

Renato Tadeu Gaspar

Modelagem Molecular da Quinase
Dependente de Ciclina 5 Complexada
com Roscovitina.



1910001169



Renato Tadeu Gaspar

**MODELAGEM MOLECULAR DA QUINASE
DEPENDENTE DE CICLINA 5 COMPLEXADA
COM ROSCOVITINA.**

Dissertação para obtenção do título de
Mestre em Biofísica Molecular, área de
concentração em Biofísica Molecular,
apresentada ao Departamento de Física do
Instituto de Biociências Letras e Ciências
Exatas da Universidade Estadual Paulista
"Júlio de Mesquita Filho" - Unesp.

Orientador: Prof. Dr. *Walter Filgueira de Azevedo Jr.*

São José do Rio Preto - SP
2003



1.169/03

577.23
G249m

20804008

UNIVERSIDADE ESTADUAL PAULISTA
IBRCE - Câmpus São José do Rio Preto



Renato Tadeu Gaspar

MODELAGEM MOLECULAR DA QUINASE
DEPENDENTE DE CICLINA 5 COMPLEXADA
COM ROSCOVITINA

Gaspar, Renato Tadeu.

Modelagem molecular da quinase dependente de ciclina 5
complexada com roscovitina / Renato Tadeu Gaspar – São José do Rio
Preto : [s.n.], 2003
85 f. : il. ; 30 cm.

Orientador: Walter Filgueira de Azevedo Júnior
Dissertação (mestrado) – Universidade Estadual Paulista. Instituto de
Biociências, Letras e Ciências Exatas

1. Biologia molecular. 2. Bioinformática. 3. Modelagem molecular.
4. Quinase. 5. Quinase - Inibidores. 6. Inibidores químicos. I. Azevedo
Júnior, Walter Filgueira. II. Universidade Estadual Paulista. Instituto de
Biociências, Letras e Ciências Exatas. III. Título.

CDU – 577.32



Agradecimentos

Ao Prof. Dr. Walter Figueira de Azevedo Jr., pela orientação, paciência e apoio necessários para a execução desse trabalho e por ter me recebido em seu grupo, possibilitando assim um voto de confiança.

A Dra. Fernanda Centuri e ao Prof. Dr. Jorge Chahina, pela participação no Exame Geral de Qualificação.

Aos Profs. Ami e José Ruggiero pela atenção e compreensão durante minha busca pelo caminho que deveria seguir.

Aos demais professores do Departamento de Física pela dedicação, colaboração e incentivo.

Aos funcionários do Departamento de Física pelo companheirismo que sempre proporcionou momentos muito agradáveis.

Aos funcionários da Biblioteca pelo auxílio sempre paciente nas mais variadas situações.

Aos funcionários da seção de Pós-Graduação pelo pronto atendimento e honestas necessidades e principalmente à Rosimar pela paciência e atenção.

Aos colegas do Laboratório de Sistemas Biométricos, Denis, Fernanda, Guilherme, Henrique, Hugo, João Carlos, Luciano, Márcio, Nelson, e Patrícia, pela ajuda e pelos bons momentos que passamos juntos.

Aos amigos, Fausto, Ricardo, Sabrina e Sérgio, pela presença, amizade e apoio nos momentos difíceis da minha vida de estudante.

Aos amigos conselheiros, Giovanni, Irene, Leandra, e Márcia, por me apoiar, ajudar e principalmente me emprestar um pouco de juízo sempre que necessário.

A Paeli pela amizade que, com tão pouco tempo, parece de uma vida inteira.

Aos demais discentes do Departamento de Física, pela convivência sempre harmoniosa.

Ao meu amigo, Dr. Marcelo Sávio, pelas horas de conversa e conselhos que foram tão importantes para a superação das principais dificuldades.

**Aos meus pais,
Luiz e Joana.**

A minhas amigas Joana D'Art e Maria Helena pela força, carinho e constante incentivo, que fazem da vida um verdadeiro prazer.

A Eliane por saber comigo essa jornada, sempre ao meu lado, por tantos anos, com amor e paciência.

Agradecimentos...

Ao Prof. Dr. Walter Filgueira de Azevedo Jr., pela orientação, paciência e apoio necessários para a execução desse trabalho e por ter me recebido em seu grupo, depositando em mim um voto de confiança.

A Dra. Fernanda Canduri e ao Prof. Dr. Jorge Chahine, pela colaboração no Exame Geral de Qualificação.

Aos Profs. Arni e José Ruggiero pela atenção e compreensão durante minha busca pelo caminho que deveria seguir.

Aos demais professores do Departamento de Física pela dedicação, colaboração e incentivo.

Aos funcionários do Departamento de Física pelo companheirismo que sempre proporcionou momentos muito agradáveis.

Aos funcionários da Biblioteca pelo auxílio sempre paciente nas mais variadas situações.

Aos funcionários da seção de Pós-Graduação pelo pronto atendimento a nossas necessidades e principalmente à Rosemar pela paciência e atenção.

Aos colegas do Laboratório de Sistemas Biomoleculares, Denis, Fernanda, Guilherme, Henrique, Hugo, João, Karisa, Luciana, Márcio, Nelson, e Patrícia, pela ajuda e pelos bons momentos que passamos juntos.

Aos amigos, Fausto, Ricardo, Sabrina e Sérgio, pela presença, amizade e apoio nos momentos complicados da nossa vida de estudante.

Aos amigos conselheiros, Giovanni, Irene, Leandra, e Márcia, por me apoiar, escutar e principalmente me emprestar um pouco de juízo sempre que necessário.

À Priscilla pela amizade que, com tão pouco tempo, parece de uma vida inteira.

Aos demais discentes do Departamento de Física, pela convivência sempre harmoniosa.

Ao meu amigo, Dr. Manoel Simão, pelas horas de conversa e conselhos, que foram tão importantes para a superação dos principais obstáculos.

A minhas amigas Joana D'Arc e Maria Helena pela força, carinho e constante incentivo, que Deus há de devolver em dobro.

À Eliane por trilhar comigo esse caminho, sempre ao meu lado, por tantos anos, com amor e paciência.



Aos meus familiares, por acreditarem nas minhas escolhas, fornecendo o apoio necessário para que os objetivos sejam sempre alcançados.

À CAPES pelo auxílio financeiro.



ÍNDICE GERAL

SUMÁRIO	1
ABSTRACT	3
1 - INIBIÇÃO DAS CDKs	5
1.1 - O que são as CDKs	5
1.2 - Função das CDKs	9
1.3 - Porque inibir as CDKs	12
1.4 - Mecanismos de Inibição	15
1.5 - A CDK5	17
1.6 - Sensitividade	19
1.7 - Objetivos	19
2 - MÉTODOS	20
2.1 - Construção do modelo	23
2.2 - Inibição da atividade	24
3 - RESULTADOS E DISCUSSÃO	26
3.1 - Qualidade dos Modelos	27
3.2 - Inibição do modelo	32
3.3 - O Coração Molecular	33
3.4 - Função das cadenas laterais	34
3.5 - Interações da rede	35
4 - CONCLUSÕES	46
REFERÊNCIAS	48
APÊNDICE A - MODELAGEM COMPARATIVA DE ESTRUTURAS	53
(1) modelo	55

Quando eu for incapaz de praticar o mal,
quando nenhuma palavra áspera ou arrogante abalar,
por um momento sequer, o meu mundo mental,
só então, e não antes,
a minha não-violência conquistará o coração de todos.

