina D no modelo. É observado um movimento do loop T, que abre o sítio ativo, expondo-o para a ligação do ligante (figura 42).

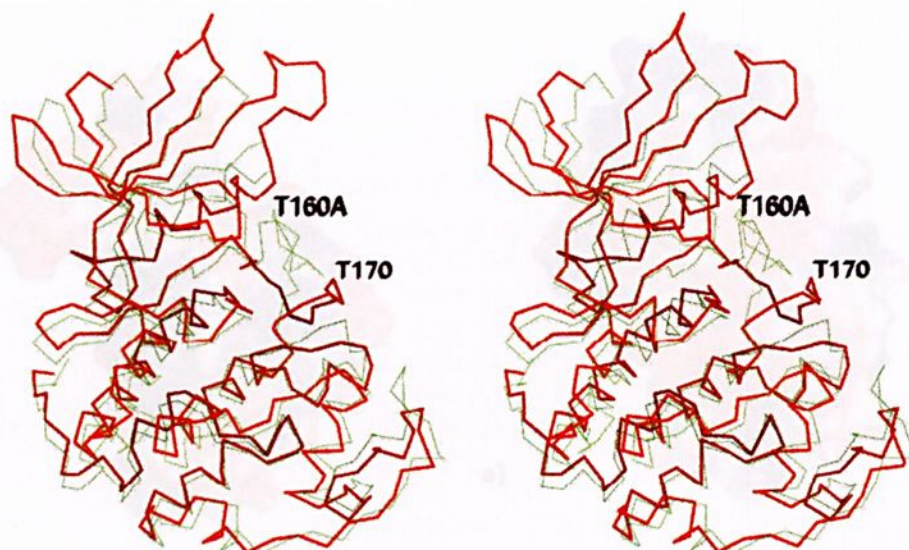


Figura 426: Sobreposição dos Carbonos α do modelo da CDK4 (linha mais grossa em vermelha) com a CDK2 hibridizada (linha mais fina em verde), mostrando a posição da Thr fosforilável

As superfícies de contato foram calculadas para as estruturas CDK2/ciclina A e CDK4/ciclinaD, procedimento esse executado com os programas AREAIMOL e RESAREA (CCP4, 1994). Os valores obtidos para molde e modelo foram $1654 \text{ \AA}^2$ e $1454 \text{ \AA}^2$, respectivamente. A área de contato da CDK4/ciclinaD é um pouco menor, provavelmente devido à mudança conformacional do loop T, já que este se abre, expondo o sítio catalítico. A superfície de contato e o potencial eletrostático da superfície das proteínas foram analisados e observado uma complementaridade de carga e forma entre CDK e ciclina (figura 43).

4.1.5. INTERAÇÃO CDK INIBIDORES

Os níveis de especificidade e afinidade entre uma enzima e seu inibidor dependem das interações iônicas, ligações de hidrogênio deslocais e da complementaridade das suas superfícies de contato (GANDURI et al., 2004; DE AZEVEDO et al., 2002; KIM et al., 1990).

A tabela 5 relaciona as interações por ligações de hidrogênio entre a CDK4 e os 10 inibidores estudados, comparando as interações observadas para o molde

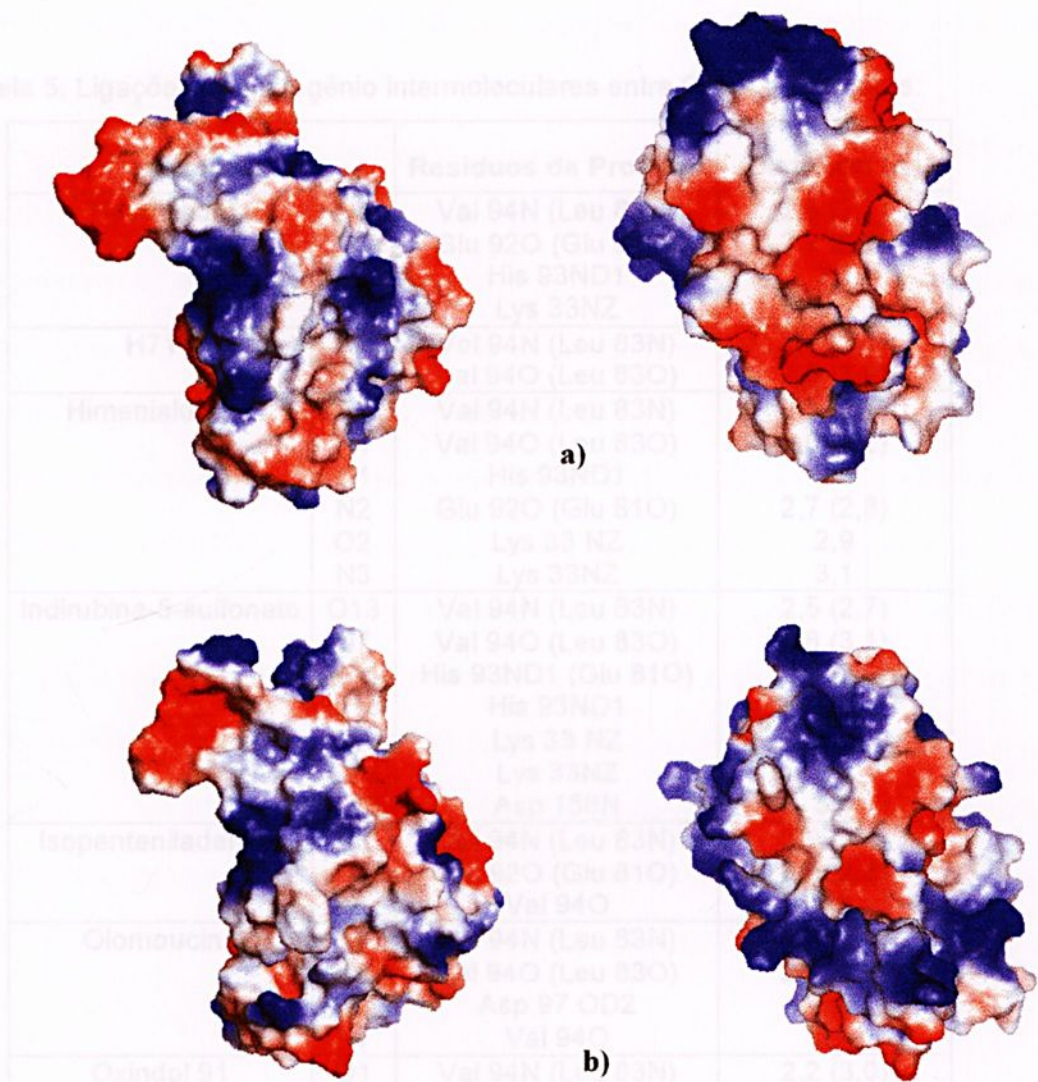


Figura 43: Imagens das regiões de interface entre CDK e Ciclina; a) CDK2 e Ciclina A; b) CDK4 e Ciclina D (Imagens geradas no Programa Pymol (DELANO, 1998-2003)).

4.1.5. INTERAÇÃO CDK_INIBIDORES

Os níveis de especificidade e afinidade entre uma enzima e seu inibidor dependem de interações iônicas, ligações de hidrogênio direcionais e da complementaridade das suas superfícies de contato (CANDURI et al., 2004; DE AZEVEDO et al., 2002; KIM et al.,1996).

A tabela 5 relaciona as interações por ligações de hidrogênio entre a CDK4 e os 10 inibidores estudados, comparados às interações observadas para o molde.

Tabela 5: Ligações de hidrogênio intermoleculares entre CDK4 e inibidores.

Inibidores		Resíduos da Proteína	Distância (Å)
Flavopiridol	O4	Val 94N (Leu 83N)	2,6 (2,7)
	O5	Glu 92O (Glu 81O)	2,9 (2,9)
	O4	His 93ND1	3,1
	O7	Lys 33NZ	3,5
H717	N7	Val 94N (Leu 83N)	2,6 (3,1)
	N18	Val 94O (Leu 83O)	2,5 (3,0)
Himenialdisina	O1	Val 94N (Leu 83N)	2,5 (2,7)
	N1	Val 94O (Leu 83O)	3,0 (3,2)
	O1	His 93ND1	2,8
	N2	Glu 92O (Glu 81O)	2,7 (2,8)
	O2	Lys 33 NZ	2,9
	N3	Lys 33NZ	3,1
Indirubina-5-sulfonato	O13	Val 94N (Leu 83N)	2,5 (2,7)
	N7	Val 94O (Leu 83O)	2,8 (3,1)
	N14	His 93ND1 (Glu 81O)	2,9 (3,0)
	O13	His 93ND1	2,9
	O24	Lys 33 NZ	2,8
	O22	Lys 33NZ	2,8
	O23	Asp 156N	3,0
Isopenteniladenina	N3	Val 94N (Leu 83N)	2,9 (3,4)
	N9	Glu 92O (Glu 81O)	3,1 (2,8)
	N3	Val 94O	3,1
Olomoucina	N7	Val 94N (Leu 83N)	2,5 (2,9)
	N6	Val 94O (Leu 83O)	2,5 (2,6)
	OH	Asp 97 OD2	3,0
	N7	Val 94O	3,4
Oxindol 91	O1	Val 94N (Leu 83N)	2,2 (3,0)
	N15	Val 94O (Leu 83O)	2,7 (3,1)
	N3	Glu 92O (Glu 81O)	3,1 (2,9)
Purvanalol B	N7	Val 94N (Leu 83N)	2,6 (3,2)
	N6	Val 94O (Leu 83O)	2,5 (2,5)
Roscovitina	N7	Val 94N (Leu 83N)	2,8 (3,4)
	N6	Val 94O (Leu 83O)	2,6 (2,8)
Estaurosporina	O5	Val 94N (Leu 83N)	2,7 (2,8)
	N1	Glu 92O (Glu 81O)	2,9 (2,7)
	O5	His 93ND1	2,9
	O5	Val 94O	2,4
	O6	Glu 142O	3,1
	N4	Asp 97OD2	3,0

Os valores em parênteses para os moldes

O maior número de ligações de hidrogênio presentes na interação da CDK4 com inibidores foi observado para himenialdisina (6), indirubina-5-sulfonato (7), flavopiridol (4)



e estaurosporina (6), sendo que para a CDK2 as ligações de hidrogênio foram em média de 2 a 3 para cada inibidor.

A comparação dos sítios de ligação dos complexos CDK4-inibidores com CDK2-inibidores, e CDK4-inibidores com CDK2-ATP possibilitou observar possíveis alterações no posicionamento das cadeias laterais dos resíduos, causadas pela ligação dos inibidores. Essa comparação foi feita após a superposição dos Ca das estruturas realizada no programa LSQKAB (CCP4, 1994).

Com a sobreposição do modelo do complexo CDK4-inibidores com a estrutura da CDK2-inibidores, e CDK2-ATP também foi possível notar que os sistemas de anéis dos inibidores e do ATP se sobrepõem aproximadamente no mesmo plano, mas apresentando diferentes orientações. Exemplo disso pode ser verificado na estrutura cristalográfica da CDK2 em complexo com Roscovitina (DE AZEVEDO et al., 1997). Para CDK4, o anel fenil da molécula dos inibidores aponta para fora do sítio de ligação de ATP e se encontra parcialmente exposto ao solvente. As figuras 44 a e b mostram a sobreposição do sítio de ligação da CDK4-Flavopiridol com CDK2-Flavopiridol e da CDK4-Flavopiridol com CDK2-ATP.

Figura 44 Sobreposições de Ca dos sítios catalíticos a) CDK4-Flavopiridol (linha mais grossa azul) com CDK2-Flavopiridol (linha mais fina vermelha); b) CDK4-Flavopiridol (linha mais grossa azul) com CDK2-ATP (linha mais fina vermelha)

Esta superposição não revelou nenhum movimento significativo das cadeias laterais dos resíduos do sítio de ligação de ATP dos complexos indicando que estes inibidores se ajustam muito bem ao sítio de ligação. Pode ser observado pela sobreposição do ATP com o Flavopiridol, a diferença de orientação com relação aos anéis. O ATP, por ser maior, posiciona-se de maneira diferente, mantendo expostos os grupos fosfato.

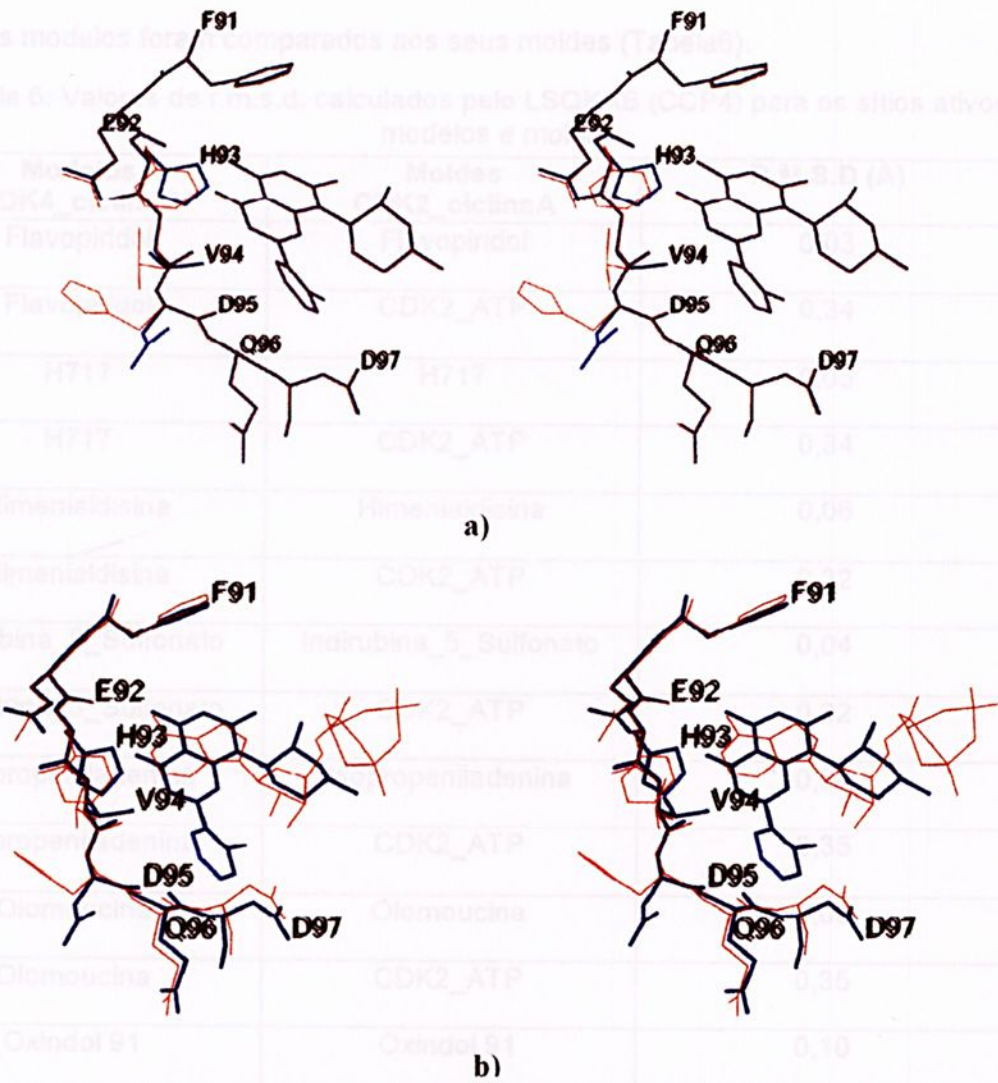


Figura 44: Sobreposições de C α dos sítios catalíticos a) CDK4-Flavopiridol (linha mais grossa azul) com CDK2-Flavopiridol (linha mais fina vermelha); b) CDK4-Flavopiridol (linha mais grossa azul) com CDK2-ATP (linha mais fina vermelha)

Esta superposição não revelou nenhum movimento significativo das cadeias laterais dos resíduos do sítio de ligação de ATP dos complexos indicando que estes inibidores se ajustam muito bem ao sítio de ligação. Pode ser observado pela sobreposição do ATP com o Flavopiridol, a diferença de orientação, com relação aos anéis. O ATP, por ser maior, posiciona-se de maneira diferente, mantendo exposto os grupos fosfato.

Os valores de r.m.s.d. para as sobreposições de Cα dos sítios de ligação para todos os modelos foram comparados aos seus moldes (Tabela6).

Tabela 6: Valores de r.m.s.d. calculados pelo LSQKAB (CCP4) para os sítios ativos dos modelos e moldes

Modelos CDK4_ciclinaD	Moldes CDK2_ciclinaA	R.M.S.D (Å)
Flavopiridol	Flavopiridol	0,03
Flavopiridol	CDK2_ATP	0,34
H717	H717	0,05
H717	CDK2_ATP	0,34
Himentialdisina	Himentialdisina	0,06
Himentialdisina	CDK2_ATP	0,32
Indirubina_5_Sulfonato	Indirubina_5_Sulfonato	0,04
Indirubina_5_Sulfonato	CDK2_ATP	0,32
Isopropeniladenina	Isopropeniladenina	0,04
Isopropeniladenina	CDK2_ATP	0,35
Olomoucina	Olomoucina	0,03
Olomoucina	CDK2_ATP	0,35
Oxindol 91	Oxindol 91	0,10
Oxindol 91	CDK2_ATP	0,45
Purvanalol B	Purvanalol B	0,06
Purvanalol B	CDK2_ATP	0,38
Roscovitina	Roscovitina	0,04
Roscovitina	CDK2_ATP	0,37
Estaurosporina	Estaurosporina	0,05
Estaurosporina	CDK2_ATP	0,35

Pela análise da tabela 6 é observado que os valores de r.m.s.d. para as sobreposições são muito baixos, como esperado, desde que os sítios de ligação das CDKs são conservados. Valores de r.m.s.d. do sítio na presença de ATP são, em média, da ordem de 10x maior. Isso se deve à diferença de orientação do anel de purina dos ligantes, resultando na acomodação do sítio de ligação.

A análise da área de contato entre a proteína e o ligante não mostrou uma relação direta com o valor de IC50 (tabela 7). Não há uma correlação clara entre área de contato e IC50, ou seja, analisando-se isoladamente a correlação entre área de contato e IC50 espera-se uma maior área de contato para inibidores que apresentam menor IC50. Obviamente tal análise é bem superficial, pois diversos outros fatores são determinantes para a especificidade de um inibidor para uma dada enzima, por exemplo, ligações de hidrogênio intermoleculares, complementaridade de carga e forma, interações salinas, forças de van der Waals e interações hidrofóbicas.