Gandhi

ÍNDICE GERAL

RESUMO.....	1
ABSTRACT.....	3
1 - INIBIÇÃO DAS CDKS	5
1.1 - O que são as CDKs.....	5
1.2 - Função das CDKs.....	9
1.3 - Porquê inibir as CDKs.....	13
1.4 - Mecanismos de inibição	15
1.5 - A CDK5	17
1.6 - Roscovitina	18
1.7 - Objetivos:	19
2 - MÉTODOS.....	20
2.1 - Construção do modelo	25
2.2 - Inclusão da roscovitina	26
3 - RESULTADOS E DISCUSSÃO.....	29
3.1 - Qualidade dos Modelos	29
3.2 - Descrição do modelo	32
3.3 - O Garfo Molecular.....	37
3.4 - Posições das cadeias laterais	40
3.5 - Interações da roscovitina com a CDK5	41
4 - CONCLUSÕES.....	46
REFERÊNCIAS:	48
APÊNDICE A – MODELAGEM COMPARATIVA DE ESTRUTURAS.....	53
O método	55

Definição das restrições espaciais	58
Derivação de pdfs básicas	58
O banco de dados local.....	60
Satisfação das restrições espaciais.....	64
Derivação das pdfs de característica a partir de pdfs básicas	64
Derivação da pdf molecular a partir das pdfs de características.....	67
Otimização da pdf molecular	68
REFERÊNCIAS:	71
APÊNDICE B – CÁLCULO DA ACESSIBILIDADE ESTÁTICA.....	72
Identificação de cavidades	74
REFERÊNCIAS:	75
APÊNDICE C – ESTRUTURAS DE CDKS DEPOSITADAS NO PROTEIN DATA BANK	76
ANEXO A – ARTIGO	84

ÍNDICE DE FIGURAS

Figura 1: Diversas CDKs e suas ciclinas associadas.....	6
Figura 2: Alinhamento de sequência das CDKs	8
Figura 3: Fases do ciclo celular em células eucariontes	10
Figura 4: Atuação das várias CDKs no controle do ciclo celular em uma célula animal.....	10
Figura 5: Atividade da CDK1 durante o ciclo celular em uma célula animal.....	13
Figura 6: Estrutura plana da molécula de Roscovitina	18
Figura 7: Resultado do alinhamento das seqüências dos 8 modelos com a seqüência da CDK5	24
Figura 8: Protocolo para modelagem do complexo CDK5-roscovitina	28
Figura 9: Parte de uma cadeia polipeptídica indicando os graus de liberdade torcional de cada unidade peptídica. Os ângulos de rotação Phi e Psi são mostrados sendo estes, rotações das ligações C α -N e C α -C respectivamente.....	30
Figura 10: Gráfico de Ramachandran para o modelo da CDK5 complexada com roscovitina.....	31
Figura 11: Gráfico de Ramachandran para as 8 CDKs usadas como padrão (templates)	31
Figura 12: Diagrama de fitas do complexo CDK5-roscovitina	33
Figura 13: Análise da estrutura secundária e nível de acessibilidade ao solvente por resíduo (PROCHECK)	34
Figura 14: Potencial eletrostático da superfície do complexo CDK5-roscovitina	35
Figura 15: Potencial eletrostático da superfície do sítio de ligação de ATP do complexo CDK5-roscovitina.....	36
Figura 16: Potencial eletrostático da superfície do sítio de ligação de ATP do complexo CDK2-roscovitina.....	36
Figura 17: Diagrama esquemático do Garfo Molecular	39
Figura 18: Ligações de hidrogênio formadas entre a CDK5 e o inibidor	42

Figura 19: Ligações de hidrogênio formadas entre a CDK2 e o inibidor	43
Figura 20: Superposição dos sítios de ligação dos complexos CDK5-roscovitina (linha grossa) e CDK2-ATP (linha fina)	45
Figura 21: Principais passos do processo de modelagem.	57
Figura 22: Configuração típica das estruturas da classe α/β	62
Figura 23: Composição do banco de dados local. (a) Classes de conformações da cadeia lateral para todos os resíduos. (b) Escore fracional de identidade de sequência para todos os pares de proteínas relacionadas. (c) Resolução da análise de Raios X para todas as proteínas. (d) Frequência de ocorrência por tipo de resíduos.	63
Figura 24: Definição das classes de conformações da cadeia principal.	63
Figura 25: Derivação de uma pdf de características a partir de pdfs básicas. .	65
Figura 26: Mudança de conformação pela otimização da função objetivo	70
Figura 27: Contornos da superfície de van der Waals no plano.....	72
Figura 28: Contornos de superfícies gerados para o cálculo da superfície acessível	73

ÍNDICE DE TABELAS

Tabela 1: Valores de IC_{50} para Roscovitina e várias CDKs.....	18
Tabela 2: Melhores resultados da busca por modelos (templates).	21
Tabela 3: Lista de estruturas iniciais usadas na modelagem (templates).	23
Tabela 4: Resultado das análises de qualidade estereoquímica.....	30
Tabela 5: Ligações de hidrogênio entre o Garfo Molecular da CDK2 e diferentes inibidores.....	38
Tabela 6: Ligações de hidrogênio formadas nos complexos CDK-roscovitina.	43
Tabela 7: Composição do banco de dados local.....	61

ÍNDICE DE ABREVIações

3D	- Tridimensional
ATP	- Adenosina Tri-fosfato
CDK	- Cyclin-Dependent Kinase, Quinase dependente de Ciclina
DNA	- Deoxyribonucleic acid, ácido desoxirribonucléico
IC ₅₀	- Inhibitory Concentration 50%, concentração inibitória mínima 50%
NMR	- Nuclear Magnetic Resonance, Ressonância Magnética Nuclear
PDB	- Protein Data Bank, Banco de dados de proteínas
PDF	- Probability density function, função de densidade de probabilidade
Rb	- Retinoblastoma
RMSD	- Root Mean Square Deviation, desvio quadrático médio
RNA	- Ribonucleic acid, ácido ribonucléico

Resumo

A fosforilação de proteínas, iniciada em resposta a sinais extracelulares, representa um mecanismo universal para o controle celular de diferentes processos, entre eles as vias metabólicas, o crescimento e diferenciação celular, transporte através da membrana, e a apoptose.

Este processo de fosforilação é catalisado por uma grande família de proteínas conhecidas como Quinases, as quais catalisam a transferência do grupo fosfato da ATP para resíduos específicos do substrato. Dentre as Quinases existe a família das Quinases dependentes de Ciclina (CDK), que tem sido extensivamente estudada por sua função essencial para o controle do ciclo de divisão celular.

A busca por inibidores para CDKs foi iniciada devido a sua atividade anti-tumoral sendo que atualmente algumas novas aplicações têm sido investigadas. Muitas doenças resultam de uma proliferação descontrolada mas não tumoral de certas células, como a Arteriosclerose, a Restenose pós-angioplastia, a Angiogênese tumoral, a Glomerulonefrite e a Psoríase. Inibidores de CDKs encontram aplicação também no controle de alguns vírus, protozoários parasitas e doenças neuro-degenerativas.

Nesta dissertação de mestrado foi estudada a CDK5 que é a única integrante da família das CDKs envolvida no controle de funções neurais e que tem o seu aumento de atividade relacionado com o aparecimento do Mal de Alzheimer.

O procedimento apresentado neste trabalho pode ser utilizado como um protocolo básico para a modelagem de complexos CDK-inibidor e posterior avaliação dos resultados obtidos, que podem fornecer informações estruturais importantes sobre esses complexos binários, de forma a melhorar o conhecimento das bases estruturais da inibição química dessa família de proteínas.

O inibidor estudado foi a Roscovitina, um derivado de purina, e os resultados obtidos indicam que esse inibidor se liga fortemente ao sítio de ligação de ATP, apresentando um maior número de ligações de hidrogênio com a CDK5 do que as ligações presentes no complexo com a CDK2, o que acarreta a maior afinidade desse inibidor pela CDK5.

O modelo da CDK5 apresentou estrutura formada por dois lóbulos principais, semelhante à encontrada para a CDK2, com o lóbulo menor, N-terminal, formado predominantemente por folhas β , e o lóbulo maior, C-terminal, formado principalmente por hélices α .

Foi possível também identificar a presença do padrão de interação que caracteriza o Garfo Molecular, o qual foi anteriormente descrito nas interações de inibidores com a CDK2 e a CDK9.



Abstract

The phosphorylation of proteins, triggered in response to extracellular signals, represents a universal mechanism for the cellular control of many different processes, including metabolic pathways, cell growth and differentiation, membrane transport and apoptosis.

A large family of proteins known as kinases catalyzes this process of phosphorylation, acting on the transferring of the ATP phosphate groups to specific substrate residues. Among kinases there is the family of Cyclin-dependent kinases (CDK), which have been extensively studied because of their essential role in the control of the cell cycle.

The search for CDK inhibitors is due to their anti-tumor activity and at present some new applications are under investigation. The uncontrolled but non-tumoral proliferation of certain cells results in several diseases, such as atherosclerosis, post-angioplastic restenosis, tumor angiogenesis, glomerulonephritis and psoriasis. CDK inhibitors may also have applications in the control of some viruses, parasitic protozoa and neurodegeneration.

In this project we studied CDK5, which is the only member of CDKs family involved with neural functions control and which has its activity increase related to the development of Alzheimer's disease.

The procedure presented in this work can be used as a basic protocol for CDK-inhibitor modeling whose properly evaluated results may lead to important structural information about these binary complexes, enhancing our knowledge on the structural basis for chemical inhibition of this family of proteins.



Roscovitine, a purine derivative, is the inhibitor studied, and our results point out that this inhibitor strongly binds to ATP binding site with more hydrogen bonds with CDK5 than with the CDK2 complex. Thus Roscovitine presents higher affinity with CDK5.

The model of the CDK5 is folded into the typical bilobal structure as was observed for CDK2 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 presence of the interaction motif characteristic of the Molecular Fork, formerly described in the interactions of CDK2 and CDK9, has been also identified.



1 - Inibição das CDKs

1.1 - O que são as CDKs

A fosforilação de proteínas, iniciada em resposta a sinais extracelulares, representa um mecanismo universal para o controle celular de diferentes processos, entre eles as vias metabólicas, o crescimento e diferenciação celular, transporte através da membrana, e a apoptose. (Hunter, 1995; Johnson & Lewis, 2001)

Este processo de fosforilação de proteínas é catalisado por uma grande família de proteínas homólogas conhecidas como Quinases, as quais catalisam a transferência do grupo fosfato da ATP para resíduos específicos do substrato. (Scapin, 2002)

Estima-se que existam, nos seres humanos, entre 1000 e 2000 diferentes Quinases, dentre as quais, a família das Quinases dependentes de Ciclina (CDK), que tem sido extensivamente estudada por sua função essencial para as células. (Meijer, 2000)

Cada CDK, durante sua breve existência, fosforila suas proteínas alvo que então, direta ou indiretamente, ativam o próximo conjunto de eventos do ciclo celular. Uma CDK, quando ativa, apresenta-se como um heterodímero composto de uma subunidade catalítica, a Quinase, e uma subunidade regulatória, a Ciclina, que ativa a Quinase (Figura 1). Em alguns casos, soma-se a esse heterodímero proteínas da família das CKs, como por exemplo a proteína p9, as quais se ligam às CDKs e tem como função modular a habilidade da CDK de reconhecer seus substratos, sendo que os mecanismos

de tal regulação não são ainda claramente conhecidos. Em células de vertebrados existem pelo menos quatro diferentes CDKs envolvidas no controle do ciclo celular, assim como outras CDKs que desempenham outras funções (Branden & Tooze, 1998).

Quinases Dependentes de Ciclinas (CDKs)

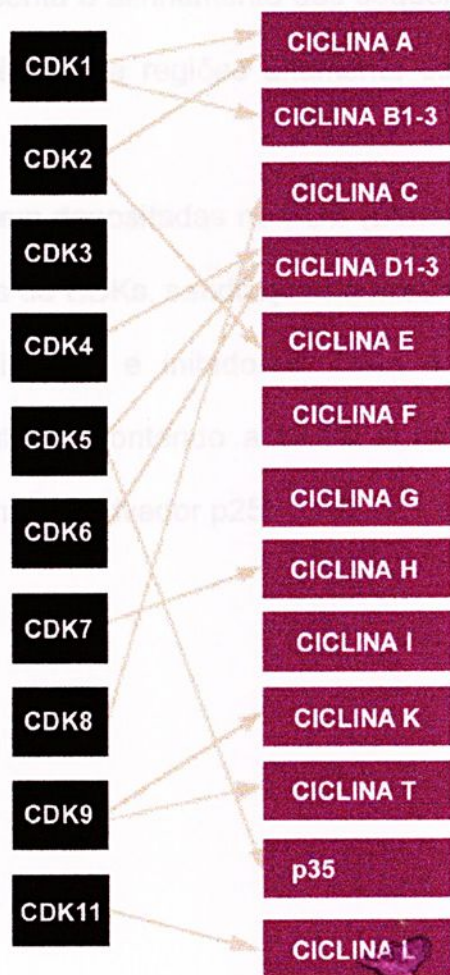


Figura 1: Diversas CDKs e suas ciclinas associadas

As diferentes subunidades catalíticas são produtos de uma família de genes fortemente relacionados. As diferentes Ciclinas, uma ou mais para cada

subunidade catalítica, são também produtos de uma família de genes. As Ciclinas que compõem as CDKs passam por um processo programado de síntese, acumulação e degradação; seus períodos de vida, que são pequenos quando comparados com a duração total do ciclo celular, garantem que as CDKs, das quais fazem parte, permaneçam ativas somente por curtos períodos de tempo e no momento correto do ciclo celular (Branden & Tooze, 1998).

A Figura 2 apresenta o alinhamento das sequências de 10 CDKs, onde pode-se notar a ocorrência de regiões altamente conservadas da estrutura primária.

Atualmente existem depositadas no PDB (Berman et al., 2000), mais de 50 estruturas resolvidas de CDKs, sendo que na maior parte são complexos da CDK2 com diversos ligantes e inibidores. Além das estruturas da CDK2 existem algumas estruturas contendo a CDK6 e uma estrutura recente da CDK5 em complexo com seu ativador p25.