Rosigoxina	>100 (0,7)	334 (300)	2 (2)
Osimercina	>1000 (7)	281 (281)	4 (2)
Furvanalil E	>10 (0,006)	233 (242)	2 (2)

Os valores entre parênteses são dos moléculas

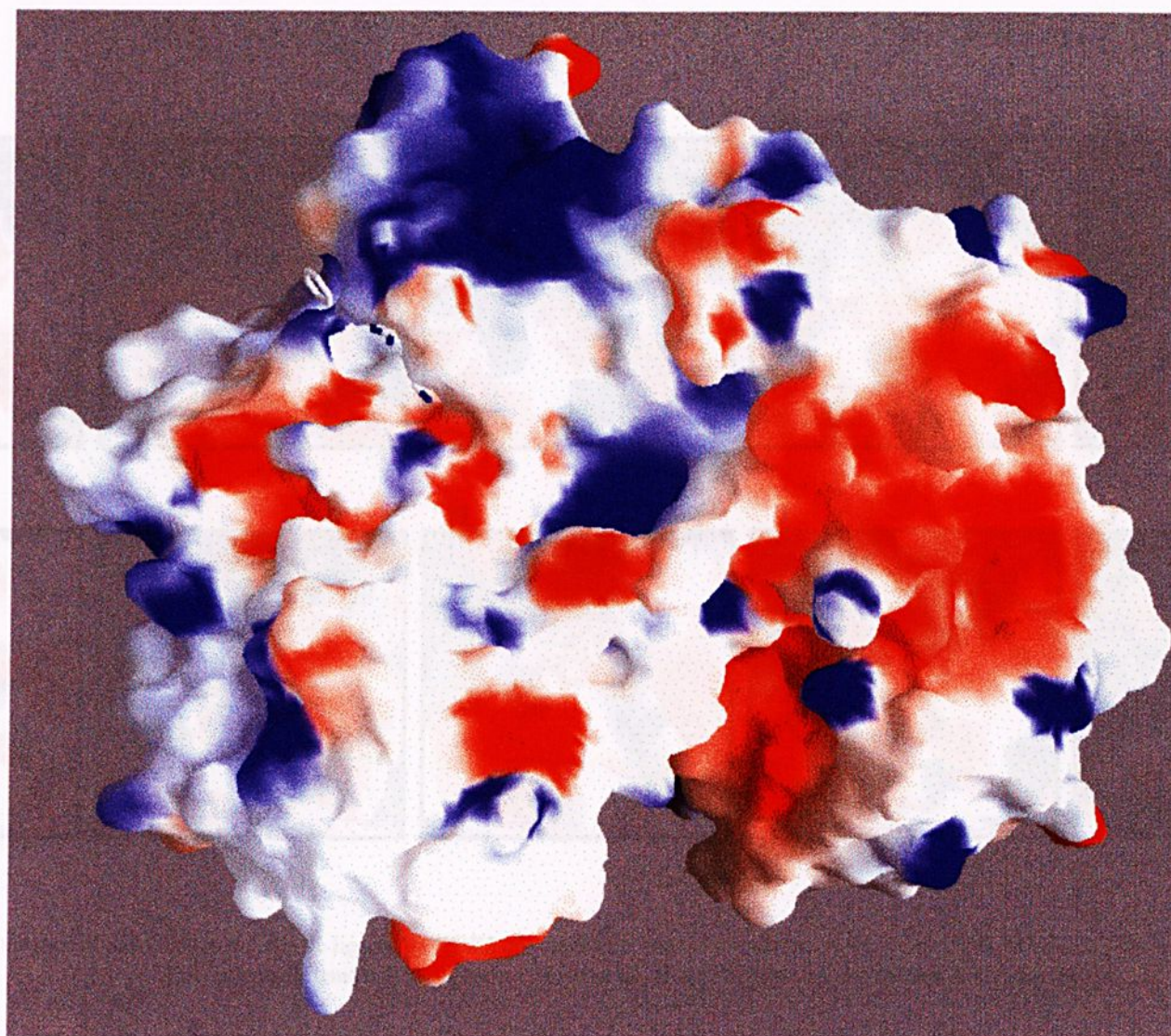
A análise do potencial eletrostático da superfície revelou a presença de certa complementaridade de cargas na região de contato entre a enzima e o inibidor sendo que a maior parte do sítio de ligação é hidrofóbica nas duas estruturas (figuras 45 a e b; 46 a e b).

Tabela 7: Resumo das análises estruturais

Inibidores	IC50 (µM)	Área de Contato (Å²)	Número de ligações de hidrogênio intermoleculares
Oxindol 91	0,13 (0,01)	283 (284)	3 (3)
Himenialdisina	0,6 (0,07)	231 (233)	6 (3)
Flavopiridol	0,4 (0,1)	319 (324)	4 (2)
Indirubina-5-sulfonato	0,3 (0,04)	265 (250)	7 (3)
Estaurosporina	<10 (0,007)	338 (330)	6 (2)
H717	2 (0,05)	341 (362)	2 (2)
Isopenteniladenina	>100 (50)	214 (224)	3 (2)
Roscovitina	>100 (0,7)	334 (360)	2 (2)
Olomoucina	>1000 (7)	281 (291)	4 (2)
Purvanalol B	>10 (0,006)	233 (242)	2 (2)

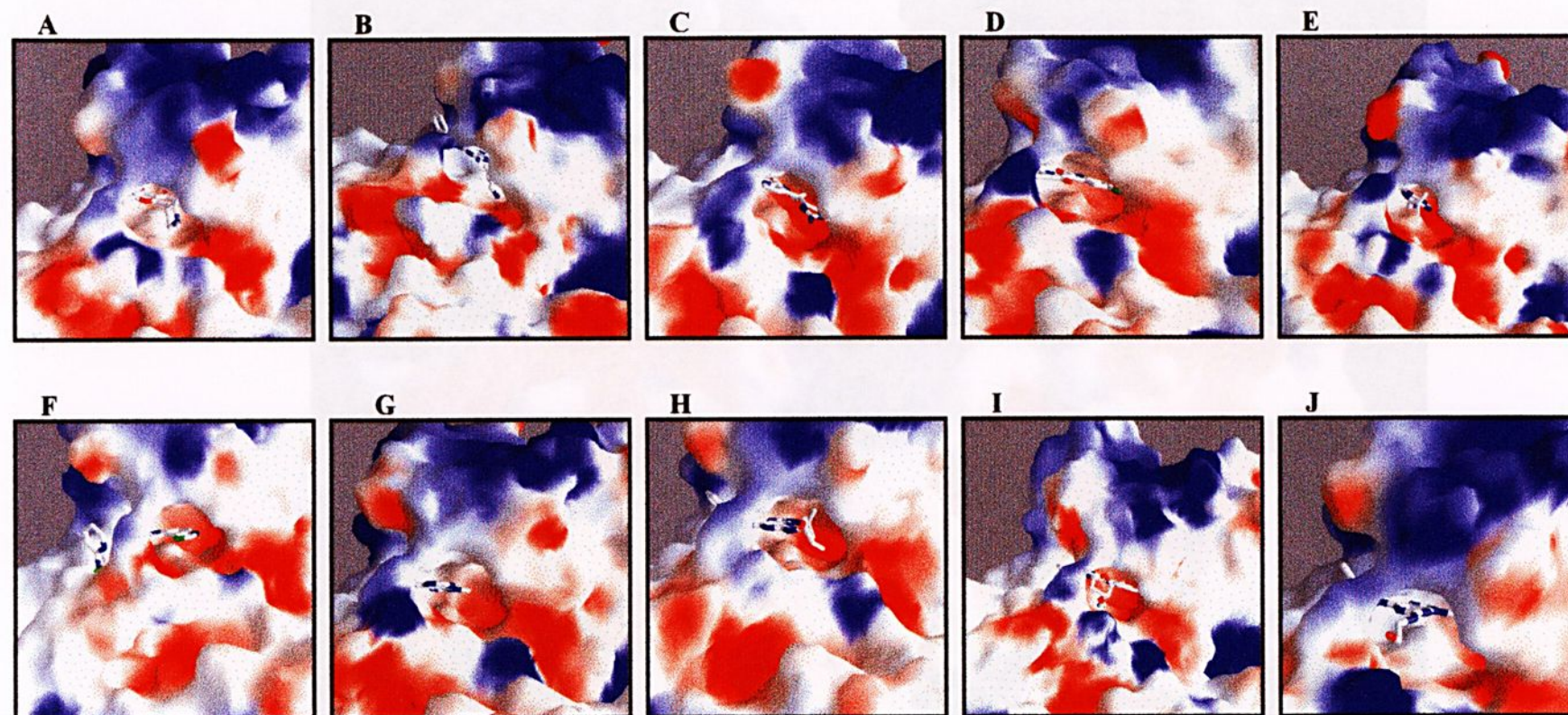
Os valores entre parênteses são dos moldes

A análise do potencial eletrostático de superfície revelou a presença de certa complementaridade de cargas na região de contato entre a enzima e o inibidor sendo que a maior parte do sítio de ligação é hidrofóbico nas duas estruturas (figuras 45 a e b; 46 a e b).



a)

Figura 45: a) Potencial eletrostático da CDK2/ciclina A complexada com o inibidor H717.



b)

Figura45: b) Potencial eletrostático do sitio de ligação da CDK2 em complexo com os inibidores: **A.** flavopiridol; **B.** H717; **C.** Himenialdisina; **D.** Indirubina; **E.** Isopenteniladenina; **F.** olomoucina; **G.** oxindol; **H.** purvanalol; **I.** roscovitina e **J.** Estaurosporina

Figura 46: a) Potencial eletrostático da CDK4/ciclina D complexada com inibidor H717.



a)

Figura 46: a) Potencial eletrostático da CDK4/ciclina D complexada com inibidor H717.

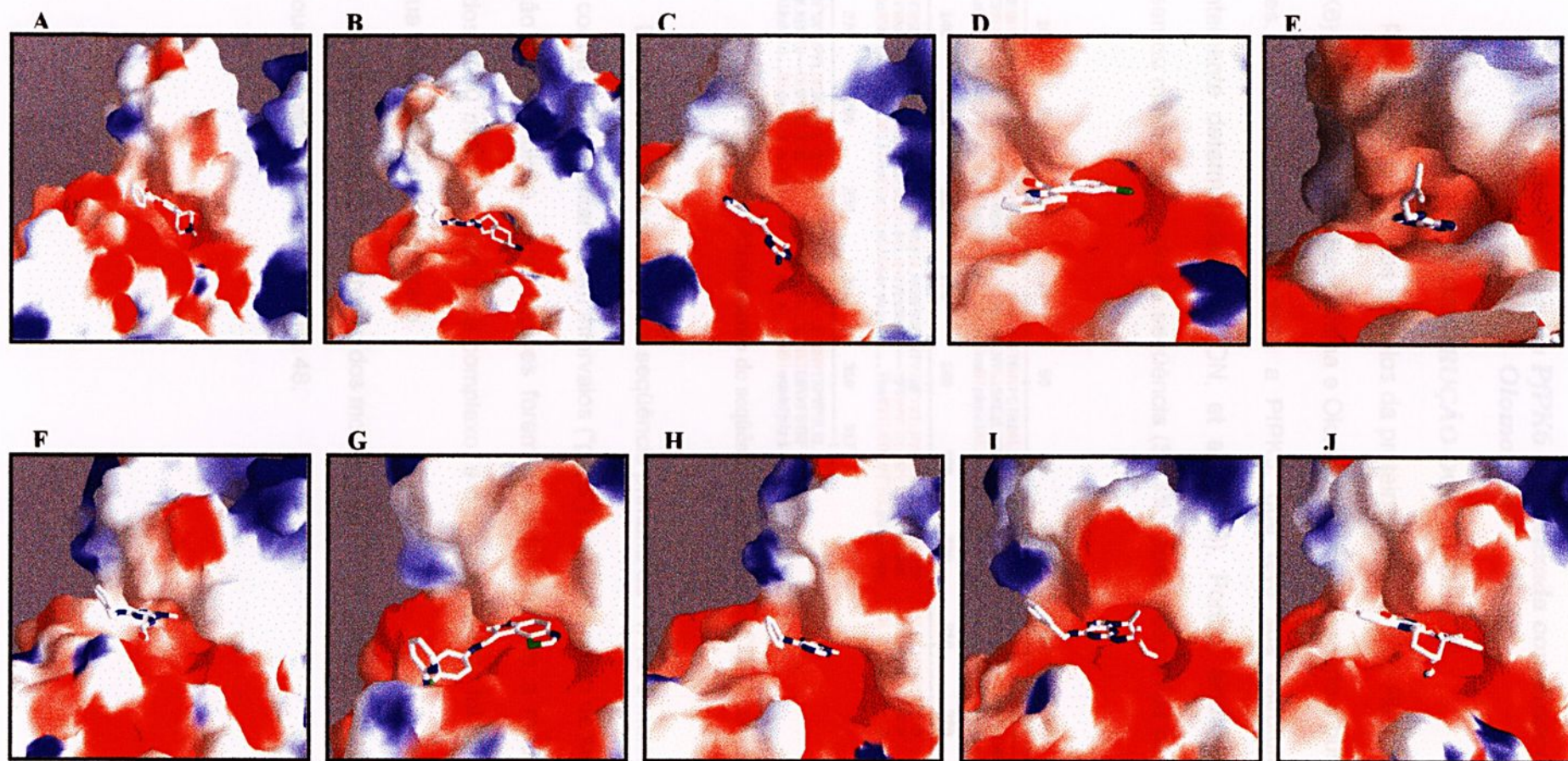


Figura46: b) Potencial eletrostático do sítio de ligação da CDK4 em complexo com os inibidores: **A.** flavopiridol; **B.** H717; **C.** Himenialdisina; **D.** Indirubina; **E.** Isopenteniladenina; **F.** olomoucina; **G.** oxindol; **H.** purvanalol; **I.** roscovitina e **J.** Estaurosporina

b)

4.2. Modelagem da PfPK6 complexada com Roscovitina e Olomoucina

4.2.1 CONSTRUÇÃO DOS MODELOS

Para a construção dos modelos da proteína quinase 6 do *Plasmodium falciparum* (PfPK6) complexada com Roscovitina e Olomoucina foi necessário a busca e seleção de moldes. Um possível molde seria a PfPK5, que teve sua estrutura cristalográfica recentemente determinada (HOLTON, et al., 2003). Porém, ainda a CDK2 humana apresentou maior identidade de sequência (39,3%), sendo escolhida como molde (figura 47).



Figura 47: Alinhamento de seqüências primárias da CDK2 com a PfPK6

É importante observar que a seqüência primária da PfPK6 é maior do que o molde (305 contra 298), resultando em intervalos (“gaps”) no alinhamento de seqüência. Apesar de não haver moldes, estas regiões foram mantidas durante a modelagem. Foram gerados 1000 modelos para cada complexo, e o melhor modelo foi escolhido, baseado em sua qualidade estereoquímica.

O protocolo para construção dos modelos de PfPK6 complexada com roscovitina e olomoucina é apresentado na figura 48.

programa PROCHECK (LASKOWSKI et al., 1993) (figura 49)

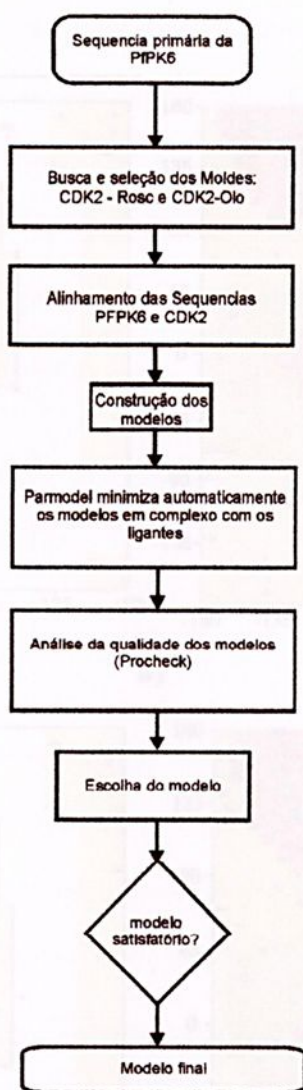


Figura 48: Protocolo para construção dos modelos Pfk6

Este método de modelagem possibilita a modelagem da proteína na presença do ligante, pois o programa faz uma otimização, ajustando a molécula do ligante.

4.2.2 QUALIDADE DOS MODELOS

Na análise da qualidade estereoquímica dos modelos e dos moldes foi usado o programa PROCHECK (LASKOWSKI et al., 1993), (figura 49).

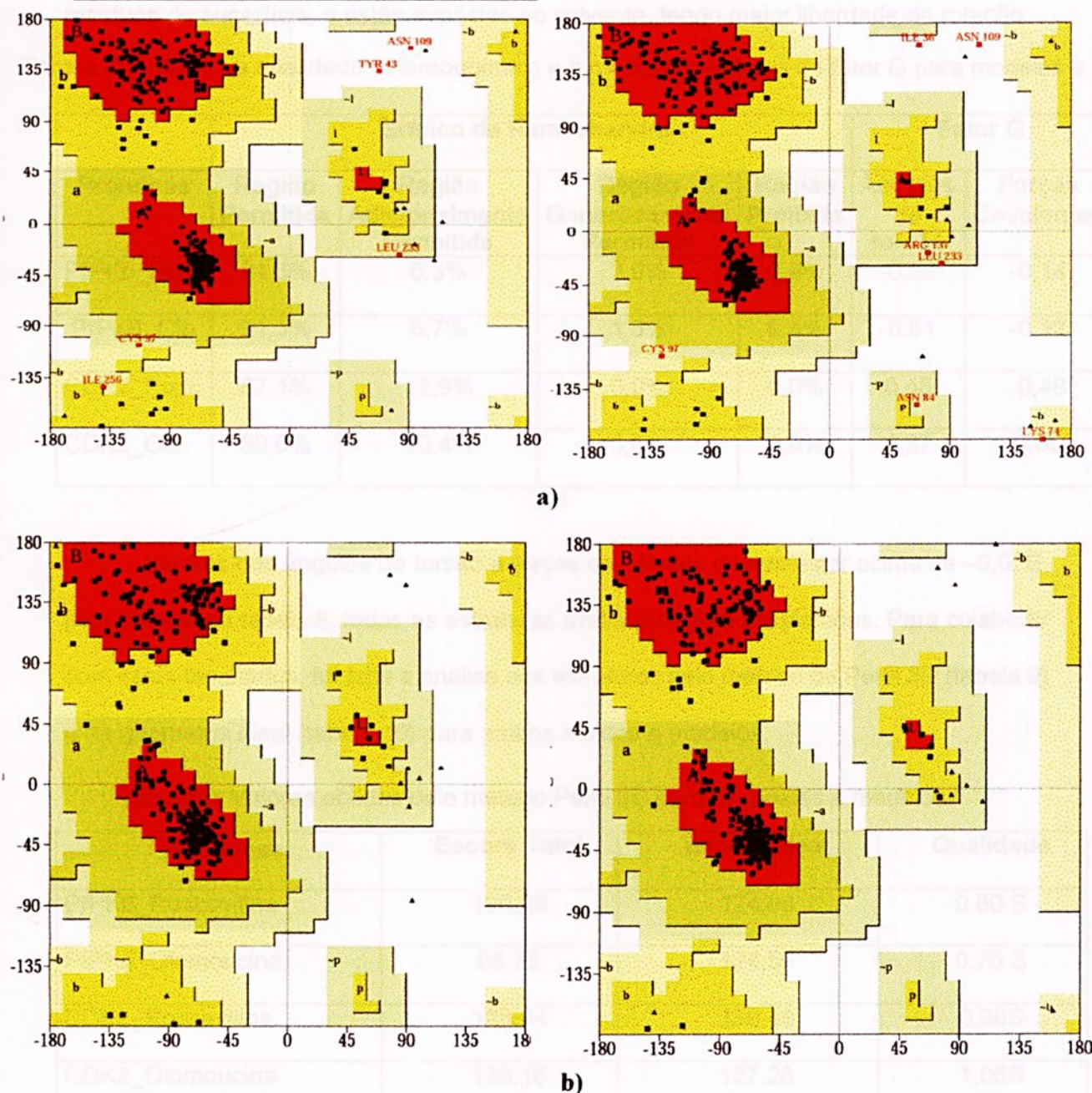


Figura 49: Gráficos de Ramachandran: a) Modelos de PfPK6 complexados com roscovitina e olomoucina, respectivamente; b) Estruturas de CDK2 complexadas com roscovitina e olomoucina, respectivamente.