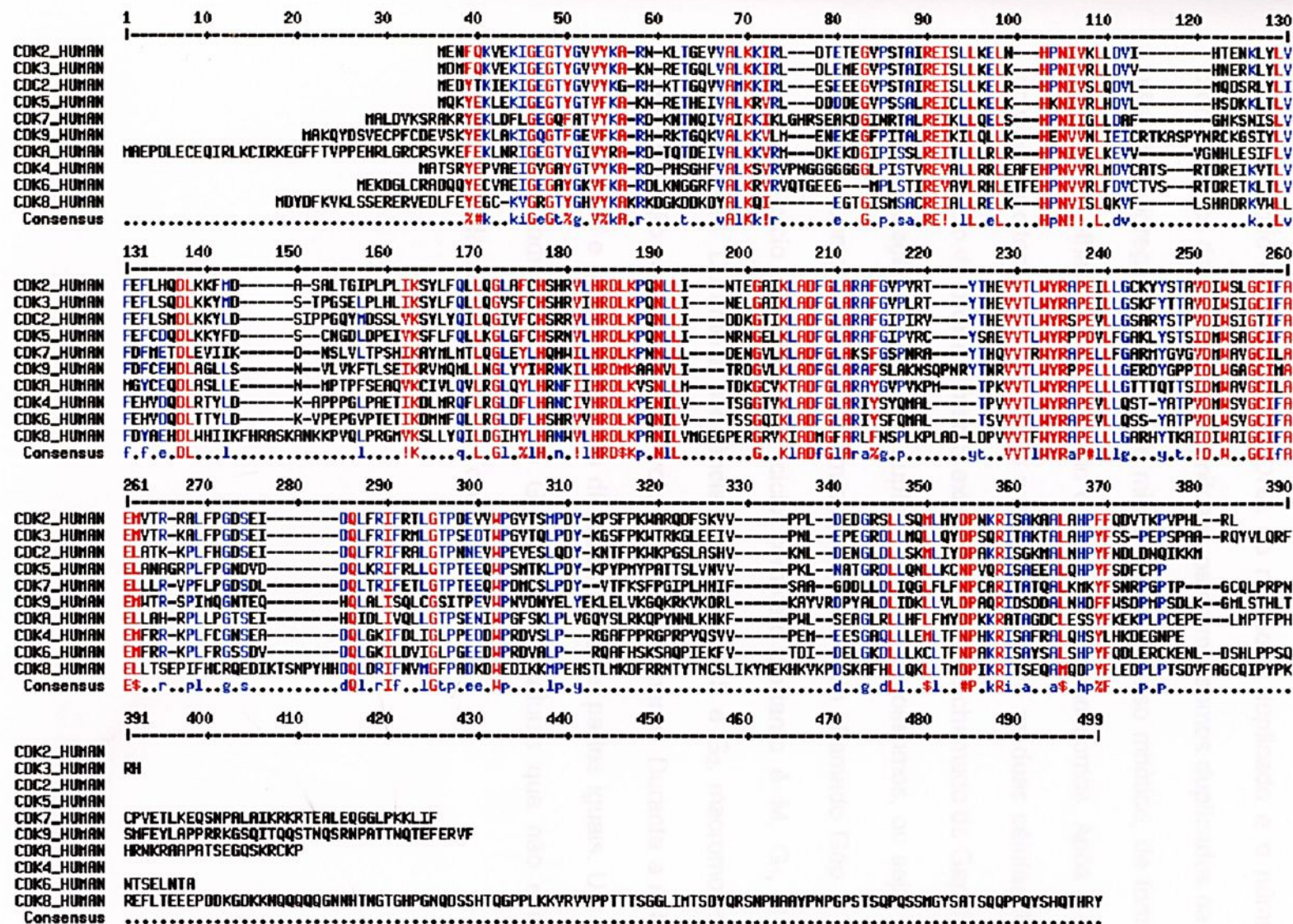


Figura 2: Alinhamento de sequência das CDKs

1.2 - Função das CDKs

O ciclo de divisão celular de todas as células eucariontes pode ser dividido em quatro principais fases: G_1 , S, G_2 , e M (figura 3). Na fase S (síntese de DNA), todo o conteúdo de DNA do núcleo é duplicado e o número de cromossomos dobra; na fase M (mitose), os cromossomos duplicados da célula mãe são segregados usando os microtúbulos do fuso mitótico, de forma que cada célula filha receba o mesmo conjunto de cromossomos. Após a mitose, incluindo a citocinese que divide o citoplasma entre as duas células filhas, e antes do início da próxima fase S existe um intervalo chamado de Gap 1 ou G_1 . Após a replicação do DNA e a duplicação dos cromossomos, ou seja, após a fase S ser completada, existe um segundo intervalo chamado Gap 2, ou G_2 , antes do início da mitose. O ciclo completo portanto é M, G_1 , S, G_2 e novamente M. Durante a sequência de fases G_1 , S e G_2 , macromoléculas e organelas são construídas, e o volume celular aumenta. Durante a mitose os cromossomos e o citoplasma são divididos em duas partes iguais. Um estado estacionário adicional, chamado G_0 ocorre em células que não estão em processo de divisão (Branden & Tooze, 1998).

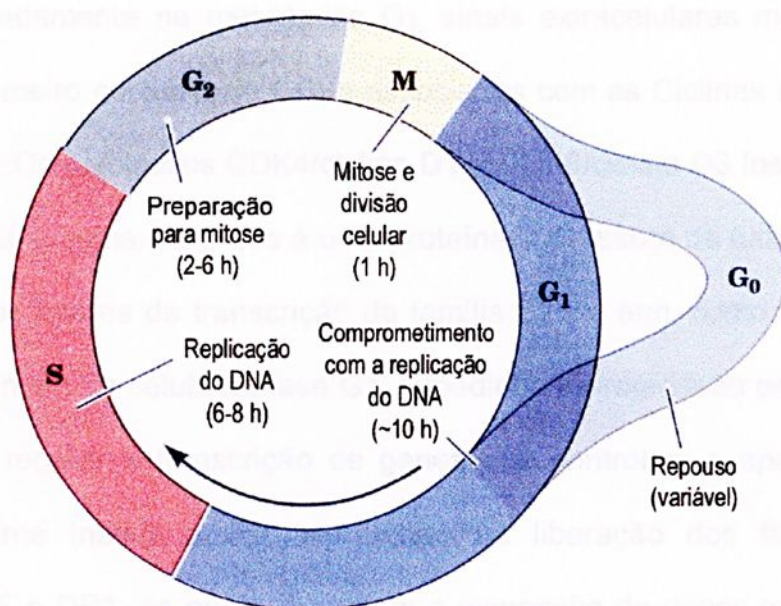


Figura 3: Fases do ciclo celular em células eucariontes
(Voet & Voet, 1995)

A progressão através das fases G₁, S, G₂, e M do ciclo de divisão celular é diretamente controlada pela ativação transitória de várias CDKs. (figura 4)

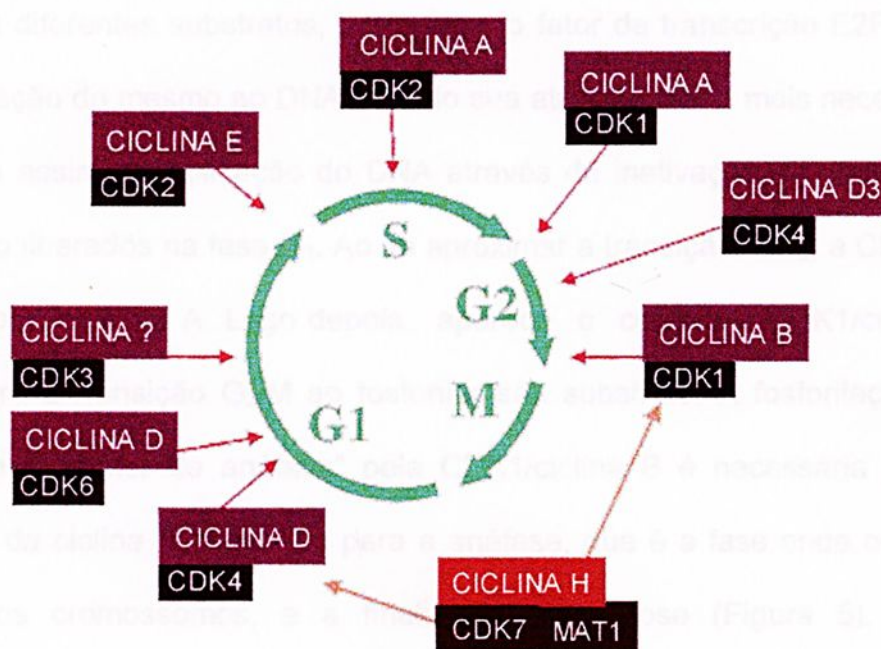


Figura 4: Atuação das várias CDKs no controle do ciclo celular em uma célula animal

Aproximadamente na metade de G_1 , sinais extracelulares modulam a ativação do primeiro conjunto de CDKs associadas com as Ciclinas do tipo D, CDK4 e CDK6. Os complexos CDK4/ciclina D1 e CDK6/ciclina D3 fosforilam a proteína Retinoblastoma (Rb), que é uma proteína supressora de tumores, que interage com os fatores de transcrição da família E2F e tem, como principais funções, a retenção da célula na fase G_1 , impedindo a progressão para a fase S, e também, regular a transcrição de genes que controlam a apoptose. A fosforilação torna inativa a Rb, resultando na liberação dos fatores de transcrição E2F e DP1, os quais controlam a expressão de genes que geram produtos necessários para a transição G_1/S e a progressão da fase S. O complexo CDK2/ciclina E, que é responsável pela transição G_1/S também causa, posteriormente, a fosforilação da Rb permitindo a liberação de uma maior quantidade de fatores de transcrição. Durante a fase S, a CDK2/ciclina A fosforila diferentes substratos, entre eles, o fator de transcrição E2F o que inibe a ligação do mesmo ao DNA quando sua atuação não é mais necessária, permitindo assim, a replicação do DNA através da inativação dos fatores de transcrição liberados na fase G_1 . Ao se aproximar a transição S/G_2 , a CDK1 se associa com a ciclina A. Logo depois, aparece o complexo CDK1/ciclina B que dispara a transição G_2/M ao fosforilar seu substrato. A fosforilação do "Complexo promotor de anáfase" pela CDK1/ciclina B é necessária para a proteólise da ciclina B, transição para a anáfase, que é a fase onde ocorre a divisão dos cromossomos, e a finalização da mitose (Figura 5). Essas sucessivas ondas de montagem, ativação e inativação dos complexos CDK/ciclina são fortemente reguladas por modificações pós-traducionais e translocações intracelulares. Elas são coordenadas e dependem da finalização



dos passos anteriores, através dos pontos de controle chamados pontos de checagem ("checkpoints"). (Meijer, 2000)

Além do seu papel no controle do ciclo celular algumas CDKs influenciam diretamente o processo de transcrição. O complexo CDK7/ciclina H/MAT1, que é um dos componentes do fator de transcrição TFIIH, é responsável pela fosforilação do domínio C-terminal de uma grande subunidade da RNA polimerase II, necessária para o processo de alongamento. (Meijer, 2000)

Para 5. Ajuste de CDK durante o ciclo celular em uma célula animal
(Meijer, 2000)

1.3 - Porquê falar as CDKs

A busca por moléculas para CDKs foi iniciada devido a sua atividade enzimática. Os primeiros resultados foram surpreendentes e revelaram que as principais moléculas para CDKs são as proteínas do ciclo de alongamento. A atividade de CDKs é regulada por uma variedade de fatores, incluindo a presença de ciclinas, a fosforilação e a interação com outras proteínas. A atividade de CDKs é essencial para o controle do ciclo celular e a progressão do ciclo de alongamento. A atividade de CDKs é regulada por uma variedade de fatores, incluindo a presença de ciclinas, a fosforilação e a interação com outras proteínas. A atividade de CDKs é essencial para o controle do ciclo celular e a progressão do ciclo de alongamento. (Meijer, 2000)

Além disso, algumas CDKs também atuam como reguladoras de outros processos celulares, como a transcrição e a tradução. A atividade de CDKs é regulada por uma variedade de fatores, incluindo a presença de ciclinas, a fosforilação e a interação com outras proteínas. A atividade de CDKs é essencial para o controle do ciclo celular e a progressão do ciclo de alongamento. (Meijer, 2000)

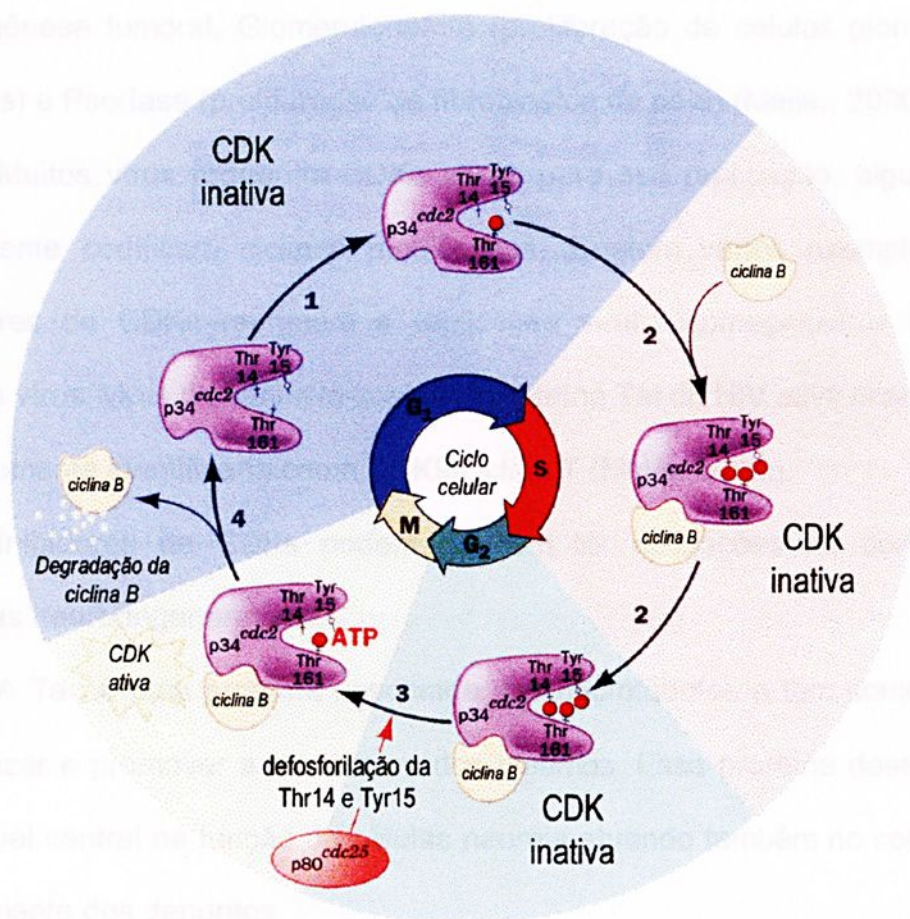


Figura 5: Atividade da CDK1 durante o ciclo celular em uma célula animal (Voet & Voet, 1995)

1.3 - Porquê inibir as CDKs

A busca por inibidores para CDKs foi iniciada devido a sua atividade anti-tumoral. Os primeiros resultados foram encorajadores e revelaram várias propriedades dos inibidores de CDKs: bloqueio da progressão do ciclo celular seguido pela indução de apoptose, reversibilidade dos efeitos em células saudáveis, efeitos independentes da Rb e interação com proteínas resistentes a drogas (Meijer, 2000).

Além disso, algumas novas aplicações têm sido investigadas. Muitas doenças resultam de uma proliferação descontrolada mas não tumoral de certas células, como a Arteriosclerose, Restenose pós-angioplastia,

Angiogênese tumoral, Glomerulonefrite (proliferação de células glomerulares dos rins) e Psoríase (proliferação de fibroblastos da pele) (Meijer, 2000).

Muitos vírus requerem CDKs ativas para sua replicação, alguns vírus atualmente codificam ciclinas modificadas. Existem vários exemplos onde inibidores de CDKs reprimem a replicação viral: citomegalovirus humano, Herpes vírus, vírus da varicella-zoster. A proteína Tat do HIV ativa uma quinase recentemente identificada como CDK9/ciclina T (Meijer, 2000).

Inibidores de CDKs podem também ter aplicações no controle de doenças neurodegenerativas.

A Tau é uma proteína associada aos microtúbulos e tem como função estabilizar e promover a montagem dos mesmos. Essa proteína desempenha um papel central na função de células neurais atuando também no controle do crescimento dos dendritos.

A hiperfosforilação da Tau é responsável pela perda da sua capacidade de se ligar aos microtúbulos e promover sua montagem. Em tal situação essa proteína forma agregados, chamados emaranhados neurofibrilares que contém os Filamentos Helicoidais Pareados (PHF) tipicamente associados com o Mal de Alzheimer. Inibidores de CDKs impedem a hiperfosforilação da Tau *in-vivo*. Eles também parecem proteger neurônios da apoptose, ainda que os mecanismos envolvidos não sejam claros. Esses dados apóiam fortemente o estudo dos inibidores de CDKs como agentes farmacológicos em potencial para o controle de doenças neurodegenerativas (Meijer, 2000).

Outra potencial aplicação atualmente investigada é o uso de inibidores de CDKs e seus derivados contra protozoários parasitas como os *Plasmodium*, *Trypanosoma cruzi*, *Leishmania*, *Toxoplasma* e Microsporídia. A multiplicação e

desenvolvimento desses parasitas provavelmente dependem da atividade de quinases relacionadas com as CDKs. A homologia dessas proteínas com as CDKs encontradas em vertebrados (usualmente entre 40 e 60% de identidade) é suficiente para permitir a sua classificação como quinases relacionadas com as CDKs, mas deixa espaço para uma divergência significativa. Isso abre uma janela terapêutica potencialmente promissora: sem dúvida, diferenças estruturais entre CDKs do parasita e do hospedeiro podem se traduzir em diferentes afinidades para moléculas inibidoras. Utilizar como alvo o sítio de ligação de ATP das quinases apresenta a vantagem de, esta parte da proteína não permitir muitas variações, fazendo com que o desenvolvimento de resistência tenha menor probabilidade de ocorrer que no caso da utilização de alvos mais variáveis. (Meijer, 2000)

Atualmente existem inibidores de CDKs, como o Flavopiridol, que se encontram já na fase de testes clínicos para uso farmacológico.

1.4 - Mecanismos de inibição

Embora o núcleo catalítico das quinases tenha sido conservado durante o processo evolutivo (Hanks & Hunter, 1995), os mecanismos para inibição de cada quinase podem variar consideravelmente. Fisiologicamente, pelo menos três mecanismos de inibição foram identificados: um mecanismo de pseudo-substrato, um mecanismo de mimetização de adenina e um mecanismo que envolve a fixação da enzima em uma conformação inativa utilizando outras superfícies fora do sítio ativo (Taylor & Radzio-Andzelm, 1997). Quimicamente, o campo da inibição das quinases se mostrou muito ativo nos últimos 10 anos

e, até hoje, vários fármacos sintéticos foram identificados como potentes inibidores de quinases. Devido à diferente natureza dos sítios de ligação de ATP e do substrato, grande parte dos esforços de síntese tem sido dedicada ao sítio de ligação de ATP e a grande maioria dos inibidores obtidos é composta por derivados de adenina, ainda que existam vários exemplos de inibidores dirigidos aos domínios não catalíticos das quinases (Broadbridge & Sharma, 2000; Vu, 2000; Cody et al., 2000; Yeh et al., 2001). O sítio onde se liga o substrato não tem sido extensivamente explorado para o desenvolvimento de inibidores (Lawrence & Niu, 1998), mesmo sabendo-se que compostos que bloqueiam a função da enzima por meio da competição com o substrato apresentam grandes vantagens sobre os compostos que imitam a ATP, isso por várias razões, incluindo o fato de que a alta concentração de ATP no interior da célula pode reduzir a eficiência de compostos que o imitam, e que inibidores que competem com o substrato tendem a ser mais específicos e mais seletivos. Recentemente, pequenos inibidores que competem com o substrato do fator de crescimento semelhante à insulina (IGF-1R) e quinase receptora de insulina (IR) têm sido descritos (Blum et al., 2000; Parang et al., 2001).