A tabela 8 mostra a estatística da qualidade stereoquímica dos moldes e modelos, e valores de Fator G para ambos. Pela análise da tabela 8, os modelos são compatíveis aos seus moldes, pois apesar dos resíduos Tyr43 e Asn109 para o modelo da PfPK6: roscovitina e resíduos Ile36 e Asn109 para o modelo da PfPK6: olomoucina

pertencerem às regiões proibidas do gráfico (0,4%), pela análise das estruturas, estes são resíduos de superfície, e estão expostos ao solvente, tendo maior liberdade de rotação.

Tabela 8: Análise da qualidade estereoquímica e a média dos valores do fator G para modelos e moldes

Proteínas	Gráfico de Ramachandran				Fator G	
	Região Permitida	Região Adicionalmente Permitida	Região Generosamente Permitida	Região Proibida	Ângulos de torsão	Forças Covalentes
PfPK6_Rosc	91,3%	6,3%	2,0%	0,4%	-0,32	-0,14
PfPK6_Olo	91,3%	6,7%	1,6%	0,4%	-0,61	-0,12
CDK2_Rosc	87,1%	12,9%	0,0%	0,0%	0,45	-0,48
CDK2_Olo	89,6%	10,4%	0,0%	0,0%	0,37	0,05

Valores dos ângulos de torsão e forças covalentes devem estar acima de -0,5. E, pela análise da tabela 8, todas as estruturas avaliadas estão satisfatórias. Para colaborar com estes resultados, foi feita a análise das estruturas pelo método do Perfil 3D (tabela 9) e da geometria ideal (tabela 10) para ambos moldes e modelos.

Tabela 9: Valores obtidos pelo método Perfil 3D para os moldes e modelos.

Proteínas	Escore Total	Escore ideal	Qualidade
PfPK6_Roscovitina	100,29	124,96	0,80 S
PfPK6_Olomoucina	95,76	124,50	0,76 S
CDK2_Roscovitina	128,94	130,48	0,98S
CDK2_Olomoucina	136,16	127,26	1,06S

O valor da qualidade é obtido através da divisão do Escore total pelo Escore ideal, sendo que seu valor ideal estabelecido pelo método para uma boa qualidade e confiabilidade deve ser obtido acima de 0,45S. Ambos moldes e modelos apresentam valores dentro do limite estabelecido pelo método.

Tabela 10: Valores de r.m.s.d. para a geometria ideal

Proteínas	Comprimento de ligação (Å)	Ângulos de ligação (Graus)	Ângulos diedros (Graus)	Ângulos Impróprios (Graus)
PfPK6_Roscovitina	0,02	3,29	23,28	1,57
PfPK6_Olomoucina	0,02	3,33	23,40	1,13
CDK2_Roscovitina	0,24	2,26	25,91	2,61
CDK2_Olomoucina	0,39	3,24	24,79	1,76

A análise das tabelas 9 e 10 mostram que os modelos de PfPK6 são bons quando comparados com as estruturas da CDK2 na presença dos ligantes, e podem ser considerados para futuros estudos estruturais.

4.2.3 DESCRIÇÃO DOS MODELOS

O modelo da PfPK6 é similar à estrutura da CDK2. É composta de dois lóbulos, um lóbulo N-terminal composto predominantemente de fitas β pregueadas e um C-terminal composto de hélices α . Trata-se de uma estrutura tipicamente bilobal. A característica das quinases é a presença do sítio catalítico localizado entre estes dois lóbulos. A PfPK6 apresenta o mesmo número de fitas β e hélices α encontrado para o modelo da CDK4. Os elementos de estrutura secundária são mostrados nas figuras 50 a e b, gerados pelo programa MolMol (KORADI et al., 1996). É possível observar as moléculas dos inibidores posicionados entre os dois lóbulos, no sítio de ligação do ATP.

Figura 50: Elementos de estrutura secundária: a) PfPK6 roscovitina; b) CDK2 roscovitina; c) PfPK6 olomoucina e d) CDK2 olomoucina

RESULTADOS E DISCUSSÕES



Figura 50: Elementos de estrutura secundária: a) PfPK6:roscovitina ; b) CDK2:roscovitina;
c) PfPK6:olomoucina e d) CDK2:olomoucina

As maiores diferenças encontradas entre as estruturas dos moldes e dos modelos estão nas alças 85 a 94; 107 a 119; 164 a 174 e 237 a 262 das regiões da PfPK6 que podem ser avaliadas no alinhamento de seqüências primárias da figura 47.

Na presença dos inibidores, que são menores, estes tendem a se acomodar dentro

4.2.4 INTERAÇÃO DOS INIBIDORES COM A PfPK6

Sabendo que a especificidade e afinidade entre enzima e seus inibidores dependem das ligações de hidrogênio e interações iônicas, tão bem quanto da complementaridade de carga e de forma entre ambos (DE AZEVEDO et al., 1996), foi feita a análise das interações intermoleculares para ambas, PfPK6 e CDK2, na presença dos inibidores. Foi observado que as estruturas de ambos, moldes e modelos, fazem duas ligações de hidrogênio com os inibidores, através dos resíduos Leu83 e Leu95, respectivamente (tabela11).

Tabela 11: Ligações de hidrogênio intermoleculares

Proteína-Inibidores		Resíduos	Distância (Å)
CDK2-Roscovitina	N7	Leu83N	3,4
	N6	Leu83O	2,8
CDK2-Olomoucina	N7	Leu83N	2,9
	N6	Leu83O	2,6
PfPK6-Roscovitina	N7	Leu95N	3,4
	N6	Leu95O	2,8
PfPK6-Olomoucina	N7	Leu95N	2,9
	N6	Leu95O	2,8

Para observar possíveis alterações com relação às cadeias laterais dos resíduos que compõem o sítio de ligação da PfPK6, este foi sobreposto aos sítios de ligação da CDK2 na presença dos mesmos ligantes, e também na presença do ATP (Tabela 12). Esta sobreposição foi baseada nos carbonos α dos resíduos que compõem o bolsão de ligação (figuras 51 e 52).

É interessante observar que ocorre uma rotação do anel de purina do inibidor, em relação ao ATP, em ambas estruturas. Isto provavelmente se deve à longa cadeia de fosfato da molécula de ATP, que fica mais exposta para estar acessível à transferência do fosfato. Na presença dos inibidores, que são menores, estes tendem a se acomodar dentro do bolsão.

Figura 51 - Sobreposição dos Carbonos α dos bolsões de ligação: a) PFKfb (vermelho) com CDK2 (azul) e b) PFKfb (vermelho) com CDK2 (azul).

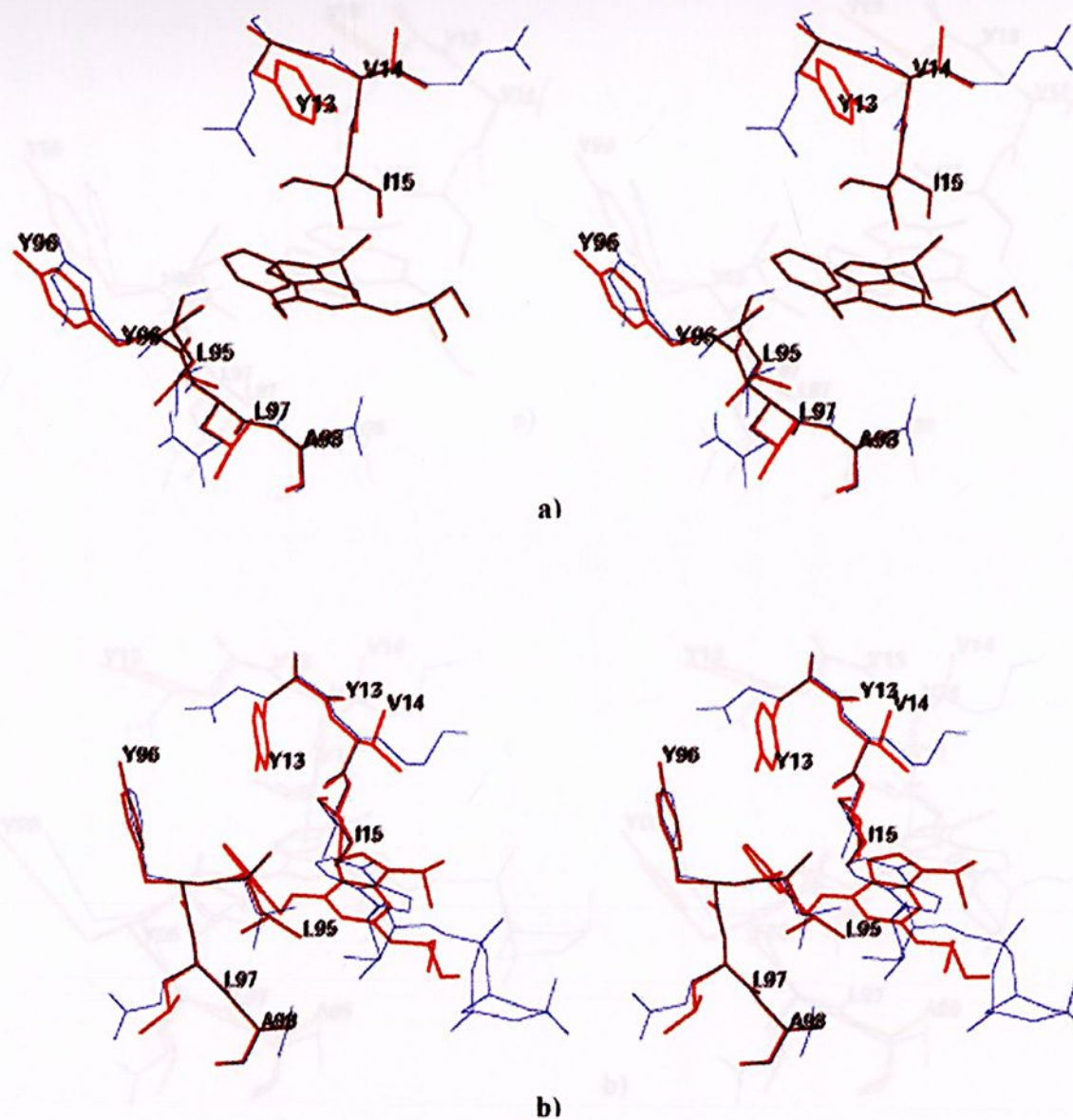


Figura 51: Sobreposições dos Carbonos α dos bolsões de ligação a) PpPK6:roscovitina (Linha vermelha) com CDK2:roscovitina (Linha azul) e b) PpPK6:roscovitina (Linha vermelha) com CDK2-ATP (Linha azul)

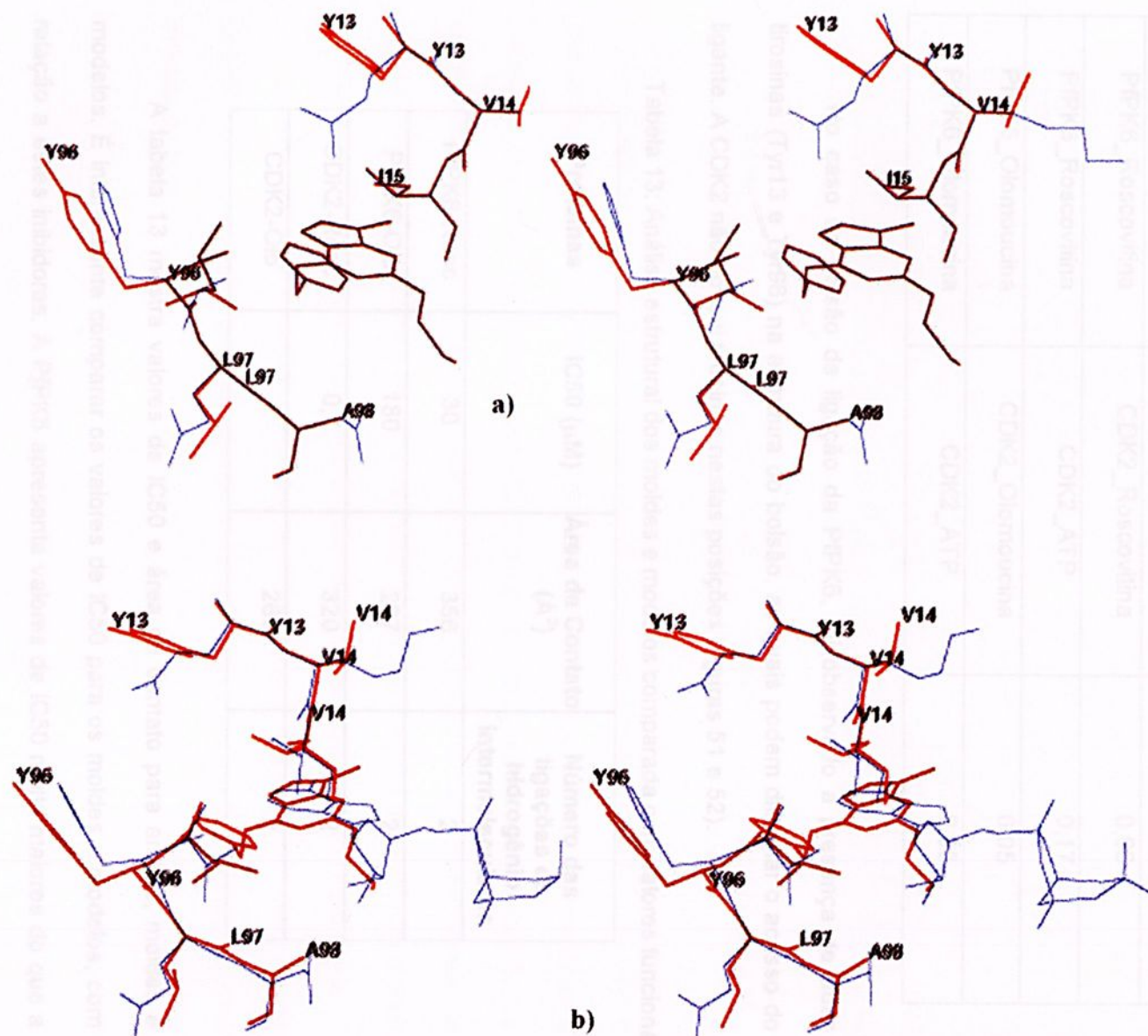


Figura 52: Sobreposições dos carbonos alfas dos bolsões de ligação a) PfPK6: olomoucina (Linha vermelha) com CDK2:olomoucina (Linha azul) e b) PfPK6:olomoucina (Linha vermelha) com CDK2:ATP (Linha azul)

Tabela 12: Valores de rmsd calculados pela sobreposição dos bolsões de ligação dos moldes e modelos

Modelos	Moldes	R.M.S.D (Å)
PfPK6_Roscovitina	CDK2_Roscovitina	0,06
PfPK6_Roscovitina	CDK2_ATP	0,17
PfPK6_Olomoucina	CDK2_Olomoucina	0,05
PfPK6_Olomoucina	CDK2_ATP	0,36

No caso do bolsão de ligação da PfPK6, foi observado a presença de duas tirosinas (Tyr13 e Tyr96) na estrutura do bolsão, as quais podem dificultar o acesso do ligante. A CDK2 não possui tirosinas nestas posições (figuras 51 e 52).

Tabela 13: Análise estrutural dos moldes e modelos comparada com valores funcionais de IC50

Proteínas	IC50 (µM)	Área de Contato (Å²)	Número das ligações de hidrogênio intermoleculares
PfPK6-Rosc	30	356	2
PfPK6-Olo	180	297	2
CDK2-Rosc	0,7	320	2
CDK2-Olo	7	269	2

A tabela 13 mostra valores de IC50 e área de contato para ambos, moldes e modelos. É interessante comparar os valores de IC50 para os moldes e modelos, com relação a estes inibidores. A PfPK6 apresenta valores de IC50 muito maiores do que a CDK2, sendo que ambas apresentam áreas de contato proteína-ligante, compatíveis, e mesmo número de ligação de hidrogênio.

A figuras 53 a e b mostram o posicionamento deste par de Tirosinas da PfPK6 representando um obstáculo para a ligação dos inibidores.

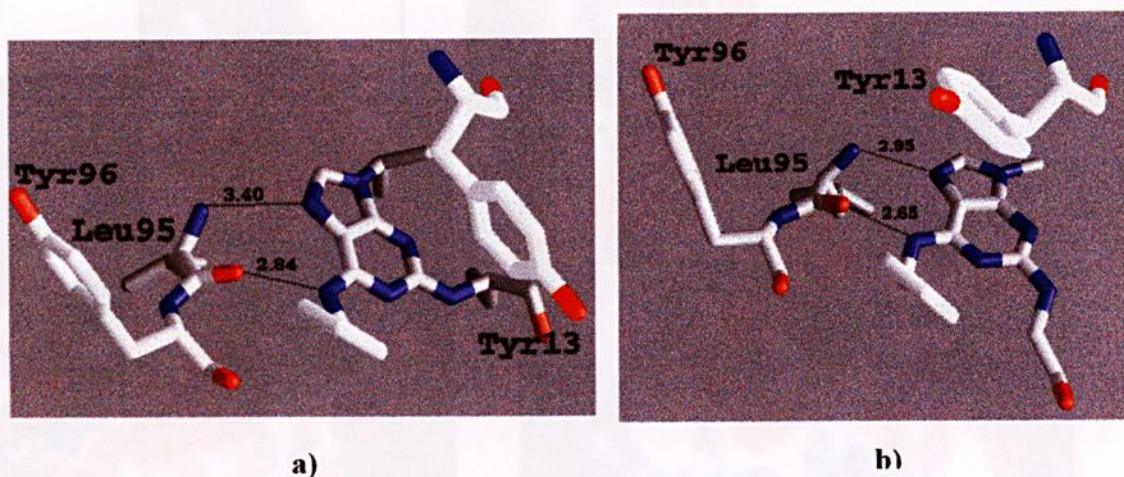


Figura 53: Par de Tirosinas na entrada do bolsão de ligação da PfPK6 a) na presença da roscovitina; b) na presença da olomoucina

A análise do potencial eletrostático da superfície das proteínas foi feita com o auxílio do programa GRASP (NICHOLLS et al., 1991). As figuras 54 a e b mostram a superfície molecular da PfPK6 e da CDK2 na presença dos inibidores roscovitina e olomoucina. A análise das figuras revelam a presença de certa complementaridade de cargas, principalmente entre a PfPK6 e os inibidores, já que o bolsão de ligação da PfPK6 é bastante negativo. No caso da CDK2, a maior parte do sítio de ligação é hidrofóbico.

Pela análise da superfície molecular da PfPK6, é possível observar apenas parte dos inibidores, pois estes estão parcialmente visíveis devido a presença das tirosinas na entrada do bolsão.