1.5 - A CDK5

A CDK5 é a única proteína integrante da família das CDKs envolvida no controle de funções neurais. Mesmo sendo largamente expressa em muitos tecidos a CDK5 tem sua atividade restrita aos neurônios. Tal especificidade se deve ao fato de seus ativadores (as proteínas p25, p35 e p39) estarem presentes apenas em neurônios. Estudos demonstraram que o complexo CDK5-p35 desempenha um papel essencial no crescimento dos axônios, na diferenciação (Nikolic et al., 1996; Ohshima et al., 1996) e migração dos neurônios, e também para a configuração laminar do córtex cerebral. Evidências recentes de que este complexo se concentra na extremidade dos axônios em crescimento indicam que o mesmo participa do controle do processo de extensão dos axônios dos neurônios em desenvolvimento (Maccioni et al., 2001).^{*} O aumento na atividade de tal complexo tem sido relacionado com o aparecimento do Mal de Alzheimer onde se verifica um nível de atividade da CDK5 duas vezes maior quando comparados os tecidos cerebrais de doentes com os de indivíduos normais de mesma idade (Lee et al., 2000).

1.6 - Roscovitina

Roscovitina [2-(R)-(1-ethyl-2-hydroxy-ethylamino)-6-benzylamino-9-isopropylpurine] é um derivado de purina (Figura 6) com um grande poder de inibição para as CDKs tendo valores de IC₅₀ que variam desde 0,16 μ M para a CDK5 até 100 μ M para CDK4 e CDK6 (Tabela 1). A estrutura cristalográfica do complexo CDK2-roscovitina indica que esse inibidor se liga muito fortemente ao sítio de ligação de ATP da CDK2. (AZEVEDO JUNIOR et al., 1997)

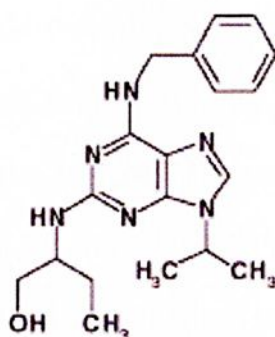


Figura 6: Estrutura plana da molécula de Roscovitina

Tabela 1: Valores de IC₅₀ para Roscovitina e várias CDKs

Enzima	IC ₅₀ para Roscovitina (μ M)
Cdc2/ciclina B	0,65
CDK2/ciclina A	0,70
CDK2/ciclina E	0,70
CDK4/ciclina D1	> 100
CDK5/p35	0,16
CDK6/ciclina D3	> 100

1.7 - Objetivos:

O presente trabalho teve como objetivo desenvolver um protocolo para a modelagem do complexo CDK5-roscovitina que possa ser posteriormente utilizado na modelagem de novos complexos.

As informações estruturais obtidas dos complexos binários são utilizadas para se conhecer as bases estruturais da inibição química dessa família de proteínas.



2 - Métodos

O programa escolhido para executar a modelagem foi o MODELLER (Sali & Blundell, 1993) o qual implementa uma abordagem automatizada para a modelagem comparativa de estruturas de proteínas baseada na satisfação de restrições espaciais.

O ponto inicial para a modelagem comparativa é a identificação de todas as estruturas de proteínas relacionadas com a sequência alvo, e então, a seleção daquelas estruturas que serão utilizadas como modelo (templates) (Marti-Renom et al., 2003).

Este processo de busca dos modelos (templates) foi executado com o próprio programa de modelagem. Para isso foi gerado um arquivo no formato FASTA contendo a sequência primária da proteína a ser modelada e um arquivo de controle contendo os comandos necessários para que o MODELLER executasse o processo de busca das melhores estruturas modelo. Os conteúdos dos dois arquivos são mostrados a seguir:

Arquivo contendo a sequência: cdk5.seq

>P1;cdk5

sequence:cdk5:::.....

MQKYEKLEKIGEGTYGTVFKAKNRETHEIVALKRVRLDDDDDEGVPSSALREICL
LKELKHKNIVRLHDLHSDKKLTLVFEFCDQDLKKYFDSCNGDLDPEIVKSFLF
QLLKGLGFCHSRNVLHRDLKPQNLLINRNGELKLADFGGLARAFGIPVRCYSAE
VVTWYRPPDVLFgaklystsIDMWSAGCIFAELANAGRPLFPGNDVDDQLK
RIFRLLGTPTEEQWPSMTKLDPYKPYMPATTSLVNVVPKLNATGRDLLQNL
LKCNPVQRISAEALQHPYFSDFCPP*

Arquivo de controle: search.top

SET SEARCH_RANDOMIZATIONS = 100

SEQUENCE_SEARCH FILE = 'cdk5.seq', ALIGN_CODES = 'cdk5'

Após o processo de busca o programa apresenta uma tabela com códigos de estruturas das 20 melhores famílias de estruturas a serem usadas como modelos (templates). Na tabela 2 vemos os 5 melhores resultados.

Tabela 2: Melhores resultados da busca por modelos (templates).

#	CODE	LEN	NID	%ID	SCORE	SIGNI
1	1hcl	294	174	59,6	206768	38,8
2	1phk	277	77	27,8	143493	8,4
3	1csn	293	65	22,3	145983	7,7
4	1kobA	353	66	22,6	148426	7,3
5	1apmE	341	62	21,2	148198	6,4

Nesta tabela temos o código (CODE) sob o qual estão depositadas as estruturas no PDB (Berman et al., 2000), o número de resíduos de cada estrutura (LEN), o número de resíduos idênticos (NID), a porcentagem de identidade com relação à sequência da proteína a ser modelada (%ID), o escore atingido na correlação entre as seqüências (SCORE) e a coluna SIGNI que indica quão significativa é a estrutura para o processo de modelagem. Valores de SIGNI maiores que 6 indicam estruturas boas para serem utilizadas no processo de modelagem (Sánchez & Sali, 2000).

Com base nos resultados obtidos na busca por modelos, optamos por utilizar estruturas da família da 1hcl (Schulze-Gahmen et al., 1995) pois a mesma apresentou alta porcentagem de identidade (59,6%), o melhor escore (206.768) e o mais alto valor de significância (38,8).

Consultando no PDB o código 1hcl vimos que se trata da estrutura da proteína CDK2 humana resolvida a uma resolução de 1,8 Å (Schulze-Gahmen et al., 1995). Como existem outras estruturas de CDK2 humana depositadas no PDB optou-se por utilizar como padrão um conjunto contendo oito dessas estruturas, todas com resolução superior a 2,2Å as quais se encontram depositadas no PDB com os seguintes códigos: 1hcl, 1aq1, 1ckp, 1dm2, 1e1v, 1e1x, 1hck e 1jsv. As estruturas de códigos 1aq1, 1ckp, 1dm2, 1e1v, 1e1x, 1hck e 1jsv contêm, além da CDK2, as coordenadas atômicas das moléculas de Staurosporina, Purvalanol B, Himenialdisina, NU20, NU60, ATP e U55 respectivamente (Tabela 3).

O último passo antes de se iniciar a modelagem é o alinhamento das sequências. Este alinhamento pode ser executado pelo próprio MODELLER mas no nosso caso utilizamos os programas CLUSTAL V (Higgins et al., 1992) e Multi-Alin (Corpet, 1988) obtendo os mesmos resultados. (Figura 7).



Tabela 3: Lista de estruturas iniciais usadas na modelagem (templates).