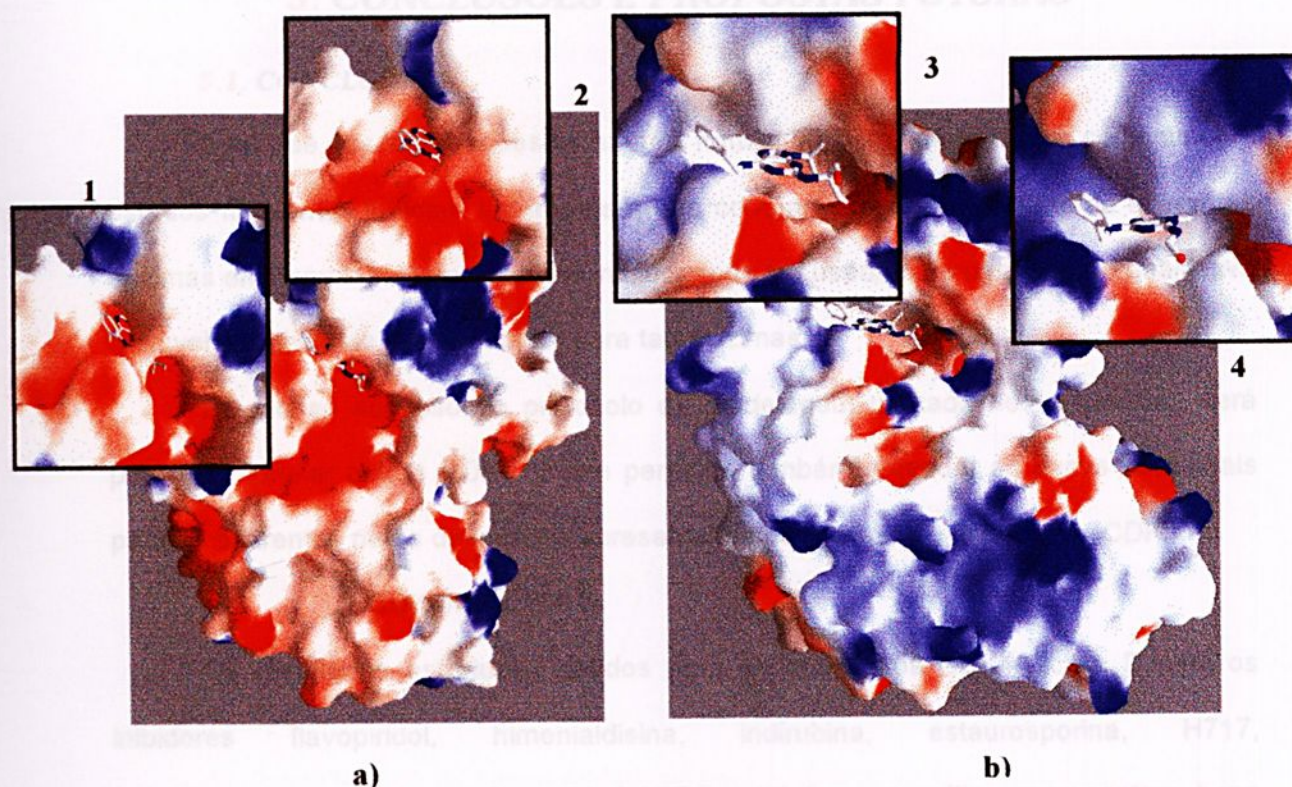


Figura 54: Superfície mostrando o potencial eletrostático de a) PIPK6 com 1-roscovitina e 2-olomoucina; b) CDK2 com 3-roscovitina e 4-olomoucina.

na cristalograficamente resolvida. Foi a primeira vez que uma CDK foi modelada na presença de ciclina, e mostrou ser possível, já que a comparação com a estrutura de CDK2/ciclina A evidenciou este fato. A presença de ciclina causa mudanças conformacionais que puderam ser observadas no modelo, as quais são importantes para a função desempenhada pelas CDKs.

Comparações da CDK4 com outras CDKs modeladas anteriormente corroboram com os resultados obtidos (DE AZEVEDO et al., 2002a,b; CANDURI et al., 2004; CANDURI & DE AZEVEDO, 2005).

Devido os complexos de CDK4/ciclina D com os inibidores estudados, flavopiridol, flavonolida, flavonolida e flavonolida mostraram ser mais específicos do que H717, roscovitina, olomoucina, oxindol, purvanol e roscovitina. Recentes estudos



5. CONCLUSÕES E PROPOSTAS FUTURAS

5.1. CONCLUSÕES

A análise das informações estruturais fornecidas pelos modelos em conjunto com estudos funcionais indicam fortemente que a comparação das estruturas de diferentes enzimas em complexo com o mesmo inibidor pode ser usada para uma análise qualitativa do nível de atividade desse inibidor para tais enzimas.

Devido ao sucesso do protocolo de modelagem utilizado neste trabalho, será possível modelar outras CDKs, o que permitirá também entender as bases estruturais para os diferentes perfis de inibição apresentados por diversos inibidores de CDKs.

5.1.1 CDK4/CICLINA D

Os resultados estruturais obtidos para os complexos CDK4/Ciclina D com os inibidores flavopiridol, himenialdisina, indirubina, estaurosporina, H717, isopenteniladenina, olomoucina, oxindol, Purvanalol e roscovitina representam bons modelos a serem usados para comparação com outras estruturas de CDKs modeladas, ou cristalograficamente resolvidas. Foi a primeira vez que uma CDK foi modelada na presença da ciclina, e mostrou ser possível, já que a comparação com a estrutura da CDK2/Ciclina A evidenciou este fato. A presença da ciclina causa mudanças conformacionais que puderam ser observadas no modelo, as quais são importantes para a função desempenhada pelas CDKs.

Comparações da CDK4 com outras CDKs modeladas anteriormente corroboram com os resultados obtidos (DE AZEVEDO et al., 2002a,b, CANDURI et al., 2004; CANDURI & DE AZEVEDO, 2005).

Dentre os complexos de CDK4/ciclina D com os inibidores estudados, flavopiridol, himenialdisina, indirubina e estaurosporina mostraram ser mais específicos do que H717, isopenteniladenina, olomoucina, oxindol, purvanalol e roscovitina. Recentes estudos



T. 1405/05



mostram que Flavopiridol encontra-se em fase 2 dos testes clínicos (FABIAN et al., 2005), com isso reforçando e evidenciando os resultados obtidos pelo presente trabalho. Novos inibidores poderão ser baseados nestas observações, a fim de se chegar a inibidores mais específicos para a CDK4.

5.1.2 PfPK6

Os resultados estruturais obtidos para a PfPK6 em complexo com roscovitina e olomoucina mostraram que inibidores mais específicos poderão ser avaliados, pois estes mostraram ser pouco eficazes para PfPK6, baseados nos valores funcionais (IC₅₀) e em análises estruturais. A presença das tirosinas 13 e 96 representam um obstáculo, o qual pode ser explorado na procura por novos inibidores.

5.2. PROPOSTAS FUTURAS DE TRABALHO

- A) Fazer uso da dinâmica molecular para conhecer as interações CDK_ciclina e na presença e na ausência de inibidores, para que possa existir uma melhor compreensão de como ocorre o processo de ativação/inibição;
- B) Propor modelagens para todas as CDKs associadas com suas respectivas ciclinas e na presença de possíveis inibidores e compará-las, fazendo uso dos dois métodos Modelagem molecular e Dinâmica Molecular, checar a qualidade desses modelos e compará-los com outros modelos, e com estruturas cristalográficas e, se possível, compará-los com dados funcionais. Isso irá ajudar a solucionar o grande problema que envolve as CDKs a procura por inibidores que apresentem maior especificidade e afinidade por uma determinada CDK, já que as CDKs possuem sítios de ligação muito conservados;
- C) Executar a dinâmica Molecular para tentar solucionar o problema apresentado no alinhamento da PfPK6 com a CDK2: a presença de intervalos que não

possuem moldes passíveis de modelagem, problema comum em modelagem

molecular comparativa.

ALBERTS, B., et al. *Biology of the cell*. 3rd ed. New York: London Garland, 1999.

ALEEM, E., et al. Cdk2 as a master of S phase entry: fact or fake? *Cell Cycle*, v.3, p. 35-37, 2004.

ALTSCHUL, S. F., et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, v.25, p. 3389-3402, 1997.

ANDERSON, M., et al. Imidazol[1,2-*a*]pyridines: a potent and selective class of cyclin-dependent kinase inhibitors identified through structure-based hybridisation. *Bioorganic & Medicinal Chemistry Letters*, v.13, p. 3021-3026, 2003.

BADGER, A. M., et al. Inhibition of interleukin-1-induced proteoglycan degradation and nitric oxide production in bovine articular cartilage/chondrocyte cultures by the natural product, hymenialdisine. *The Journal of pharmacology and experimental therapeutics*, v. 290, N. 2, p.587-593, 1999.

BAIROCH, A.; APWEILER, R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.*, v. 28, p. 45-48, 2000.

BAIROCH, A.; APWEILER, R. The SWISS-PROT protein sequence database: its relevance to human molecular medical research. *J. Mol. Med.*, v. 75, p. 312-318, 1997.

BAYAT, A. Science, medicine, and the future: Biotechnology. Clinical review. *BMJ.*, v. 324, p. 1018-1022, 2002.

BARVIAN M., et al. Pyrido[2,3-*d*]pyrimidin-7-one inhibitors of Cyclin-Dependent Kinases. *J. Med. Chem.*, v. 43, n. 24, p. 4606 - 4616, 2000.

BERMAN, H. M., et al. The Protein Data Bank. *Nucleic Acids Res.*, v. 28, p. 235-242, 2000.

BERNSTEIN, F. C., et al. The Protein Data Bank: a computer-based archival file for macromolecular structures. *J Mol Biol.*, v. 112, n. 3, p. 535-542, 1977.

BERTHET C., et al. Cdk2 knockout mice are viable. *Curr. Biol.*, v.13, p. 1775-1785, 2003.

6. REFERÊNCIAS

- ALBERTS, B., et al. Molecular biology of the cell. 3rd ed. New York: London Garland, 1999.
- ALEEM, E., et al. Cdk2 as a master of S phase entry: fact or fake? *Cell Cycle.*, v.3, p. 35-37, 2004.
- ALTSCHUL, S. F., et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, v.25, p.3389-3402, 1997.
- ANDERSON, M., et al. Imidazol[1,2-a]pyridines: a potent and selective Class of cyclin-dependent kinase inhibitors identified through structure-based hybridisation. *Bioorganic & Medicinal Chemistry Letters*, v.13, p. 3021-3026, 2003.
- BADGER, A. M., et al. Inhibition of interleukin-1-induced proteoglycan degradation and nitric oxide production in bovine articular cartilage/chondrocyte cultures by the natural product, hymenialdisine. *The Journal of pharmacology and experimental therapeutics*, v. 290, N. 2, p.587-593, 1999.
- BAIROCH, A.; APWEILER, R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.*, v. 28, p. 45-48, 2000.
- BAIROCH, A.; APWEILER, R. The SWISS-PROT protein sequence database: its relevance to human molecular medical research. *J. Mol. Med.*, v. 75, p. 312-316, 1997.
- BAYAT, A. Science, medicine, and the future: Bioinformatics. Clinical review. *BMJ.*, v. 324, p. 1018-1022, 2002.
- BARVIAN M., et al. Pyrido[2,3-d]pyrimidin-7-one Inhibitors of Cyclin-Dependent Kinases. *J. Med. Chem.*, v. 43, n. 24, p. 4606 - 4616, 2000.
- BERMAN, H. M., et al. The Protein Data Bank. *Nucleic Acids Res.*, v. 28, p. 235-242, 2000.
- BERNSTEIN, F. C., et al. The Protein Data Bank: a computer-based archival file for macromolecular structures. *J Mol Biol.*, v. 112, n. 3, p. 535-542, 1977.
- BERTHET C., et al. Cdk2 knockout mice are viable. *Curr. Biol.*, v.13, p. 1775-1785, 2003.



BOWIE, J. U., et al. A method to identify protein sequences that fold into a known three-dimensional structure. *Science*, v. 253, p.164 – 170, 1991.

BOWMAN, S., et al. The complete nucleotide sequence of chromosome 3 of *Plasmodium falciparum*, v. 400, p.532-538, 1999.

BRACCHI-RICHARD, V., et al. PfPK6, a novel cyclin-dependent kinase/mitogen-activated protein kinase-related protein kinase from *Plasmodium falciparum*, v. 347, p. 255-263, 2000.

BRAMSON, H.N., et al., Oxindole-based inhibitors of cyclin-dependent kinase 2 (CDK2): design, synthesis, enzymatic activities, and X-ray crystallographic analysis. *J. Med. Chem.*, v.44, p.4339-4358, 2001.

BROOKS, C.L. III., et al. CHARMM: A program for macromolecular energy minimization and dynamics calculations. *J. Comp. Chem.*, v. 4, p. 187-217, 1983.

BRUNGER, A. T. X-PLOR a system for X-ray Crystallography and NMR. Yale University Press New Haven USA, 1992.

CANDURI, F., et al. Molecular models of cyclin-dependent kinase 1 complexed with inhibitors. *Biochem. Biophys. Res. Commun.*, v. 324, p.661-666, 2004.

CANDURI, F.; DE AZEVEDO, W. F. JR. Structural basis for interaction of Inhibitors with Cyclin-Dependent Kinase 2. *Current Computer-Aided Drug Design*, v.1, p. 53-64, 2005.

CHOTHIA, C.; LESK, A. M. Canonical structures for the hypervariable regions of immunoglobulins. *J. Mol. Biol.*, v.196, p. 901 – 917, 1987.

COLLABORATIVE COMPUTACIONAL PROJECT Nº 4. The CCP4 Suite: Programs for protein crystallography. *Acta Crystallogr.*, v. 50, p. 760-763, 1994.

CORPET, F. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.*, v. 16, p.10881 – 10890, 1988.

DAÍ, Y.; GRANT, S. Cyclin-dependent kinase inhibitors. *Current Opinion in Pharmacology*, v. 3, p. 362-370, 2003.



DAVIES, T. G., et al. Structure-based desing of a potent purine-based cyclin-dependent kinase inhibitor. *Nature structural biology*, v. 9, p. 745-749, 2002.

DAVIES, T. G., et al. Inhibitor binding to active and inactive CDK2: the crystal structure of CDK2-cyclin A/Indirubin-5-Sulphonate. *Structure*, v. 9, p. 389-397, 2001.

DE AZEVEDO JR., W. F., et al. Molecular Model of Cyclin-Dependent Kinase 5 Complexed with Roscovitine. *Biochem. Biophys. Res. Commun.*, v. 297, n. 1, p.1454-1458, 2002 a.

DE AZEVEDO JR., W. F.; CANDURI, F.; SILVEIRA, N. J. F. Structural basis for inhibition of Cyclin-Dependent Kinase 9 by flavopiridol. *Biochem. Biophys. Res. Commun.*, v. 293, n. 1, p. 566-571, 2002 b.

DE AZEVEDO JR., W. F., et al. Structural basis for specificity and potency of a flavonoid inhibitor of human CDK2, a cell cycle kinase., v. 93, p. 2735-2740, 1996.

DE AZEVEDO JR., W. F., et al. Inhibiton of cyclin-dependent kinases by purine analogues crystal structure of human cdk2 complexed with roscovitine. *Eur. J. Biochem.*, v.243, p. 518-526, 1997.

DE BONDT, H.L., et al. Crystal structure of cyclin dependent kinase 2. *Nature*. v. 363, p. 592-602, 1993.

DELANO, W.L. The Pymol. Reference Manual in Copyright,. Delano Scientific LLC, San Carlos, California. U.S.A. All Rights Reserved,1998-2003.

DOOLITTLE, R. F. The grand assault. *Nature*. v. 419, p. 493-494, 2002.

DREYER, M.K., et al. Crystal structure of human cyclin-dependent kinase 2 in complex with the adenine-derived inhibitor H717. *J Med Chem.*, v. 44, n. 4, p.524-30, 2001.

FILHO SANTOS, A. O.; DE ALENCASTRO, R. D. Modelagem de proteínas por homologia. *Quim. Nova.*, v. 26, n. 2, 2003.

FABIAN, M. A., et al. A small molecule kinase interaction map for clinical kinase inhibitors. *Nature Biotechnology.*, v.23, p.329 – 336, 2005.

FISER, A.; SALI, A. Comparative protein structure modeling. *In Protein Structural*. p. 167-206, 2003.

FISER, A.; SALI, A. Comparative protein structure modeling with Modeller: A practical approach. *In protein structural.*, p. 1-73, 2000.

GARDNER, M., et al. Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature*. v. 419, p. 498-511, 2002.

GASTEIGER, E., et al. SWISS-PROT: connecting biomolecular knowledge via a protein database. *Curr. Issues Mol. Biol.*, v. 3, n. 3, p. 47-55, 2001.

GOULD, K. L. In Cyclin-dependent kinases in Protein Kinase Functions. Jim Woodgett Ed.; Frontiers in molecular biology. Oxford University Press: GreadBritain., p. 276-302, 2000.

GRAY, N. S., et al. Exploiting Chemical libraries, Structure, and Genomics in the search for kinase inhibitors. *Science*., v. 281, p. 533-538, 1998.

HALL, M., et al. Evidence for different modes of action of cyclin dependent Kinase inhibitors: p15 and p16 bind to Kinases, p21 and p27 bind to cyclins. *Oncogene*, v. 11. p. 1581-1588, 1995.

HARDCASTLE, I. R., et al. Designing inhibitors of cyclin-dependent kinases. *Annu. Rev. Pharmacol. Toxicol.*, v.42, p. 325-348, 2002.

HARMSE, L., et al. Structurae-activity relationships and inhibitory effects of various purine derivaties on the *in vitro* growth of *Plasmodium falciparum*. *Biochemical Pharmacology*. v. 62, p. 341-348, 2001.

HOESSEL, R., et al. Indirubin, the active constituent of a Chinese antileukaemia medicine, inhibits cyclin-dependet kinases. *Nature cell biology*, v. 1, p. 60-67, 1999.

HOLTON, S., et al. Structures of *P. falciparum* PfPK5 test the CDK regulation Paradigm and suggest mechanisms of small molecule inhibiton. *Structure*, v. 11, p. 1329-1337, 2003.

JOHNSON N. L., et al. Structural studies with inhibitors of the cell cycle regulatory kinase cyclin-dependent protein kinase 2. *Pharmacology & Therapeutics*., v. 93, p. 113-124, 2002.

KIM, S. H., et al. Structural basis for chemical inhibitor of cdk2. *Prog. Cell Cycle Res.*, v. 2, p.137-145, 1996.

KINOSHITA, K.; NAKAMURA, H. Protein informatics towards function identification. *Current Opinion in Structural Biology*, v. 13, p. 396-400, 2003.

KNOCKAERT, M., et al. Pharmacological inhibitors of cyclin-dependent kinases. *Trends in Pharmacological Sciences*, v. 23, n. 3, p. 417-425, 2002.

KORADI, R., et al. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graphics*, n. v.14, p. 51-55, 1996.

KOPP, J. & SCHWEDE, T. The SWISS-MODEL Repository of annotated three-dimensional protein structure homology models. *Nucleic Acids Research*, v. 32, DOI: 10.1093, 2004.

LASKOWSKI, R. A., et al. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.*, v. 26, p. 283-291, 1993.

LAWRIE, A. M., et al. Protein kinase inhibition by staurosporine revealed in details of the molecular interaction with CDK2. *Nature structural biology*, v. 4, n. 9, p. 797-800, 1997.

LE ROCH, K., et al. Activation of a *plasmodium falciparum* cdc2-related kinase by heterologous p25 and cyclin H. *The Journal of biological Chemistry*, v. 275, n. 12, p. 8952-8958, 2000.

LUTHY, R., et al. Assessment of protein models with three-dimensional profiles. *Nature*, v. 356, p.83-85, 1992.

MALUMBRES, M.; BARBACID, M. To cycle or not to cycle: a critical decision in cancer. *Nature*, v. 1, p. 222-231, 2001.

MALUMBRES, M., et al. Mammalian cells cycle without the D-type cyclin-dependent kinases Cdk4 and Cdk6. *Cell*, v. 118, p. 493-504, 2004.

MARTI-RENOM, M. A.; MADHUSUDHAN, M. S.; SALI, A. Alignment of protein sequences by their profiles, *Protein Sci.*, v. 13, p. 1071-1087, 2004.