Cód. PDB	Resolução [Å]	R-fator	Referência	Res. ausentes
1hcl	1,80	0,181	Schulze-Gahmen et al., 1995	37 - 40
1aq1	2,00	0,220	Lawrie et al., 1997	36 - 43 e 149 - 161
1ckp	2,05	0,192	Gray et al., 1998	36 - 43 e 153 - 163
1dm2	2,10	0,192	Meijer et al., 2000	36 - 44 e 149 - 163
1e1v	1,95	0,195	Arris et al., 2000	36 - 43
1e1x	1,85	0,213	Arris et al., 2000	36 - 43
1hck	1,90	0,185	Schulze-Gahmen et al., 1996	72 - 75 e 153 - 162
1jsv	1,96	0,185	Clare et al., 2001	37 - 46

$$R - fator : \frac{\sum |F_0 - F_c|}{\sum F_0}$$

F_0 = Fator de Estrutura observado

F_c = Fator de Estrutura calculado



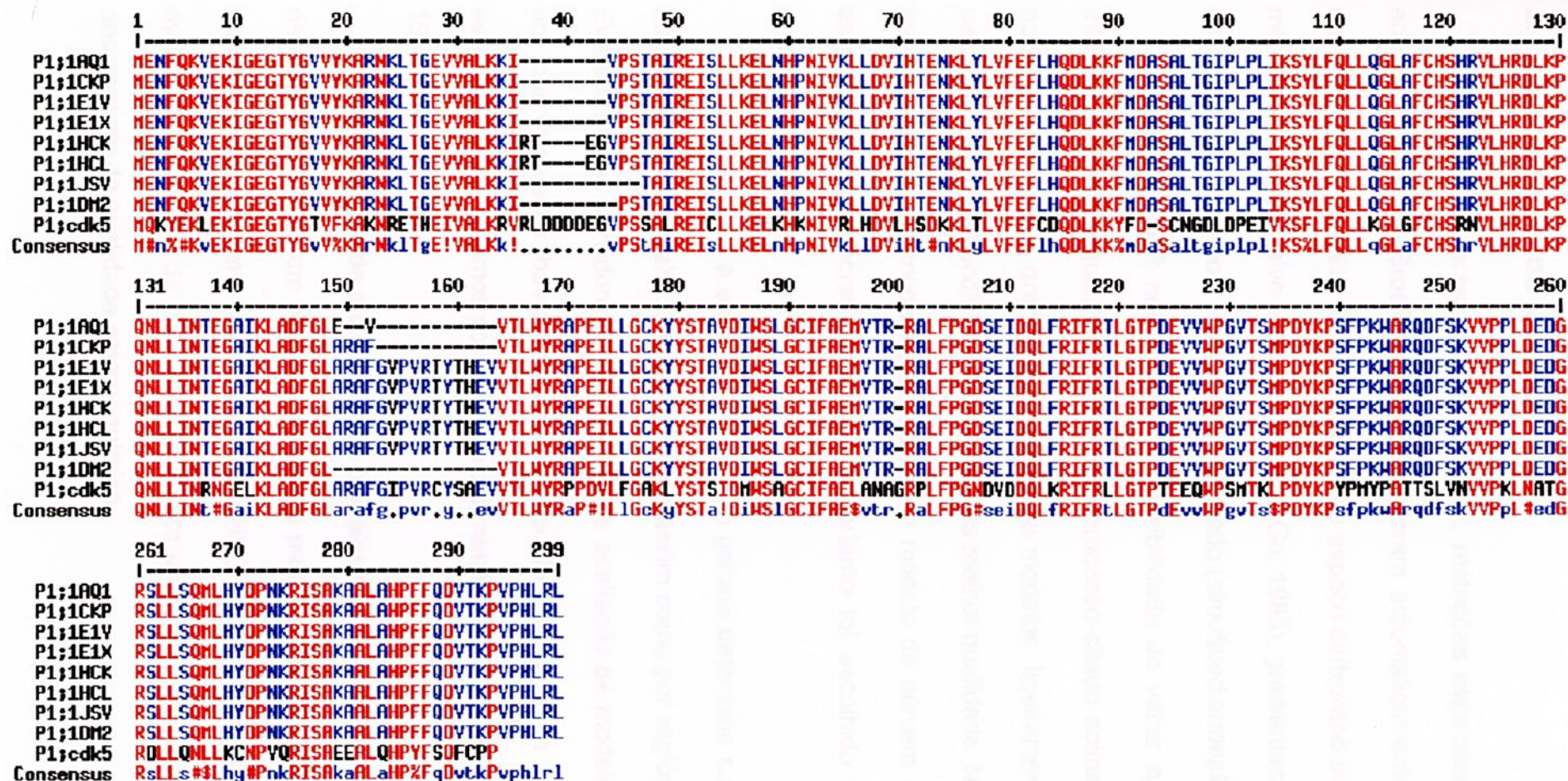


Figura 7: Resultado do alinhamento das sequências dos 8 modelos com a sequência da CDK5

2.1 - Construção do modelo

Durante o processo de modelagem, restrições espaciais e termos de energia CHARMM (Brooks et al., 1983) foram automaticamente combinados em uma função objetivo a ser otimizada no espaço cartesiano utilizando-se os métodos de função alvo variável (Braun & Go, 1985), gradientes conjugados e dinâmica molecular com recozimento simulado (*simulated annealing*).

O MODELLER nos fornece a possibilidade de variar a conformação inicial que é em seguida otimizada pelo processo citado acima e, com isso, podemos obter uma grande quantidade de modelos, ligeiramente diferentes, para que se possa escolher o que apresenta melhor qualidade. No nosso caso, foram gerados 500 modelos, sendo que o modelo de número 423 foi o que apresentou os melhores resultados e portanto foi escolhido como modelo definitivo.

Sabemos que a exatidão do modelo gerado determina sua utilidade. O modelo pode ser avaliado como um todo assim como por regiões individuais. Existem muitos servidores e programas de avaliação de modelos mas, como um dos requisitos básicos é que o modelo possua uma boa qualidade estereoquímica, optamos por utilizar o programa PROCHECK (Laskowski et al., 1993).

Devido à grande quantidade de modelos gerados tornou-se necessário o desenvolvimento de um pequeno programa para automatizar este processo de análise. Este programa executa o PROCHECK para cada um dos modelos, avalia os resultados de cada análise e gera uma lista, classificada em ordem decrescente de qualidade estereoquímica.



2.2 - Inclusão da roscovitina

O complexo CDK5-roscovitina foi construído a partir da superposição da estrutura do complexo CDK2-roscovitina com o modelo gerado para a CDK5. Esse passo foi executado com o auxílio do programa LSQKAB que consta do pacote CCP4 (Collaborative Computational Project 4, 1994) o qual executa o ajuste de conjuntos de coordenadas atômicas em arquivos no formato PDB.

Os átomos a serem ajustados podem ser selecionados por grupos pré-definidos no programa, que são, átomos da cadeia principal, átomos das cadeias laterais, carbonos alfa, ou selecionados por sua posição nos arquivos de entrada. Uma das estruturas, a ser usada como referência, permanecerá com coordenadas fixas, enquanto a outra estrutura terá suas coordenadas alteradas.

Para executar o ajuste, o programa define o centróide do conjunto de átomos a ser movido e ajusta ao centróide do conjunto de átomos de referência, e então, uma matriz de rotação é derivada, utilizando o método sugerido por Kabsch em 1976. Dessa forma, todo o conjunto de coordenadas pode ser transformado, e como resultado teremos as duas estruturas, superpostas e corretamente orientadas de acordo com o critério selecionado.

Outras duas funções adicionais que este programa executa são a alteração das coordenadas de arquivos de estrutura, de acordo com matrizes de rotação e translação fornecida pelo usuário, e também, o cálculo de desvio médio quadrático para conjuntos de átomos. Nessa modalidade, o programa retorna uma tabela contendo as distâncias RMS entre pares de átomos selecionados.



No presente caso, as estruturas da CDK2-roscovitina e o modelo de estrutura da CDK5 foram superpostos com base nos carbonos α , com isso a molécula de roscovitina ficou corretamente posicionada com relação ao modelo da CDK5, bastando então retirar-se os átomos da CDK2 para se obter as coordenadas atômicas da molécula de roscovitina em complexo com a CDK5.

Devido ao fato de o modelo da CDK5 ter sido gerado sem a presença da molécula de roscovitina foi preciso executar-se novamente o processo de otimização, desta vez otimizando-se todo o complexo, para que as interações entre as duas moléculas pudessem ser corretamente avaliadas no modelo final.

O próprio MODELLER foi utilizado para executar esta otimização, procedimento este que é descrito nos tutoriais do programa e utiliza as mesmas técnicas que são implementadas no processo de modelagem, só que neste caso, trabalhando sobre o modelo fornecido pelo usuário. Com isso, obtemos como resultado um novo modelo para o complexo, mas este leva em conta as interações entre a molécula do inibidor e as cadeias laterais dos resíduos presentes no bolsão de ligação, resíduos esses que podem eventualmente ser deslocados de sua posição inicial pela presença dos átomos do inibidor.