MCREE, D. Copyright © by The Scripps Research Institute. 1999.

MEIJER, L.; RAYMOND, E. Roscovitina and other purines as kinase inhibitors. From starfish Oocytes to clinical trials. *Accounts of chemical research*, v. xx, p. A-1, 2002.

MEIJER, L., et al. Inhibitor of cyclin-dependent kinases, GSK-3beta and CK1 by hymenialdinsine, a marine sponge constituent. *Chem Biol.*, v. 7 n. 1, p. 51-63, 2000a.

MEIJER, L. Cyclin-dependent kinases inhibitors as potential anticancer, antineurodegenerative, antiviral and antiparasitic agents. *Drug Resist Updat.*, v.3, n.2, p.83-88, 2000b.

MONACO III, E. A., et al. Roscovitine, olomoucine, purvalanol: inducers of apoptosis in maturing cerebellar granule neurons. *Biochemical Pharmacology*, v. 67, p. 1947-1964, 2004.

MORGAN, D.O. Principles of CDK regulation. *Nature*, v. 374, p. 131-134, 1995.

MORRIS, A. L., et al. Stereochemical quality of protein structure coordinates. *J. Appl. Cryst.*, v. 12, p. 345-364, 1986.

MURTHI, K., et al. Structure-activity relationship studies of Flavopiridol Analogues. *Bioorganic & Medicinal Chemistry Letters.*, v. 10, p. 1037-1041, 2000.

NICHOLLS, A., et al. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, v.11, n. 4, p. 281-296, 1991.

PATEL, V., et al. Flavopiridol, a novel cyclin-dependent kinase inhibitor, Suppresses the growth of head and neck squamous cell carcinomas by inducing apoptosis. *The Journal of Clinical Investigation*, v. 102, n. 9, p. 1674-1681, 1998.

PEARSON, W. R. Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods Enzymol.*, v. 183, p. 63 – 98, 1990.

POLYCHRONOPOULOS, P., et al. Structural basis for the synthesis of indirubins as potent and selective inhibitors of glycogen synthase kinase-3 and cyclin-dependent kinases., v.47, p. 935-946, 2004.

PRIZE, W. J. J. Mechanisms of cyclin-dependent kinase regulation: Structures of CDKs, their cyclin activators, and Cip and INK4 Inhibitors. *J.Mol. Biol.*, v.287, p. 821-828, 1999.

RAMAKRISHNAN, C.; RAMACHANDRAN, G. N. Stereochemical criteria for polypeptide and protein chain conformations. II. Allowed conformations for a pair of peptide units. *Biophys J.*, v. 5, n. 6, p. 909 – 933, 1965.

RUSSO, A., et al. Structural basis of cyclin dependent kinase activation by phosphorylation. *Nat. Struct. Biol.*, v. 3, p. 696-700, 1996.

SALI, A.; BLUNDELL, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.*, v. 234, p.779-815, 1993.

SALI, A.; OVERINGTON, J. P. Derivation of rules for comparative protein modeling from a database of protein structure alignments. *Protein Sci.*, v. 3, n. 9, p.1582-1596, 1994.

SAUSVILLE, E. A. Complexities in the development of cyclin-dependent kinase inhibitor drugs, *A trends guide to cancer Therapeutics*, v. 8, n. 4, p. S32-S37, 2002.

SCHULZE-GAHMEN, U., et al. Multiple modes of ligand recognition: cristal structures of cyclin-dependent protein kinase 2 in complex with ATP and Two inhibitors, olomoucine and Isopentenyladenine. *Proteins. Structure, Function an genetics*, v. 22, p. 378-391, 1995.

SCHUTTE, B., et al. The effect of the cyclin-dependent kinase inhibitor olomoucine on cell cycle kinetics. *Experimental Cell Research.*, v. 236, p. 4-15, 1997.

SIELECKI, T. M., et al. Cyclin-dependent kinase inhibitors: Useful targets in cell cycle regulation. *Journal of Medicinal Chemistry*, v. 43, n. 1, p.1-16, 2000.

SWANTON, C. Cell-cycle targeted therapies. *The Lancet Oncology*. v. 5, p. 27-36, 2004.

THOMPSON, J. D., et al. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, v. 22, p.4673-4680, 1994.

UCHÔA, H. B., et al. Parmodel: a web server for automated comparative modeling of proteins. *Biochemical and Biophysical Research Communications.*, v. 325, p. 1481-1486, 2004.

VERMEULEN, K., et al. The cell cycle: a review of regulation, deregulation and therapeutic targets in cancer. *Cell prolifer.*, v.36, p. 131-149., 2003.

VOET, D., et al. Fundamentos de Bioquímica. ARTMED editora, 2002.

WARD, P., et al. Protein kinases of the malaria parasite *Plasmodium falciparum*: the kinome of a divergent eukaryote. *BMC Genomics*, doi: 10.1186/1471-2164-5-79. 2004.

WOODARD, C. L.; et al., Oxindole-based compounds are selective inhibitors of *Plasmodium falciparum* cyclin dependent protein kinases. *J. Med. Chem.*, v. 46, p. 3877-3882, 2003.

MODELLER (SALI & BLUNDELL, 1993; FISER & SALI, 2003)

XIN, S., et al. The role of RBF in developmentally regulated cell proliferation in the eye disc and in cyclin D/Cdk4 induced cellular growth. *Development*, v. 129, p. 1345-1356, 2002.

na biologia. Esta questão é geralmente facilitada pela ocorrência de estruturas tridimensionais nas proteínas em estudo. Estrutura tridimensional de proteínas naturais é guiada por dois distintos grupos de princípios operando em escalas de tempo vastamente diferentes: as leis físicas e a teoria da evolução. Cada um desses grupos de princípios de seleção de seqüências de proteínas naturais ocasionou métodos de predição para estruturas protéicas.

A primeira classe de método de predição de estrutura é conhecido como de novo ou ab initio, que prediz a estrutura de seqüências desconhecidas, sem dependência de estruturas conhecidas a nível de enovelamento de seqüências a ser modelada e alguma das estruturas conhecidas. Os métodos de novo assumem que a estrutura nativa corresponde ao mínimo de energia livre global acessível a proteína e tentam para encontrar este mínimo através de uma exploração de muitas conformações de proteínas concebíveis.

A segunda classe de métodos, incluindo "threading" é modelagem comparativa, descreve modo de detectar similaridade entre a seqüência a ser modelada e a estruturas conhecidas. Quando a seqüência de um membro de família de proteínas é conhecida experimentalmente, outro membro da família pode ser modelado baseado na similaridade de suas seqüências primárias.

A modelagem molecular comparativa por homologia consiste em modelar um novo membro de família de proteínas baseado em uma ou mais proteínas de estruturas conhecidas relacionadas (molde).

O método de modelagem foi desenvolvido tanto para uso em diferentes tipos de dados que seguem a desconhecida quanto possível. O método consiste de três estágios:

7. APÊNDICE: Informações sobre o programa **MODELLER** (SALI & BLUNDELL, 1993; FISER & SALI, 2000)

A caracterização funcional de uma proteína é um dos problemas mais freqüentes na biologia. Esta questão é geralmente facilitada pela ocorrência de estruturas tridimensionais nas proteínas em estudo. Estrutura tridimensional de proteínas naturais é guiada por dois distintos grupos de princípios operando em escalas de tempo vastamente diferentes: as leis físicas e a teoria da evolução. Cada um destes grupos de princípios aplicados em seqüências de proteínas naturais ocasionou métodos de predição para estruturas protéicas.

A primeira classe de método de predição de estrutura é conhecido como *de novo* ou *ab initio*, que prediz a estrutura da seqüência sozinha, sem dependências de similaridade a nível de enovelamento entre a seqüência a ser modelada e alguma das estruturas conhecidas. Os métodos *de novo* assumem que a estrutura nativa corresponde ao mínimo de energia livre global acessível a proteína e atentam para encontrar este mínimo através de uma exploração de muitas conformações de proteínas concebíveis.

A segunda classe de métodos, incluindo "threading" e modelagem comparativa, depende muito de detectar similaridade entre a seqüência a ser modelada e a estrutura conhecida. Quando a estrutura de um membro de família de proteínas tem sido determinada experimentalmente, outro membro da família pode ser modelado baseado no alinhamento de suas seqüências primárias.

A modelagem molecular comparativa ou por homologia constrói um modelo tridimensional para uma proteína desconhecida baseada em uma ou mais proteínas de estruturas conhecidas relacionadas (moldes).

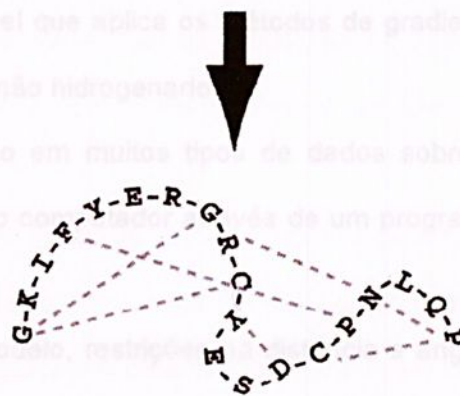
O método de modelagem foi desenvolvido tanto para uso em diferentes tipos de dados que cercam o desconhecido quanto possível. O método consiste de três estágios:

a) alinhamento da seqüência a ser modelada com a seqüência da proteína com estrutura conhecida e segmentos relacionados; b) extração de restrições espaciais em seqüências usando o alinhamento; e c) satisfação de restrições para obter um modelo tridimensional (figura 55).

1. Alinhamento da Seqüência com a Seqüência da estrutura conhecida

3D	GKIFYERGFQGHCEYSDC-NLQP
Seq	GKIFYERG---RCYESDCPNLQP

2. Extração das Restrições espaciais



3. Satisfação das Restrições espaciais



Figura 55: Principais passos do processo de modelagem (FISER & SALI, 2000)

As restrições espaciais da sequência a ser modelada são obtidas de análises estatísticas de relações entre várias características de estrutura protéica. Estas relações foram descritas como uma função de densidade condicional de probabilidade (pdfs) para as características a serem preditas. Os cálculos do modelo tridimensional pela satisfação de restrição espacial é atingido pela minimização da pdf molecular. Esta função é uma combinação das pdfs individuais restringindo características espaciais. O método de minimização é um método de função alvo variável que aplica os métodos de gradiente conjugado para as posições de todos os átomos não hidrogenados.

A aproximação foi desenvolvida para uso em muitos tipos de dados sobre a sequência alvo possível. Ele foi implementado no computador através de um programa chamado MODELLER.

No primeiro passo para construção do modelo, restrições na distância e ângulo diedral da sequência alvo são derivadas de seus alinhamentos com seqüências de estruturas tridimensionais conhecidas.

Se um número suficiente de fortes restrições é especificada, a estrutura tridimensional da proteína é bem determinada. Uma restrição é muito mais precisamente definida em termos de uma função de densidade de probabilidade, $p(x)$, para a característica x que é restringida. Ela deve integrar em 1 para todas as escalas de valores possíveis para x . A probabilidade finita atual de um evento $x_1 \leq x < x_2$ é obtida pela integração de p :

$$p(x_1 \leq x < x_2) = \int_{x_1}^{x_2} p(x) dx \quad (1)$$

Por exemplo, a probabilidade para valores diferentes de ângulos diedrais em cadeias principais são calculados: para um tipo de resíduo considerado, para uma conformação de cadeia principal de um resíduo equivalente, e para a similaridade de

seqüência entre as duas proteínas. Outro exemplo é a pdf para uma certa distância de Ca-Ca dada distâncias equivalentes em duas estruturas de proteínas relacionadas. Uma característica importante do método é que as formas de restrições espaciais foram obtidas empiricamente através de um banco de dados de alinhamentos de estruturas de proteínas.

Uma pdf útil para restringir certa característica x pode ser escrita como:

$$p(x/a,b,...,c) \quad (2)$$

Isto é uma pdf condicional que nos dá a densidade de probabilidade para x , quando $a,b,...,c$ são conhecidos. Isso pode ser visto como uma pdf para x que também depende dos valores de outras variáveis.

Para que uma pdf seja útil na modelagem, todas as características associadas ($x/a,b,...,c$) exceto x devem ser conhecidas no momento da predição. Além disso, x deve ser uma característica espacial no momento da predição. A característica conhecida mais importante é aquela do mesmo tipo de x e associada com posições equivalentes nas estruturas conhecidas relacionadas.

Na verdade, não é possível se obter a verdadeira função p e sim, suas aproximações:

$$p(x/a,b,...,c) \approx W_{x,a,b,...,c} \approx f(x,a,b,...,c,q) \quad (3)$$

onde, $W_{x,a,b,...,c}$ é uma tabela baseada nos valores de $x,a,b,...,c$ e que contém, com seus elementos, as freqüências relativas observadas para a ocorrência de determinados valores de x dados $a,b,...,c$. A função f é uma função analítica ajustada ao W observado. Assim como qualquer pdf, W e f devem satisfazer os critérios de não negatividade, e integrar para 1.

A tabela multidimensional de freqüências relativas W é calculada com base nas freqüências absolutas W' , usando-se a seguinte equação:

$$W_{x,a,b,\dots,c} = \frac{W'_{x,a,b,\dots,c}}{\sum_x W'_{x,a,b,\dots,c}} \quad (4)$$

As frequências absolutas W' , são obtidas por contagem direta do número de ocorrências de cada combinação de valores $(x,a,b,\dots,c)$ presentes na amostra, e tal amostra é derivada do banco de dados local de estruturas de proteínas conhecidas e seus alinhamentos.

O segundo passo, as restrições espaciais e os termos de campo de força que cumprem propriedades estereoquímicas e são combinadas dentro de uma função objetivo. A forma geral da função objetivo é similar para aqueles programas de dinâmica molecular, tais como CHARMM22 (BROOKS, et al., 1983). A forma funcional de cada termo é simples; ela inclui uma função quadrática, oscilador harmônico, cossenos e uma soma pesada de algumas funções Gaussianas, lei de Coulomb e potencial Lennard-Jones. As características geométricas presentemente incluem uma distância, um ângulo, um ângulo diedral, um par de ângulos diedrais entre dois, três e quatro átomos e oito átomos, respectivamente, a distância mais curta num grupo de distâncias, acessibilidade ao solvente em Å e densidade de átomo que é expressa como o número de átomos ao redor do átomo central. Algumas restrições podem ser usadas como para restringir pseudo-átomos, tais como, o centro de gravidade de vários átomos.

Estas satisfações de restrições espaciais mostram como a restrição espacial em seqüência podem ser expressas como pdfs. Estas pdfs foram obtidas de considerações estereoquímicas e de formas simples de estruturas homólogas. Nesta seção será descrito como combinar restrições de diversas estruturas homólogas e como usar estas restrições para derivar um modelo tridimensional. O princípio é obter o modelo mais provável para uma certa seqüência dada seu alinhamento com as estruturas relacionadas. O modelo

tridimensional é obtido por uma minimização de pdf molecular que dependem de seus modelos e de suas restrições.

A pdf molecular é reunida de características de pdfs que são obtidas de pdfs básicas.

7.1. Derivação de uma característica pdf através de pdfs básicas

Uma característica de uma estrutura de proteína é definida como alguma quantidade associada com um particular grupo de átomos $ijk...l$. Por exemplo, a distância entre átomos i e j é uma característica, a distância entre átomos j e k é uma outra característica e um ângulo ijk é ainda outra característica. As pdfs básicas, $p^f(f)$ são aquelas já descritas anteriormente. De uma forma geral, várias características estruturais f podem ser restringidas por diversas pdfs básicas $p_k^f(f)$ para $k = 1, 2, \dots$. A característica pdf, $p^F(f)$, é uma pdf que combina todas as informações a respeito dos valores possíveis que a característica f pode assumir.

O exemplo seguinte ilustra perfeitamente estas definições descritas acima. O objetivo é construir uma pdf característica para uma particular distância C α -C α em uma dada seqüência. Suponha que duas estruturas conhecidas e relacionadas com distâncias equivalentes são avaliadas; então neste caso têm-se duas pdfs básicas correspondentes para a distância C α -C α na seqüência alvo. Em adição, sabe-se que cada uma destas duas restrições deve satisfazer o critério de van de Waals, isto é, a distância tem que ser maior do que a soma dos dois raios de van der Waals. De forma a poder combinar todas essas informações, combinou-se três pdfs básicas dentro de uma única pdf característica. Para que fosse possível, alinhar as três proteínas em um banco de dados local.

O peso ω presente é proporcional à média da diferença na vizinhança do resíduo entre a estrutura correspondente e a seqüência a ser modelada. Os dados são ajustados de acordo com o seguinte modelo para $\omega(s)$:

$$\omega(s) = \frac{w(s)}{\sum_j w(s_j)} \quad (5)$$

onde

$$w(s) = a + \exp(bs^c); \sum_j \omega(s_j) = 1 \quad (6)$$

sendo **a**: acessibilidade do resíduo ao solvente, **b**: B_{ISO} médio do resíduo, **c**: classe χ dos ângulos diedrais da cadeia lateral, **s**: diferença de resíduos vizinhos entre duas proteínas.

O último passo da derivação da pdf característica é incluir as restrições de van de Waals. Desde que todas as restrições estereoquímicas tenham que ser satisfeitas em todas as estruturas, estas restrições são multiplicadas dentro da pdf característica e assim se obtém a pdf de característica final:

$$p^p(d) = [\omega_1 p_1'(d) + \omega_2 p_2'(d)] p^r(d) \quad (7)$$

Esta aproximação simples por combinação de duas das pdfs básicas é utilizada para qualquer número de pdfs básicas que foram derivadas de estruturas relacionadas.

7.2. Derivação de uma pdf molecular a partir das pdfs características

A meta é encontrar a estrutura tridimensional que é consistente com muitos valores prováveis de características f_i individuais. A pdf molecular deve fornecer uma probabilidade de ocorrência de alguma destas características simultaneamente. Então o modelo para a estrutura tridimensional desconhecido poderá corresponder ao máximo da pdf molecular. Assumindo que pdfs características são independentes uma das outras, a pdf molecular é simplesmente um produto das pdfs de características:

$$P = \prod_i p^r(f_i) \quad (8)$$

Sendo assim, maximizando a função P pode-se encontrar o modelo mais provável para a estrutura 3D da seqüência, dado seu alinhamento com as estruturas conhecidas.

Finalmente, no terceiro e último passo, o modelo é obtido pela otimização da função objetivo, em um espaço cartesiano. A otimização faz uso do método de função alvo variável, empregando métodos de gradientes conjugados e dinâmicas moleculares com recozimento simulado. Diversos modelos ligeiramente diferentes podem ser calculados, variando a estrutura inicial, e a variabilidade entre estes modelos pode ser usada para estimar a baixa ligação sobre os erros em regiões correspondentes do enovelamento.

7.3. Otimização da pdf molecular

Este programa tem uma variedade de funções que facilitam a modelagem, incluindo construção de estruturas através de seqüência, múltipla comparação de proteínas, comparações de características de proteína com restrições, e várias rotinas gráficas que tem como saída arquivos PostScript (PROCHECK (LASKOWSKI, et al., 1993). O MODELLER foi implementado como uma interpretação de uma linguagem de alto-nível especializada para trabalhar com as estruturas de proteína.

A função que é realmente otimizada é uma transformação da pdf molecular de P :

$$F = -\ln(p) \quad (9)$$

onde todas as características são expressas em termos de coordenadas atômicas no espaço Cartesiano. A função F é chamada de função objetivo. O mesmo conjunto de coordenadas Cartesianas que maximiza P também minimiza F , pelo fato de que se substitui a multiplicação de termos em P pela adição de termos em F , o que reduz o problema de "overflow" em ponto flutuante.

A segunda transformação da pdf molecular original, que é útil para a otimização, é a mudança de escala dos desvios padrões das restrições básicas. Isso permite a



mudança de escala independente para cada tipo de característica a se restringir (tamanhos de ligação, ângulos de ligação, distâncias entre C α -C α , etc). Aumentando o desvio padrão, as restrições se tornam menos poderosas, e violações maiores tendem a ocorrer.

O método da função alvo variável foi inicialmente utilizado para o cálculo de estruturas tridimensionais de proteínas consistentes com restrições obtidas de NMR bidimensional. A principal diferença entre o método e a atual implementação é que a otimização atual é feita no espaço Cartesiano e o procedimento original otimiza os ângulos diédricos.

Seguindo o método da função alvo variável, o ponto ótimo da pdf molecular é encontrado após sucessivas otimizações de funções alvo, gradativamente mais complexas, resultando na verdadeira pdf molecular ao final. Esta série é obtida partindo-se das restrições locais e introduzindo-se mais e mais restrições de longa distância, finalmente chegando à verdadeira pdf molecular que incorpora todas as restrições. Mais precisamente, a função alvo $P(\Delta r)$ é definida como uma função de variável inteira $\Delta r = 1, \dots, N$, onde N é o número de resíduos da seqüência que está sendo modelada. A função alvo $P(\Delta r)$ é obtida da mesma forma que a pdf molecular, exceto que somente aquelas restrições cujos átomos que as originam pertençam a resíduos que distem menos que Δr do início da seqüência, são incluídas. O cálculo como um todo consiste de um grande número de otimizações de gradiente conjugado da função alvo $P(\Delta r)$ com valores crescentes de Δr . A conformação inicial $P(1)$ a ser otimizada pode ser uma estrutura estendida ou uma conformação derivada de uma cadeia estendida através de rotação dos ângulos diedrais das cadeias principal e lateral. Nos passos subseqüentes do método de função alvo variável a conformação inicial a ser otimizada é o modelo final do passo anterior. Um conjunto de diferentes modelos finais pode ser obtido utilizando-se diferentes conformações iniciais (figura 56)

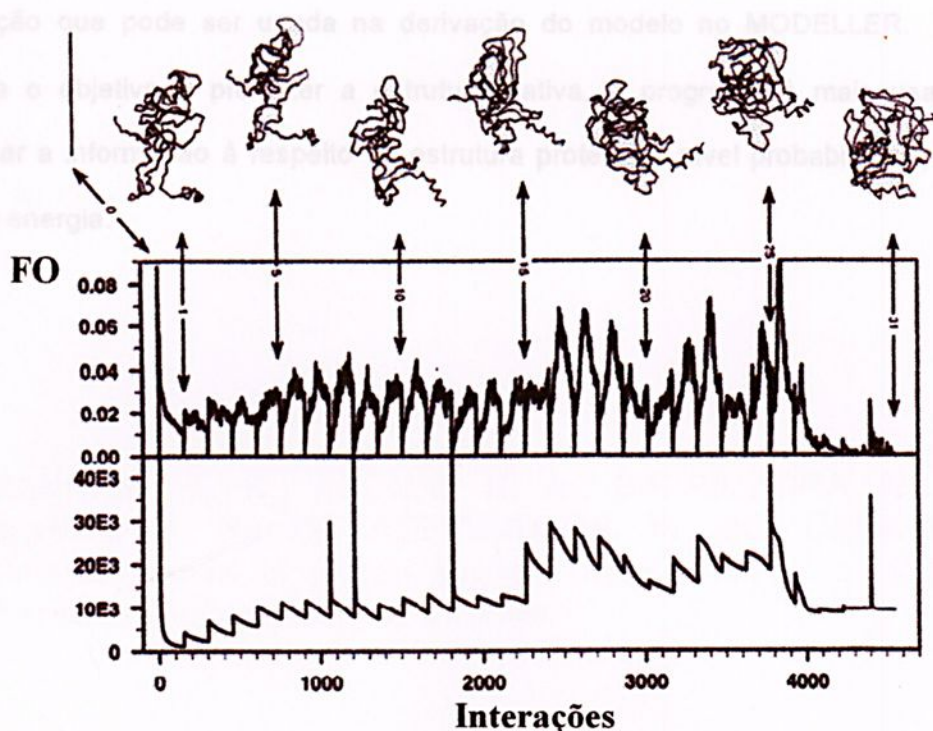


Figura56: Mudança de conformação pela otimização da função objetivo (FISER & SALI, 2000)

Essa otimização parte de uma conformação aleatória inicial e inicialmente usa-se somente restrições sequencialmente locais, e procedem em um número de passos para aumentar o número de restrições até que, finalmente, todas as restrições são incluídas em suas violações minimizadas. Ambas as extrações das restrições dos alinhamentos e a otimização da pdf molecular são completamente automatizadas.

O MODELLER calcula o modelo através de uma provável estrutura para uma certa sequência, dado seu alinhamento com relação às estruturas resolvidas, enquanto os métodos de minimização de energia calcula o modelo com a estrutura de energia mais baixa dado um campo de força.

Sendo assim a otimização do MODELLER não possui nenhuma significância física, tais como aqueles termos de energia individual, energia total e mecanismos moleculares em dinâmica molecular. Esta diferença estende-se a quantidade de

informação que pode ser usada na derivação do modelo no MODELLER. Quando somente o objetivo é pré-dizer a estrutura nativa, o programa é mais usado para processar a informação à respeito da estrutura protéica a nível probabilístico, e não à nível de energia.

A) MANHANI, K. K.; ARCURI, H. A.; DA SILVEIRA, N. J. F.;
UCHOA, H. B.; DE AZEVEDO, W. F. JR.; CANDURI, F.
Molecular model of protein kinase 6 from Plasmodium falciparum
Journal Modelling Molecular (in press)

B) MANHANI, K. K.; DA SILVEIRA, N. J. F.; UCHOA, H. B.; DE
AZEVEDO, W. F. JR.; CANDURI, F. *Molecular models of the*
Cyclin-Dependent Kinase 4 associated the Cyclin D and Complexed
with Inhibitors (Submitted)

Anexo: Artigos

KARLA KARLA MANHANI

HELEN ASSURADE ARCURI

NELSON JOSÉ FREITAS DA SILVEIRA

- A) MANHANI, K. K.; ARCURI, H. A.; DA SILVEIRA, N. J. F.; UCHOA, H. B.; DE AZEVEDO, W. F. JR.; CANDURI, F.** Molecular model of protein kinase 6 from *Plasmodium falciparum*. *Journal Modelling Molecular (In press)*

FERNANDA CANDURI (fcanduri@webmail.biotec.unesp.br)

- B) MANHANI, K. K.; DA SILVEIRA, N. J. F.; UCHOA, H. B.; DE AZEVEDO, W. F. JR.; CANDURI, F.** Molecular models of the Cyclin-Dependent Kinase 4 associated the Cyclin D and Complexed with Inhibitors (*Submetido*)

Universidade Estadual Paulista - UNESP
Departamento de Física
Rua Cristóvão Colombo, 2365

Código Postal: 13054-000

São José do Rio Preto - São Paulo - Brasil

Telefone: (017) 321-2463 - Fax: (017) 321-3249

Molecular Models of Protein Kinase 6 from *Plasmodium falciparum*

KARISA KARLA MANHANI, HELEN ANDRADE ARCURI, NELSON JOSÉ FREITAS DA
SILVEIRA, HUGO BRANDÃO UCHOA, WALTER FILGUEIRA DE AZEVEDO JR.,
FERNANDA CANDURI

KARISA KARLA MANHANI

HELEN ANDRADE ARCURI

NELSON JOSÉ FREITAS DA SILVEIRA

HUGO BRANDÃO UCHOA

WALTER FILGUEIRA DE AZEVEDO JR.

FERNANDA CANDURI (canduri@webmail.ibilce.unesp.br)

Abstract: Cyclin Dependent Kinases (CDKs) have been identified as potential targets for development of drugs, mainly against cancer. These studies generated a vast library of chemical inhibitors of CDKs, and some of these molecules can also inhibit Kinases identified in the *Plasmodium falciparum* genome. Here we describe structural models for Protein Kinase 6 from *Plasmodium falciparum* (PfPK6) complexed with Roscovitine and Olanocicline. These models show clear structural evidence for differences observed in the inhibition, and may help designing inhibitors for PfPK6 generating new potential drugs against malaria.

Key Words: Kinases, *Plasmodium falciparum*, Roscovitine and Olanocicline.

Universidade Estadual Paulista - UNESP
Departamento de Física
Rua Cristovão Colombo, 2265

Código Postal: 15054-000

São José do Rio Preto – São Paulo – Brasil

Telefone: (017) 221-2463 - Fax: (017) 221-2247



Molecular Models of Protein Kinase 6 from *Plasmodium falciparum*

KARISA KARLA MANHANI, HELEN ANDRADE ARCURI, NELSON JOSÉ FREITAS DA SILVEIRA, HUGO BRANDÃO UCHÔA, WALTER FILGUEIRA DE AZEVEDO JR. FERNANDA CANDURI

Programa de Pós-Graduação em Biofísica Molecular. Departamento de Física, UNESP, São José do Rio Preto, SP 15054-000, Brasil

Abstract: Cyclin Dependent Kinases (CDKs) have been identified as potential targets for development of drugs, mainly against cancer. These studies generated a vast library of chemical inhibitors of CDKs, and some of these molecules can also inhibit kinases identified in the *Plasmodium falciparum* genome. Here we describe structural models for Protein Kinase 6 from *Plasmodium falciparum* (PfPK6) complexed with Roscovitine and Olomoucine. These models show clear structural evidence for differences observed in the inhibition, and may help designing inhibitors for PfPK6 generating new potential drugs against malaria.

Key Words: Kinases, Bioinformatics, *Plasmodium falciparum*, Roscovitine and Olomoucine.



Introduction

Unicellular parasites such as *Plasmodium falciparum* (malaria), *Leishmania major* (leishmaniasis), *Trypanosoma brucei* (sleeping sickness), *Trypanosoma cruzi* (Chagas disease) and *Toxoplasma gondii* (toxoplasmosis) are responsible for the world's most widespread diseases. These microorganisms have genuine CDKs or CDK-related kinases. Several *Plasmodium falciparum* genes encoding cdc2-related protein kinases have been identified, but the modalities of their regulation remains largely unexplored [1,2].

Malaria remains a major and growing threat to the public health and economic development of countries in the tropical and subtropical regions of the world. Approximately 40% of the world's population lives in areas where malaria is endemic. There are an estimated 300–500 million cases and up to 2.7 million deaths from malaria each year. The mortality levels are greatest in sub-Saharan Africa, where children under 5 years of age account for 90% of all deaths due to malaria [1,2]. Human malaria is caused by infection with intracellular parasites of the genus *Plasmodium* (parasite) that is transmitted by *Anopheles gambiae* mosquitoes. Among four species of *Plasmodium* that infect humans, *Plasmodium falciparum* is the most lethal [3].

While the stages of parasitic infection are well documented, until recently, little has been known concerning the regulation of the parasitic life cycle. However, the identification of a family of kinases from the *P. falciparum* with a high degree of sequence conservation to the mammalian CDKs has allowed researchers to begin investigation of the mechanisms that control passage through the parasitic life cycle. *P. falciparum* Protein Kinase 5, PfPK5, was isolated in 1994 using oligonucleotides based on the sequences of CDK2 and CDK1 [4,5]. In a sequence comparison with members of the human CDK family, it shares the greatest degree of sequence identity about 60% with human CDK1 and CDK5. The similarity between PfPK5 and the mammalian CDKs in terms of their mechanism of kinase activation and inhibition has led researchers to investigate the effects of known small molecule CDK inhibitors on the growth of *P. falciparum* [4]. The PfPK5 structure was recently determined (PDB code: 1OB3) [6].



In addition to the observed sequence conservation, evidence suggests that PfPK5 is activated in a manner similar to the activation of the mammalian CDKs. First, research into the mechanism of PfPK5 activation demonstrates the necessity of Thr158 phosphorylation of PfPK5 for activation [7]. Similar phosphorylation events have been shown to be necessary for the activation of the mammalian CDKs. For example, activation of CDK2 requires phosphorylation of Thr160 [8]. Second, the *in vitro* activation of PfPK5 is enhanced as a result of co-incubation with the human cyclins [3]. Full activation of the mammalian CDKs requires the association of the catalytic kinases with a regulatory cyclin subunit [9]. These data strongly suggest a role for PfPKs in the life cycle of *P. falciparum* which is similar to that observed for CDK2 in the mammalian cell cycle [4].

PfPK6 was recently isolated by differential display RT-PCR (DDRT-PCR) of mRNA obtained from different asexual erythrocytic stages of *Plasmodium falciparum*, which shows sequence similarity to cyclin-dependent kinase (CDK) [10]. Over-expression of PfPK6 in *Escherichia coli* allowed studying the biochemical properties of the recombinant enzyme. In addition, immunolocalization and immunoblot analysis were performed to localize PfPK6 protein in various developmental stages [10].

Since there is no structural study about PfPK6, this article describes three-dimensional models of PfPK6 in complex with roscovitine and olomoucine [10], by applying computational homology modeling techniques and utilizing the high-resolution crystal structure of the CDK2 as template. It was made a comparison between the structural features of ATP-binding sites of PfPK6 and CDK2 in order to gain further insight into the structural basis for chemical inhibition of CDKs.

Methods

Molecular modeling

Homology modeling is usually the method of choice when there is a clear relationship of homology between the sequence of a target protein and at least one known structure. Model building of the complexes was carried out using the program Parmodel [11], which is a web server for automated modeling and protein structural assessment. Parmodel runs a parallelized version of



MODELLER [12]. MODELLER is an implementation of an automated approach to comparative modeling by satisfaction of spatial restraints. The modeling procedure begins with alignment of the sequence to be modeled (target) with the sequence of related known three-dimensional structures in complex with the inhibitor (templates). This alignment is usually the input to the program. The output is a three-dimensional model for the target sequence containing all mainchain and sidechain non-hydrogen atoms [12].

The homology models of PfPK6 with inhibitors were based on the atomic coordinates of CDK2 complexed with Roscovitine and Olomoucine. The alignment of CDK2 (template) and PfPK6 (target) is shown in Figure 1.

Figure 1

The modeling was performed in the presence of inhibitors from the template. These inhibitors were refined in the active site of the models. Several slightly different models can be calculated by varying the initial structure. A total of 1000 models were generated by the program Parmodel [10], the final model was select based on stereochemical quality. All optimization process was performed on a Beowulf cluster with 16 nodes (BioComp, AMD Athlon XP 2100+).

Analysis of the model

The overall stereochemical quality of the final models PfPK6 complexed with Roscovitine and Olomoucine was assessed by program PROCHECK [13]. The cutoff for hydrogen bonds and salt bridges was 3.4 Å. The contact surfaces for the binary complexes were calculated using AREAIMOL and RESAREA [14].

For superposition of C α we used the program LSQKAB from CCP4 [14]. The root mean square deviations (RMSD) from ideal geometries for bond lengths, bond angles, dihedrals and impropers were calculated with X-PLOR [15]. The program VERIFY-3D was used to assess structural quality of the homology models [16-17]. G-factor values were calculated using PROCHECK [13].

Results and Discussion

Quality of the model

The analysis of the Ramachandran plots for the templates and models indicates that over 90.0% of the residues are in the most favorable regions. Analysis of the structural quality of the homology models using PROCHECK [13], XPLOR [15] (for R.M.S.D. from ideal geometry), and VERIFY-3D [16,17] strongly indicates that the models are good enough for structural studies (Table 1).

Table 1

Overall description and Interactions with Inhibitors

The models for *Plasmodium falciparum* Protein Kinase complexed with inhibitors are folded into the typical bilobal structure, with the smaller N-terminal lobe consisting predominantly of β -sheet structure and the larger C-terminal lobe consisting primarily of α -helices. The N-terminal lobe of PfPK6, as observed for CDKs, consists of a sheet of five antiparallel β -strands (β 1– β 5) and a single large helix (α 1). The C-terminal lobe contains a pseudo-4-helical bundle (α 2,3,4,6), a small β -ribbon (β 6– β 8), and two additional helices (α 5,7). Figure 2 shows schematic drawings of the complex PfPK6:Roscovitine and PfPK6:Olomoucine. The Roscovitine and Olomoucine molecules are found in the cleft between the two lobes. The core (the β -sheet and the helical bundle) of the CDK1, CDK5 and PfPK6 structures are very similar to that of CDK2 [18-23].

Figure 2

The larger differences between the structures of CDK2 and PfPK6 are in the loops 85-94; 107-119; 164-174 and 237-262 of the regions of PfPK6, which are also observed in the alignment of primary sequences (Figure 1).

The specificity and affinity between enzyme and its inhibitor depend on directional hydrogen bonds and ionic interactions, as well as on shape complementarity of the contact surfaces of both partners [20-22,24-25]. Superpositions of the CDK2-ATP onto CDK1, CDK5 and PfPK6 structures complexed with roscovitine and olomoucine indicate that the two ring systems of roscovitine with ATP and olomoucine with ATP overlap approximately in the

same plane [20-23,26], however with different orientations. As observed in the crystallographic structures of CDK2:roscovitine and CDK2:olomoucine the region of CDK1, CDK5 and PfPK6 occupied by the phenyl ring of roscovitine and the phenyl ring of olomoucine are pointing away from the ATP-binding pocket, and partially exposed to solvent in the both complexes.

For the complexes CDK2:roscovitine and CDK2:olomoucine, the contact areas between inhibitor and CDK2 are 320 Å² and 269 Å², respectively. This analysis was used for comparison of the CDK1, CDK5 and PfPK6 complexed with the same ligands. The table 2 summarizes some structural results and IC50 values.

Table 2

Interactions of Roscovitine and Olomoucine with PfPK6

The active site of PfPK6 is structurally similar to the CDKs. Two hydrogen bonds between PfPK6 and inhibitors, involving the residue Leu95 were observed for the binary complexes. Table 3 shows the intermolecular hydrogen bonds for the structures of PfPK6, CDK1, CDK2 and CDK5. The figures 3 show the binding pocket for PfPK6:Roscovitine and PfPK6:olomoucine complexes, and the figure 4 shows the superposition among them, and CDK2 template.

Table 3

Figure 3a

Figure 3b

The contact areas for the complexes of PfPK6 with roscovitine and PfPK6 with olomoucine are 356 Å² and 297 Å², respectively (Table 1), compatible with the values observed to CDK2:roscovitine and CDK2:olomoucine complexes. However IC50 is higher for PfPK6 than the observed for CDK2. The structural basis for this higher IC50 relies on the presence of two tyrosine residues in the entry of the ATP-binding pocket observed in the complex PfPK6:roscovitine and PfPK6:olomoucine (Figures 3ab). This pair of tyrosines is not observed in the CDK2 complexes. We suggest that these tyrosines offer further hindrance to the

docking of roscovitine and olomoucine to the ATP-binding pocket of PfPK6, justifying the higher IC50 observed for the inhibition of PfPK6 by roscovitine and olomoucine (30 μ M and 180 μ M, respectively) [10], when compared with the inhibition of CDK2 by roscovitine and olomoucine (0.7 μ M and 7 μ M, respectively) [10]. The values of contact area to CDK1 and CDK5 are in agreement with IC50 values. The RMSD values from the ideal geometry of PfPK6 and CDKs 1, 2 and 5 are in table 4.

Figure 4

Table 4

Conclusion

Analysis of the charge distribution of the binding pockets indicates the presence of charge and shape complementarity between CDK1 complexed with roscovitine, CDK5 complexed with roscovitine and PfPK6 complexed with roscovitine and olomoucine. Figures 5a and b show the electrostatic potential surface of the structures of PfPK6 compared with CDK2 structure. Analysis of the present structural models indicates that roscovitine and olomoucine are more specific for CDK1, CDK2 and CDK5 than for PfPK6.

Figure 5a

Figure 5b

It was observed one significant difference in the PfPK6 models in complex with roscovitine and olomoucine. These models present a pair of tyrosines, which make a barrier for the ligand (figures 3ab). The presence of Tyr13 e Tyr96 in the entrance of the ATP-binding pocket reduces the volume available for ligand binding in the PfPK6 active site (Figure 4). The presence of this pair of tyrosines is the structural basis for the high values of IC50 for PfPK6 when compared with CDK1, CDK2 and CDK5. Further inhibition experiments may confirm this hypothesis.

Acknowledgments

This work was supported by grants from FAPESP (SMOLBNet 01/07532-0, 02/04383-7, 04/00217-0), CNPq, CAPES and Instituto do Milênio (CNPq-MCT). WFA (CNPq, 300851/98-7) is researcher for Brazilian Research Council (CNPq).

References

1. Doerig C, Meijer L, Mottram JC, (2002) *Trends Parasitol* 18:366-371
2. Ward P, Equinet L, Packer J, Doerig C (2004) *BMC Genomics* 5:79-97
3. Le Roch K, Sestier C, Dorin D, Waters N, Kappes B, Chakrabarti D, Meijer L, Doerig C (2000) *J Biol Chem* 275:8952-8958
4. Keenan SM, Welsh WJ (2004) *J Mol Graph Model*. 22:241-247
5. Gardner MJ *et al.* (2002) *Nature* 419:498-511
6. Holton S, Merckx A, Burgess D, Doerig C, Noble M, Endicott J (2003) *Structure* 11:1329-1337
7. Graeser R, Franklin RM, Kappes B (1996) *Mol Biochem Parasitol* 79:125-127
8. Gu Y, Rosenblatt J, Morgan DO (1992) *EMBO J* 11:3995-4005
9. Sherr CJ (1996) *Science* 274:1672-1677
10. Bracchi-Ricard V, Barik S, DelVecchio C, Doerig C, Chakrabarti R, Chakrabarti D (2000) *Biochem J* 347:255-263
11. Uchoa HB, Jorge GE, da Silveira NJ, Camera JC, Canduri F, De Azevedo WF (2004) *Biochem Biophys Res Commun* 325:1481-1486
12. Sali A, Blundell TL (1993) *J Mol Biol* 234:779-815
13. Laskowski RA, MacArthur MW, Smith DK, Jones DT, Hutchinson EG, Morris AL, Naylor D, Moss DS, Thornton JM (1994). PROCHECK v.3.0 - Program to check the stereochemistry quality of Protein structures - Operating instructions.
14. Collaborative Computational Project No. 4. (1994) *Acta Crystallogr D*50:760-763
15. Brünger AT (1992) *X-PLOR Version 3.1: A system for crystallography and NMR*. Yale University Press, New Haven.
16. Lüthy R, McLachlan AD, Eisenberg D (1991) *Proteins* 10:229-239
17. Bowie JU, Lüthy R., Eisenberg D (1991) *Science* 253:164-170
18. De Azevedo WFJr, Leclerc S, Meijer L, Havlicek L, Strnad M, Kim S-H (1997) *Eur J Biochem* 243:518-526
19. De Azevedo WFJr, Mueller-Dieckmann, HJ, Schulze-Gahmen U, Worland PJ, Sausville E, Kim S-H (1996) *Proc Natl Acad Sci USA*. 93:2735-2740
20. Canduri F, Uchoa HB, De Azevedo WFJr (2004) *Biochem Biophys Res Commun* 324:661-666
21. De Azevedo WFJr, Gaspar RT, Canduri F, Camera JCJr, da Silveira NJF (2002) *Biochem Biophys Res Commun* 297:1154-1158
22. Canduri F, De Azevedo WFJr (2005) *Cur Computer-Aided Drug Design* 1:53-64
23. Taylor SS, Radzio-Andzelm E (1994) *Structure* 2:345-55
24. De Azevedo WFJr, Canduri F, Fadel V, Teodoro LGVL, Hial V, Gomes RAS (2001) *Biochem Biophys Res Commun* 287:277-281
25. Canduri F, Teodoro LGVL, Lorenzi CCB, Hial V, Gomes RAS, Ruggiero Neto J, de Azevedo WFJr (2001) *Acta Crystallogr D*57:1560-1570



26. Schulze-Gahmen U, Brandsen J, Jones HD, Morgan DO, Meijer L, Vesely J, Kim S-H (1995) Proteins 4:378-391
26. Corpet F (1988) Nucl Acids Res 16:10881-10890
27. Koradi R, Billeter M, Wüthrich K (1996) J Mol Graphics 14:51-55
28. Nicholls A, Sharp KA, Honig B (1991) Proteins 11:281-296

Figure 3. Stereo view of binding pocket of the PIPK6 in complex with (a) roscovitine and (b) olomoucine.

Figure 4. Superposition of the complexes PIPK6-roscovitine (in pink) and PIPK6-olomoucine (in blue), with CDK2 (in light pink). In detail is shown the active site with the tyrosine residues.

Figure 5. Electrostatic potential surfaces for ATP binding pocket of a) PIPK6-roscovitine and b) PIPK6-olomoucine. Figures were generated with GRASP [28] shown from -10kT to +10kT. Uncharged regions are white. It is shown only part of inhibitor due to tyrosines in the entrance.

Figure 1. Sequence alignment of CDK2 with PfPK6 (39.7% of identity). The alignment was performed with the program MULTALIN [26].

Figure 2. Ribbon diagram of a) PfPK6 complexed with roscovitine and b) PfPK6:olomoucine. Figures were generated by MolMol [27].

Figure 3. Stereo view of binding pocket of the PfPK6 in complex with (a) roscovitine and (b) olomoucine.

Figure 4. Superposition of the complexes PfPK6:roscovitine (in pink) and PfPK6:olomoucine (in blue), with CDK2 (in light pink). In detail is shown the active site with the tyrosine residues.

Figure 5. Electrostatic potential surfaces for ATP binding pocket of a) PfPK6:roscovitine and b) PfPK6:olomoucine. Figures were generated with GRASP [28] shown from -10kT to +10kT. Uncharged regions are white. It is shown only part of inhibitor due to tyrosines in the entrance.

Complex	Most	Additional	Generously	Disallowed	3D	Q Factor	
	(%)	(%)	(%)	(%)			
CDK1_Ross	87.8	10.7	1.5	0.0	0.975	-0.23	-0.27
CDK1_Olo	87.8	10.7	1.5	0.0	0.975	-0.23	-0.27
CDK3_Ross	87.8	10.7	1.5	0.0	0.975	-0.23	-0.27
CDK3_Olo	87.8	10.7	1.5	0.0	0.975	-0.23	-0.27
PfPK6_Ross	81.3	6.3	2.0	0.4	0.808	-0.37	-0.19
PfPK6_Olo	81.3	6.7	1.6	0.4	0.769	-0.61	-0.12
CDK2_Ross	87.1	12.9	0.0	0.0	0.985	-0.45	-0.48
CDK2_Olo	89.6	10.4	0.0	0.0	1.065	-0.37	-0.05

*Ideally, scores should be above -0.5. Values below -1.0 may need investigation.

Table1: Analysis of the stereochemical quality, 3D PROFILE and Average G-factor values for the PfPK6 model.

Complex	Region of the Ramachandran Plot				Analysis Structural		
	Most Favorable (%)	Additional Allowed (%)	Generously Allowed (%)	Disallowed (%)	3D Profile*	G Factor	
CDK1_Rosc	87.8	10.7	1.5	0.0	0.97S	-0.23	-0.27
CDK1_Olo	95.0	4.6	0.4	0.0	0.89S	-0.26	-0.04
CDK5_Rosc	90.2	7.8	3.0	2.0	0.93S	-0.32	-0.07
CDK5_Olo	93.4	6.2	0.4	0.0	0.87S	-0.23	-0.03
PfPK6_Rosc	91.3	6.3	2.0	0.4	0.80S	-0.37	-0.18
PfPK6_Olo	91.3	6.7	1.6	0.4	0.76S	-0.61	-0.12
CDK2_Rosc	87.1	12.9	0.0	0.0	0.98S	-0.45	-0.48
CDK2_Olo	89.6	10.4	0.0	0.0	1.06S	-0.37	-0.05

*Ideally, scores should be above - 0.5. Values below -1.0 may need investigation.



Table2: Summary of structural and function results.

Complex	IC50 (μM)	Contact Area (Å ²)	Number of intermolecular hydrogen bonds
CDK1-Rosc	0.45	339	6
CDK1-Olo	7	285	2
CDK5-Rosc	0.16	334	5
CDK5-Olo	3	289	2
CDK2-Rosc	0.7	320	2
CDK2-Olo	7	269	2
PfPK6-Rosc	30	356	2
PfPK6-Olo	180	297	2

Table 3: Intermolecular hydrogen bonds.

Protein	Residues	Inhibitor	Distance (Å)
CDK1-Roscovitrine	Leu83N	N7	3.4
	Leu83O	N6	3.1
	Gln132NL2	O1	2.8
	Asp86OD2	N1	3.2
	Asp86OD2	N2	3.3
CDK2-Roscovitrine	Leu83N	N7	3.4
	Leu83O	N6	2.8
CDK5-Roscovitrine	Cys83N	N7	3.5
	Cys83O	N6	3.1
	Gln130NE2	O1	3.3
	Asp86OD2	N2	3.5
	Asp86OD2	N1	3.2
CDK1-Olomoucine	Leu83N	N7	2.9
	Leu83O	N6	2.6
CDK2-Olomoucine	Leu83N	N7	2.9
	Leu83O	N6	2.6
CDK5-Olomoucine	Cys83N	N7	2.9
	Cys83O	N6	2.6
PfPK6-Roscovitrine	Leu95N	N7	3.4
	Leu95O	N6	2.8
PfPK6-Olomoucine	Leu95N	N7	2.9
	Leu95O	N6	2.8



Molecular Models of the Cyclin-Dependent Kinase 4 associated the Cyclin D and Complexed

Table 4: Summary of the R.M.S.D. values from ideal geometry for the complexes.

Complex	Bonds lengths (Å)	Bond angles (degrees)	Dihedrals (degrees)
CDK1_Roscovitrine	0.021	2.273	24.404
CDK1_Olomoucine	0.019	2.728	23.396
CDK5_Roscovitrine	0.042	3.224	24.294
CDK5_Olomoucine	0.019	3.074	23.516
PfPK6_Roscovitrine	0.018	3.295	23.283
PfPK6_Olomoucine	0.021	3.328	23.403
CDK2_Roscovitrine	0.240	2.256	26.911
CDK2_Olomoucine	3.389	3.242	24.790

Universidade Estadual Paulista - UNESP

Departamento de Física

Rua Cristóvão Colombo, 2263

Código Postal: 13054-000

São José do Rio Preto - São Paulo - Brasil

Telefone: (017) 221-2403 - Fax: (017) 221-2247



Molecular Models of the Cyclin-Dependent Kinase 4 associated the Cyclin D and Complexed with Inhibitors

KARISA KARLA MANHANI, NELSON JOSÉ FREITAS DA SILVEIRA, HUGO BRANDÃO
UCHOA, WALTER FILGUEIRA DE AZEVEDO JR, FERNANDA CANDURI

KARISA KARLA MANHANI

*Programa de Pós-Graduação em Biofísica Molecular, Departamento de Física,
UNESP, 15054-000, Brasil*

NELSON JOSÉ FREITAS DA SILVEIRA

HUGO BRANDÃO UCHOA

WALTER FILGUEIRA DE AZEVEDO JR.

FERNANDA CANDURI (fcanduri@yahoo.com.br)

Abstract: Cancer is often regarded as a disease of uncontrolled cell growth, so it is not surprising that many of the cell cycle regulatory mechanisms have been found to be altered in tumors. It is now known that proteins that regulate proliferation are frequently mutated, deleted or over-expressed in cancer cell lines, and that this leads to the deregulated growth of the tumor cell. Overexpression of cyclin D is commonly observed in breast cancer, and mutation of cyclin D has been observed in melanomas. Such proteins therefore represent potential targets for therapeutic intervention. CDK4 associated with cyclin D has been modelled by homology in presence of inhibitors showing structural differences depending of the ligand.

*Universidade Estadual Paulista - UNESP
Departamento de Física*

Key Words: Abstract, Cyclin D, Kinase 4, Potential Targets

Código Postal: 15054-000

São José do Rio Preto – São Paulo – Brasil

Telephone: (017) 221-2463 - Fax: (017) 221-2247



Molecular Models of the Cyclin-Dependent Kinase

4 associated the Cyclin D and Complexed with

Inhibitors

KARISA KARLA MANHANI, NELSON JOSÉ FREITAS DA SILVEIRA, HUGO BRANDÃO UCHÔA, WALTER FILGUEIRA DE AZEVEDO JR, FERNANDA CANDURI

Programa de Pós-Graduação em Biofísica Molecular. Departamento de Física, UNESP, São José do Rio Preto, SP 15054-000, Brasil

Abstract: Cancer is often regarded as a disease of uncontrolled cell growth, so it is not surprising that many of the cell cycle regulatory mechanisms have been found to be altered in tumors. It is now known that proteins that regulate proliferation are frequently mutated, deleted or over-expressed in cancer cell lines, and that this leads to the deregulated growth of the tumor cell. Overexpression of cyclin D is commonly observed in breast cancer, and mutation of cdk4 has been observed in melanomas. Such proteins therefore represent potential targets for therapeutic intervention. CDK4 associated with cyclin D has been modelled by homology in presence of inhibitors showing structural differences depending of the ligand.

Key Words: *Kinases, Cyclins, Inhibitors, Cancer and Potential Targets.*



Introduction

The cyclin-dependent kinases are a family of serine/threonine kinases that are important in controlling entry into and transition through each phase of the cell cycle. Its cellular activity is tightly regulated through a number of mechanisms: phosphorylation by kinases and dephosphorylation by phosphatases; the binding of specific activating proteins called cyclin; and inhibitors. These inhibitors must be either sequestered or destroyed, and the CDK appropriately phosphorylated and bound to its cyclin, before the enzyme is fully active. The synthesis and degradation of the activating cyclins is tightly controlled; this, together with their specificity for a particular CDK partner, causes first one, and then another member of the CDK family to become active and drive progression through the cell cycle [1].

A member of the family of CDKs initiates each of the transitions. In response to the interaction of a mitogenic factor (e.g., epidermal growth factor), with its receptor at the outside of the cell, the signalling pathway to the nucleus results in expression of the D-type cyclins that associate with and activate the CDK4 and CDK6. These CDK complexes phosphorylate the tumour suppressor protein pRb and promote the expression of cyclin E. Cyclin E associates with CDK2 to drive cells from the G₁ to S phase through phosphorylation of a limited number of targets, including pRb. Phosphorylation of pRb releases the restraining activity on the transcription factor E2F. On entry into the S phase, cyclin E is ubiquitinated and abruptly destroyed by the proteasome. Cyclin A, expressed in response to the CDK2/cyclin E activities, associates with CDK2 to drive cells through the S phase. CDK2/cyclin A phosphorylates a large number of target proteins, including pRb, p107, transcription factors, regulators of transcription factors, and pre-replication complexes. Cyclin A is present throughout G₂, and associates with CDK1 during the transition from G₂ to M phase, after which it is abruptly degraded. Cyclin B then associates with CDK1 to drive cells through the M phase in concert with other kinases and downstream regulator enzymes [2].

Recent studies have also suggested that CDK4/Cyclin D can function to promote growth. In *Drosophila*, CDK4/Cyclin D overexpression in postmitotic cells of the developing eye causes cell enlargement, suggesting that CDK4/Cyclin D can promote growth and lead to hypertrophy in non-dividing cells. In contrast, overexpression of CDK4/Cyclin D in the proliferating wing discs leads to



increased growth and increased rate of cell proliferation, with normal cell size and cell-cycle phasing (hyperplasia) [3].

Tumour expression of all nuclear proteins was scored according to the percentage of positive cancer nuclei. Positive p53 expression was detected in 27/53 (50.94%) cases. Lack of pRb immunostaining was observed in 39/53 (73.58%) cases. Overexpression of Cyclin D₁ was shown in 21/53 (39.62%) tumours. The overexpression of CDK4 was detected in 43/53 (81.13%) cases. This implies that the aberration of an important cell cycle regulator may be sufficient to disrupt regulatory mechanism in a manner favouring tumourigenesis [4]. The CDKs are increasingly recognized as important targets for therapeutic intervention in various proliferative disease states [5].

Crystallographic studies of human CDK2 have shown that the CDK subunit is inactive when monomeric and unphosphorylated, because a conformational change is necessary to provide a correct access to the binding sites within the catalytic cores [6]. Until cyclin binding, the substrate binding site is blocked by a domain called T loop that prevents binding of the target protein and access to ATP [7].

Flavopiridol, a inhibitor, is a potential anti-cancer therapeutic agent currently being tested in phase I and II clinical trials because of its antiproliferative properties [8]. Recombinant CDK4-cyclin D1 was inhibited potently by Flavopiridol (K_i, 65 nM), competitive with respect to ATP [9]. Roscovitine is a potent CDK inhibitor and was tested against various CDKs, with IC₅₀ values ranging from 0.16 μ M (for CDK5) [10], 0.7 μ M (for CDK2) [11,12], to over 100 μ M (for CDK4) [11,12].

Since there is not structural study about CDK4/Cyclin D, this article describes the modeling of the human CDK4/Cyclin D in presence of ten inhibitors: flavopiridol, H717, hymenialdisine, indirubin-5-sulfonate, N-isopentenyladenine, olomoucine, oxindole 91, purvalanol B, roscovitine, and staurosporine.

Methods

Molecular modeling

Homology modeling is usually the method of choice when there is a clear



relationship of homology between the sequence of a target protein and at least one known structure. Model building of the complexes were carried out using the program Parmodel [13], which is a web server for automated modeling and protein structural assessment. Parmodel runs a parallelized version of MODELLER [14]. MODELLER is an implementation of an automated approach to comparative modeling by satisfaction of spatial restraints. The modeling procedure begins with alignment of the sequence to be modeled (target) with the sequence of related known three-dimensional structures (templates). This alignment is usually the input to the program. The output is a three-dimensional model for the target sequence containing all mainchain and sidechain non-hydrogen atoms [14].

The models of CDK4/Cyclin D were based on the atomic coordinates of CDK2/Cyclin A (PDB code: 1H1Q) [5]. This crystallographic structure was solved in presence of the inhibitor v6094. The choice of ligands for this work was based in the IC50 values. The coordinates of these ligands were obtained in the Data Bank Protein (PDB) (Table 1). Firstly was necessary to prepare the templates through of the superposition of C α of the CDK2/Cyclin A (PDB code 1H1Q) with the CDK2 structure bound with the inhibitor of interesting, keeping the atomic coordinates of 1H1Q fixed. An energy minimization of the templates associated with Cyclin A and complexed with inhibitors was performed. These templates were used to model the CDK4/Cyclin D complexed with inhibitors (table 1). This table relates the inhibitors and IC50 values.

Table 1

Several slightly different models can be calculated by varying the initial structure. A total of 1000 models were generated by the program Parmodel [13], the final model was selected based on stereochemical quality. All optimization process was performed on a Beowulf cluster with 16 nodes (BioComp, AMD Athlon XP 2100+).

Analysis of the model

The overall stereochemical quality of the final models CDK4/Cyclin D complexed with inhibitors was assessed by program PROCHECK [15]. The cutoff



for hydrogen bonds and salt bridges was 3.1 Å. The contact surfaces for the binary complexes were calculated using AREAIMOL and RESAREA [16].

In the superposition of Cα was used the program LSQKAB from CCP4 [16]. The root mean square deviations (RMSD) from ideal geometries for bond lengths, bond angles, dihedrals and impropers were calculated with X-PLOR [17]. The program VERIFY-3D was used to assess structural quality of the homology models [16]. G-factor values were calculated using PROCHECK [15].

Results and Discussion

Quality of the model

The analysis of the Ramachandran plots for the templates and models indicates that over 91.7% and 92.5% of the residues are in the most favorable regions, respectively. Analysis of the structural quality of the models using XPLOR [17] for r.m.s.d. from ideal geometry was obtained for CDK4/Cyclin D. Values of bonds lengths, bond angles, dihedrals and impropers are 0.598 Å; 2.709°; 22.394° and 2.631°, respectively, in comparison with values for CDK2/Cyclin A, 0.605 Å; 3.525°; 22.879° and 2.668°, respectively. Values of VERIFY-3D [18] were obtained for the model. The quality calculated of the S = Total Score/Ideal Score obtained from S Score was 0.92. The cutoff values is 0.45. Values above of 0.45 strongly indicate that the models are good enough for structural studies.

Overall description and Interactions with Inhibitors

The models for Cyclin Dependent Kinase 4 complexed with ten inhibitors are folded into the typical bilobal structure, with the smaller N-terminal lobe consisting predominantly of β-sheet structure and the larger C-terminal lobe consisting primarily of α-helices. The N-terminal lobe of CDK consists of a sheet of five antiparallel β-strands (β1–β5) and a single large helix (α1). The C-terminal lobe contains a pseudo-4-helical bundle (α2,3,4,6), a small β-ribbon (β6–β8), and two additional helices (α5,7). The inhibitor molecule is found in the cleft between the two lobes. The core (the β-sheet and the helical bundle) of the CDK4 structure is very similar to that of CDK2 [8;19-21], as shows the figure 2a.

The minor differences between the structures of CDK2 and CDK4 are in N-terminal (residues 1-2) and C-terminal (residues 301-303). The CDK4 sequence

is longer than CDK2. The larger differences are in the loops 46-48; 65-67 and 81-85 of CDK4, which can be observed in the alignment of primary sequences (Figure 1 a).

Figure 1a

Figure 1b

The specificity and affinity between enzyme and its inhibitor depend on directional hydrogen bonds and ionic interactions, as well as on shape complementarity of the contact surfaces of both partners [11;22]. Superpositions of the CDK2-ATP onto CDK4 complexes indicate that the two ring systems of inhibitors overlap with ATP approximately in the same plane [8;23-24], however with different orientations. As observed in the crystallographic structures of CDK2-inhibitor the region of CDK4 occupied by the phenyl ring of inhibitors are pointing away from the ATP-binding pocket, and partially exposed to solvent in both complexes.

The contact areas between inhibitors and enzyme, IC50 values and number of hydrogen bonds are summarized in table 2. This analysis was used for comparison of the CDK4 complexes with their templates. The contact areas for the complexes of CDK4 are compatible with the values observed to CDK2-inhibitors complexes. Table 3 shows the intermolecular hydrogen bonds for the structures of CDK4 and CDK2 with the inhibitors.

TABLE 2

TABLE 3

Description and Interactions with Cyclins

The activation of cyclin-dependent kinase 2 (CDK2) structure by phosphorylation and association with cyclin A presents conformational changes that can be equally observed in the CDK4 model associated with cyclin D [1-2;5].

The alignment between cyclin A and cyclin D, was performed as shown in figure 1b, presenting 29.2% of identity. The main differences are in the N-terminal (residues 1-15) and three loops (21-24; 204-207 and 247-250) observed in the sequence of cyclin D.

In the CDK4/cyclin D model was notable that the N-terminal lobe (residues 1-90) is comprised mostly of β -pleated sheet with one helix, the C helix or PISTVRE helix, where in CDK2/Cyclin A structure the β -pleated sheet are the residues 1-79, the C helix is in PSTAIRE region and both contains a Gly-rich loop between β strand 1 and 2. The N-terminal lobe is joined to the C-terminal lobe by a hinge region (residues 91-96 in CDK4 and residues 80-85 in CDK2). The C-terminal lobe (residues 97-301 in CDK4 and residues 86-294 in CDK2) is mostly α helical and contains the activation segment from residues 156-182 in CDK4 and residues 145-172 in CDK2, between the conserved DFG and APE motifs in both model and template. The activation segment is important in the substrate binding and carries the phosphorylated residue Thr170 for CDK4/Cyclin D and residue Thr160 for CDK2/Cyclin A both conserved, as shown in figures 2a and b. In the inactive non-phosphorylated CDK2, the two lobes are held in an open conformation and the activation segment partially occludes the catalytic site. Phosphorylation of Thr160 in monomeric CDK2 results in displacement and some disordering of the activation segment, but no major change in the structure [2]. Association of CDK2 with cyclin A leads to major changes in structure of CDK2, but without phosphorylation, there is still only basal activity [2;5].

Figure 2 a

Figure 2 b

On the association of CDK4 with Cyclin D, the helix both rotates and shifts to make more intimate contact with the N-terminal lobe and to bring the key residues for ATP phosphate binding (Lys33, where in both structures were conserved), and Glu54 (Glu51 in CDK2/Cyclin A structure) into their correct positions. Both CDK4 and CDK2 are pliable, but Cyclin D and Cyclin A are rigid. The final component of the activation process is the combination of phosphorylation of Thr170 for CDK4/Cyclin D and Thr160 for CDK2/Cyclin A.

Anderson et al., (2003) suggested a hybridization strategy between CDK2 and CDK4. These authors have substituted the binding site residues of the CDK2 by residues of the CDK4, based in functional studies with one selective class of inhibitors: the imidazo[1,2-*a*]pyridines [1]. This hybridized structure was used to compare with the active site of the CDK4 model in presence of cyclin D.

with the hybridized structure due to conformational change of T-loop (figure 3)

The major difference observed in CDK4/Cyclin D model in comparison with the structure-based hybridization (code PDB 1OIR) [1] is the presence of differences in the surface of the binding site (CDK2/Lys9 for CDK4/Glu11 and His84 for Asp97). They were realized superposition of C α from LSQKAB (CCP4) [16] between CDK4 model and hybridized structure, and were obtained r.m.s.d. values of 3.270 Å for all C α and 3.998 Å for the T-loop. These high r.m.s.d. values are related to conformation change due to the presence of the cyclin D in the model. It can be observed a movement of the T-loop, opening the active site (figure 3).

Figure 3

The contact area between CDK4 and cyclin D is 1454 Å², compatible with the value observed to CDK2/cyclin A (1654 Å²). Figures 4a and b show the electrostatic potential surface for interface between CDK4 and cyclin D model compared with CDK2 and cyclin A structures. The charge distribution and the shape observed in the interface of both template and model indicates a complementarity between CDK and cyclin.

Figure 4 a

Figure 4 b

Conclusion

Analysis of the charge distribution of the binding pocket indicates that this site is strongly negative, and there are shape complementarity between CDK4 and ligands. Analysis of the structural models associated with IC50 values and contact area indicates that Flavopiridol, Hymedialdisine, Indirubin-5-sulfonate, Oxindole91 and Staurosporine are more specific for CDK4 than the others. Flavopiridol has been in phase 2 of clinical trial [25]. Based in these facts, new inhibitors for CDK4 can be designed.

The interface between CDK4 with Cyclin D shows a shape and charge complementary, as observed in the CDK2/cyclin A. It was observed some significant differences in the CDK4 associated with cyclin D models compared with the hybridized structure due to conformational change of T-loop (figure 3).

Acknowledgments

This work was supported by grants from FAPESP (SMOLBNet 01/07532-0, 02/04383-7), CNPq, CAPES and Instituto do Milênio (CNPq-MCT). WFA (CNPq, 300851/98-7) is researcher for the Brazilian Council for Scientific and Technological Development, and FC is post-doctoral fellow under FAPESP fellowship.

References

- 1 Anderson M, Beattie JF, Breault GA, Breed J, Byth KF, Culshaw JD, Ellston RPA, Green S., Minshull CA, Norman RA, Pauptit RA, Satanway J, Thomas AP, Jewsbury PJ (2003) *Bioorg Med Chem Lett* 13:3021-3026
- 2 Johnson LN, De Moliner E, Brown NR, Song H, Barford D, Endicott JA, Noble MEM (2002) *Pharmacol Ther* 93:113-124
- 3 Xin S, Weng L, Xu J, Du W (2002) *Development* 129:1345-1356
- 4 Koontongkaew S, Chareonkitkajorn L, Chanvitan A, Leelakriangsak M (2000) *Oral Oncol* 36:334-339
- 5 Davies TG, Bentley J, Arris CE, Boyle FT, Curtin NJ, Endicott JA, Gibson AE, Golding BT, Griffin RJ, Hardcastle IR, Jewsbury P, Johnson LN, Mesguiche V, Newell DR, Noble ME, Tucker JA, Wang L, Whitfield HJ (2002) *Nat Struct Biol* 9:745-749
- 6 Jeffrey PD, Russo AA, Polyak K, Gibbs E, Hurwitz J, Massague J, Pavletich NP (1995) *Nature* 376:313-320
- 7 Coqueret O (2002) *Gene* 299:35-55
- 8 Senderowicz AM, Sausville EA (2000) *J Natl Cancer Inst* 92:376-387
- 9 Carlson BA, Dubay MM, Sausville EA, Brizuela L, Worland PJ (1996) *Cancer Res* 56:2973-2978
- 10 Meijer L, Borgne A, Mulner O, Chong JP, Blow JJ, Inagaki N (1997) *Eur J Biochem* 243:527-536
- 11 De Azevedo WFJR, Leclerc S, Meijer L, Havlicek L, Strnad M, Kim S-H (1997) *Eur J Biochem* 243:518-526
- 12 Knockaert M, Greengard P, Meijer L (2002) *Trends Pharmacol Sci* 23:417-25
- 13 Uchoa HB, Jorge GE, da Silveira NJ, Camera JC, Canduri F, De Azevedo WF (2004) *Biochem Biophys Res Commun* 325:1481-1486
- 14 Sali A, Blundell TL (1993) *J Mol Biol* 234:779-815
- 15 Laskowski RA, MacArthur MW, Smith DK, Jones DT, Hutchinson EG, Morris AL, Naylor D, Moss DS, Thorton JM (1994) PROCHECK v.3.0 - Program to check the stereochemistry quality of Protein structures - Operating instructions.
- 16 Collaborative Computational Project No. 4. (1994) *Acta Crystallogr* D50:760-763
- 17 Brünger AT (1992) X-PLOR Version 3.1: A system for crystallography and NMR. Yale University Press, New haven.
- 18 Bowie JU, Lüthy R, Eisenberg D (1991) *Science* 253:164-170
- 19 Schulze-Gahmen U, Brandsen J, Jones HD, Morgan DO, Meijer L, Vesely J, Kim S-H (1995) *Proteins* 4:378-391



- 20 Canduri F, Uchoa HB, De Azevedo WFJr (2004) *Biochem Biophys Res Commun* 324:661-666
- 21 Canduri F, De Azevedo WFJr (2005) *Current Computer-Aided Drug Design* 1:53-64
- 22 Canduri F, Teodoro LGVL, Lorenzi CCB, Hial V, Gomes RAS, Ruggiero Neto J, De Azevedo WFJr (2001) *Acta Crystallogr D* 57:1560-1570
- 23 Meijer L, Raymond E (2003) *Acc Chem Res* 36:417-425
- 24 De Azevedo, WFJr, Mueller-Dieckmann HJ, Schulze-Gahmen U, Worland PJ, Sausville E, Kim S-H (1996) *Proc Natl Acad Sci USA* 93:2735-2740
- 25 Fabian MA, Biggs WH 3rd, Treiber DK, Atteridge CE, Azimioara MD, Benedetti MG, Carter TA, Ciceri P, Edeen PT, Floyd M, Ford JM, Galvin M, Gerlach JL, Grotzfeld RM, Herrgard S, Insko DE, Insko MA, Lai AG, Lelias JM, Mehta SA, Milanov ZV, Velasco AM, Wodicka LM, Patel HK, Zarrinkar PP, Lockhart DJ (2005) *Nat Biotechnol* 23:329-336
- 26 Corpet F (1988) *Nucl Acids Res* 16:10881-10890
- 27 Koradi R, Billeter M, Wüthrich K (1996) *J Mol Graphics* 14:51-55
- 28 McRee DE (1999) *J Struct Biol* 125:156-165
- 29 Nicholls A, Sharp KA, Honig B (1991) *Proteins* 11:281-296
- 30 De Azevedo WFJr, Mueller-Dieckmann HJ, Schulze-Gahmen U, Worland PJ, Sausville E, Kim S-H (1996) *Proc Natl Acad Sci USA* 93:2735-2740
- 31 Dreyer MK, Borchering DR, Dumont JA, Peet NP, Tsay JT, Wright PS, Bitonti AJ, Shen J, Kim S-H. (2001) *J Med Chem* 44:524-530
- 32 Meijer L, Thunnissen AM, White AW, Garnier M, Nikolic M, Tsai LH, Walter J, Cleverley KE, Salinas PC, Wu YZ, Biernat J, Mandelkow EM, Kim S-H, Pettit GR (2000) *Chem Biol* 7:51-63
- 33 Hoessel R, Leclerc S, Endicott JA, Nobel ME, Lawrie A, Tunnah P, Leost M, Damiens E, Marie D, Marko D, Niederberger E, Tang W, Eisenbrand G, Meijer L (1999) *Nat Cell Biol* 1:60-67
- 34 Davis ST, Benson BG, Bramson HN, Chapman DE, Dickerson SH, Dold KM, Eberwein DJ, Edelstein M, Frye SV, Gampe RTJr, Griffin RJ, Harris PA, Hassell AM, Holmes WD, Hunter RN, Knick VB, Lackey K, Lovejoy B, Luzzio MJ, Murray D, Parker P, Rocque WJ, Shewchuk L, Veal JM, Walker DH, Kuyper LF (2001) *Science* 291:134-137
- 35 Gray NS, Wodicka L, Thunnissen AM, Norman TC, Kwon S, Espinoza FH, Morgan DO, Barnes G, LeClerc S, Meijer L, Kim S-H, Lockhart DJ, Schultz PG (1998) *Science* 281:533-538
- 36 Lawrie AM, Noble ME, Tunnah P, Brown NR, Johnson LN, Endicott JA (1997) *Nat Struct Biol* 4:796-801



Figure 1: Sequence alignment of a) CDK2 with CDK4 (45.2% of identity) and b) Cyclin A with Cyclin D (29.2% of identity). The alignment was performed with the program MULTALIN [26].

Figure 2: Ribbon diagram of a) CDK4 associated with Cyclin D and complexed with Flavopiridol; b) Conformational change of the T-loop and activation segment. Figure was generated by MolMol [27]

Figure 3: Stereo view of the superposition of the CDK4/Cyclin A with hybridized structure showing the T170 (T160 in CDK2) of the T-loop. It can be observed the conformation change between then. Thick line is CDK4 and thin line is hybridized CDK2. Figures were generated by XtalView [28].

Figure 4: Electrostatic potential surfaces for a) interface CDK4/Cyclin D, and b) interface CDK2/Cyclin A. Figures were calculated with GRASP [29] shown from -10kT to +10kT. Uncharged regions are in white.

Table 1: Inhibitors of CDKs obtained from data base.

Inhibitors	PDB Code	References
Flavopiridol	ND	30
H717	1G5S	31
Hymenialdisine	1DM2	32
Indirubin-5-sulfonate	1E9H	33
N-Isopentenyladenine	ND	20
Olomoucine	ND	20
Oxindole 91	1FVV	34
Purvalanol B	1CKP	35
Roscovitine	ND	11
Staurosporine	1AQ1	36

ND: inhibitors were not deposited in Protein Data Bank.

Table 2: Summary of the data obtained for CDK4:inhibitors.

Inhibitors	IC50 (μ M)	Contact Area ($\text{\AA}^2$)	Number of intermolecular hydrogen bonds
Flavopiridol	0.4	319 (324)	4 (2)
H717	2	341 (362)	2 (2)
Hymenialdisine	0.6	231 (233)	6 (3)
Indirubin-5-sulfonate	0.3	265 (250)	7 (3)
N-Isopentenyladenine	>100	214 (224)	3 (2)
Olomoucine	>1000	281 (291)	4 (2)
Oxindole 91	0.13	283 (284)	3 (3)
Purvalanol B	>10	233 (242)	2 (2)
Roscovitine	>100	334 (360)	2 (2)
Staurosporine	<10	338 (330)	6 (2)

In parenthesis are the values obtained for templates used in the modeling.

Table 3: Intermolecular hydrogen bonds between CDK4 and inhibitors.

Inhibitor		Residues	Distance (Å)
Flavopiridol	O4	Val 94N (Leu 83N)	2.6 (2.7)
	O5	Glu 92O (Glu 81O)	2.9 (2.9)
	O4	His 93ND1	3.1
	O7	Lys 33NZ	3.5
H717	N7	Val 94N (Leu 83N)	2.6 (3.1)
	N18	Val 94O (Leu 83O)	2.5 (3.0)
Hymenialdisine	O1	Val 94N (Leu 83N)	2.5 (2.7)
	N1	Val 94O (Leu 83O)	3.0 (3.2)
	O1	His 93ND1	2.8
	N2	Glu 92O (Glu 81O)	2.7 (2.8)
	O2	Lys 33 NZ	2.9
	N3	Lys 33NZ	3.1
Indirubin-5-sulfonate	O13	Val 94N (Leu 83N)	2.5 (2.7)
	N7	Val 94O (Leu 83O)	2.8 (3.1)
	N14	His 93ND1 (Glu 81O)	2.9 (3.0)
	O13	His 93ND1	2.9
	O24	Lys 33 NZ	2.8
	O22	Lys 33NZ	2.8
	O23	Asp 156N	3.0
N-Isopentenyladenine	N3	Val 94N (Leu 83N)	2.9 (3.4)
	N9	Glu 92O (Glu 81O)	3.1 (2.8)
	N3	Val 94O	3.1
Olomoucine	N7	Val 94N (Leu 83N)	2.5 (2.9)
	N6	Val 94O (Leu 83O)	2.5 (2.6)
	OH	Asp 97 OD2	3.0
	N7	Val 94O	3.4
Oxindole 91	O1	Val 94N (Leu 83N)	2.2 (3.0)
	N15	Val 94O (Leu 83O)	2.7 (3.1)
	N3	Glu 92O (Glu 81O)	3.1 (2.9)
Purvalanol B	N7	Val 94N (Leu 83N)	2.6 (3.2)
	N6	Val 94O (Leu 83O)	2.5 (2.5)
Roscovitine	N7	Val 94N (Leu 83N)	2.8 (3.4)
	N6	Val 94O (Leu 83O)	2.6 (2.8)
Staurosporine	O5	Val 94N (Leu 83N)	2.7 (2.8)
	N1	Glu 92O (Glu 81O)	2.9 (2.7)
	O5	His 93ND1	2.9
	O5	Val 94O	2.4
	O6	Glu 142O	3.1
	N4	Asp 97OD2	3.0

In parenthesis are the values obtained for templates used in the modeling