O protocolo utilizado para a modelagem do complexo CDK5-roscovitina pode ser visto na figura 8.

Por último foi feito o cálculo das superfícies de contato dos complexos CDK2-roscovitina e CDK5-roscovitina, procedimento esse executado com os programas AREAIMOL e RESAREA também integrantes do pacote CCP4 (Collaborative Computational Project 4, 1994).



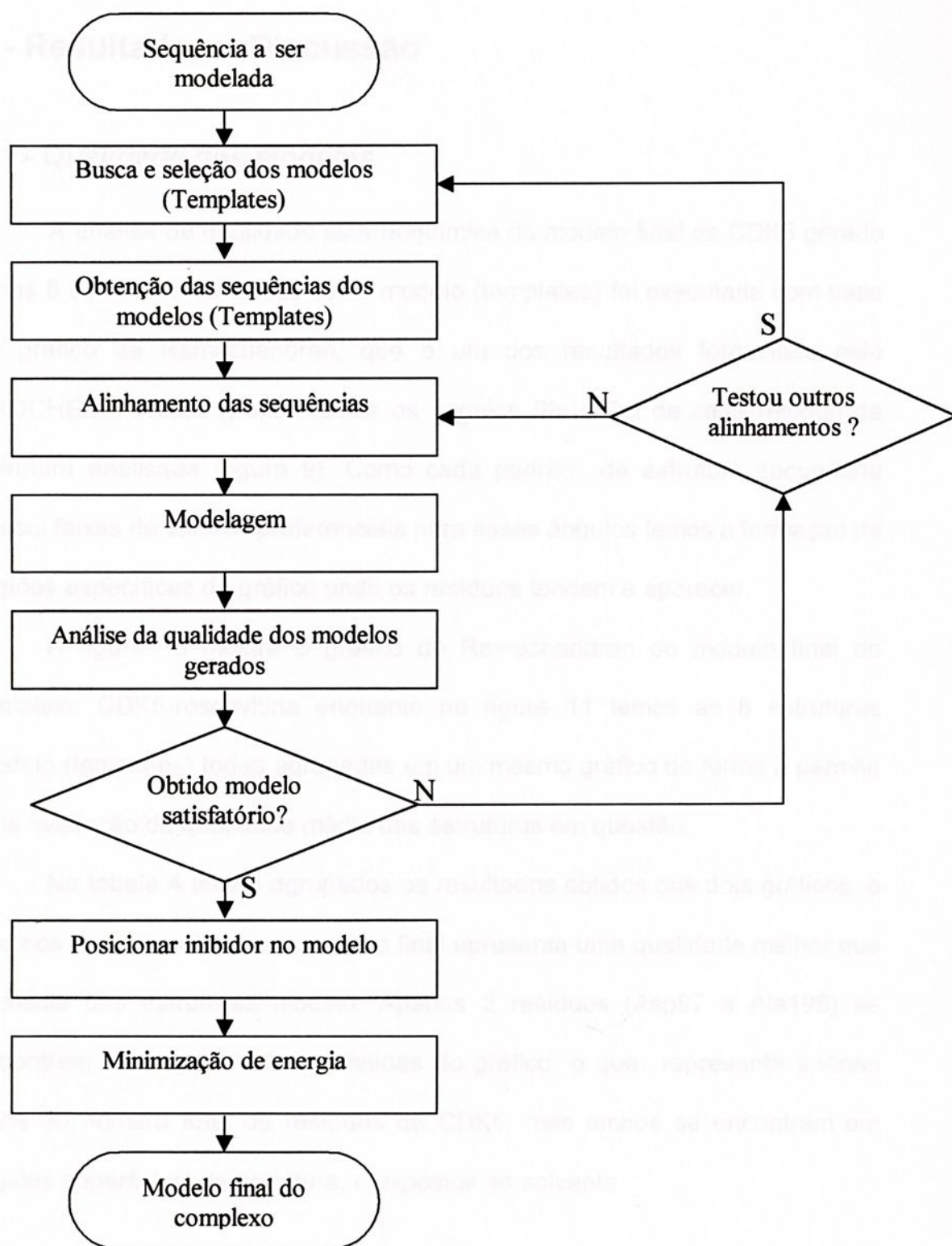


Figura 8: Protocolo para modelagem do complexo CDK5-roscovitina

3 - Resultados e Discussão

3.1 - Qualidade dos Modelos

A análise da qualidade estereoquímica do modelo final da CDK5 gerado e das 8 estruturas utilizadas como modelo (templates) foi executada com base no gráfico de Ramachandran, que é um dos resultados fornecidos pelo PROCHECK. Nesse gráfico temos os ângulos Phi e Psi de cada resíduo da estrutura analisada (figura 9). Como cada padrão de estrutura secundária possui faixas de valores preferenciais para esses ângulos temos a formação de regiões específicas do gráfico onde os resíduos tendem a aparecer.

A figura 10 mostra o gráfico de Ramachandran do modelo final do complexo CDK5-roscovitina enquanto na figura 11 temos as 8 estruturas modelo (templates) todas agrupadas em um mesmo gráfico de forma a permitir uma avaliação da qualidade média das estruturas em questão.

Na tabela 4 temos agrupados os resultados obtidos dos dois gráficos, o que nos permite notar que o modelo final apresenta uma qualidade melhor que a média das estruturas modelo. Apenas 2 resíduos (Asp97 e Ala198) se encontram em regiões não permitidas do gráfico, o que representa apenas 0,8% do número total de resíduos da CDK5, mas ambos se encontram em regiões superficiais da estrutura, e expostos ao solvente.

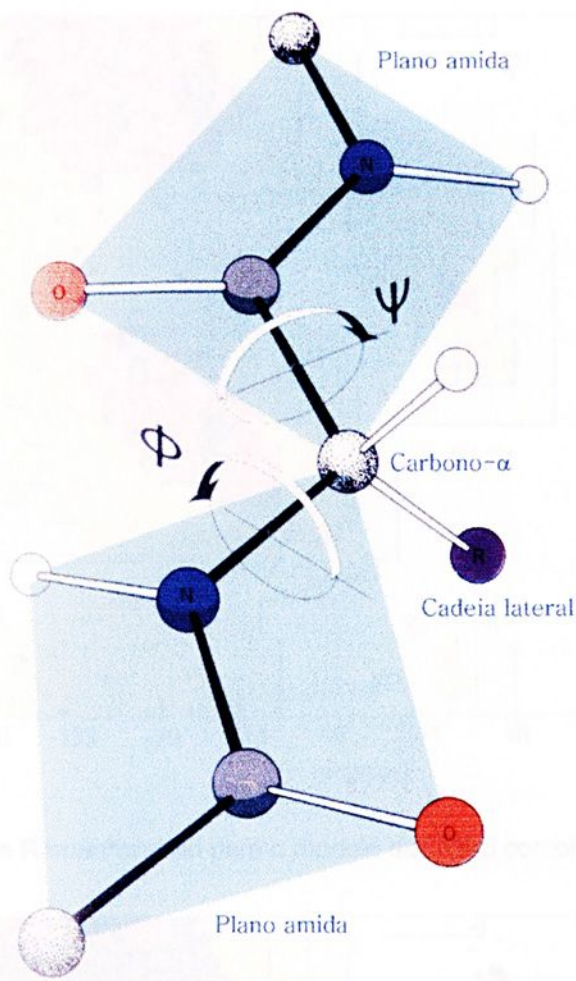


Figura 9: Parte de uma cadeia polipeptídica indicando os graus de liberdade torcional de cada unidade peptídica. Os ângulos de rotação Phi e Psi são mostrados sendo estes, rotações das ligações Cα-N e Cα-C respectivamente.

Figura 9: Parte de uma cadeia polipeptídica indicando os graus de liberdade torcional de cada unidade peptídica. Os ângulos de rotação Phi e Psi são mostrados sendo estes, rotações das ligações Cα-N e Cα-C respectivamente.

Tabela 4: Resultado das análises de qualidade estereoquímica

Região	8 CDKs	CDK5-roscovitina
Mais Favorável	89,8 %	91,0 %
Adicionalmente permitida	9,7 %	7,4 %
Generosamente permitida	0,5 %	0,8 %
Não permitida	0 %	0,8 %

Figura 11: Gráfico de Ramachandran para os 8 CDKs (dados como parte da estrutura)

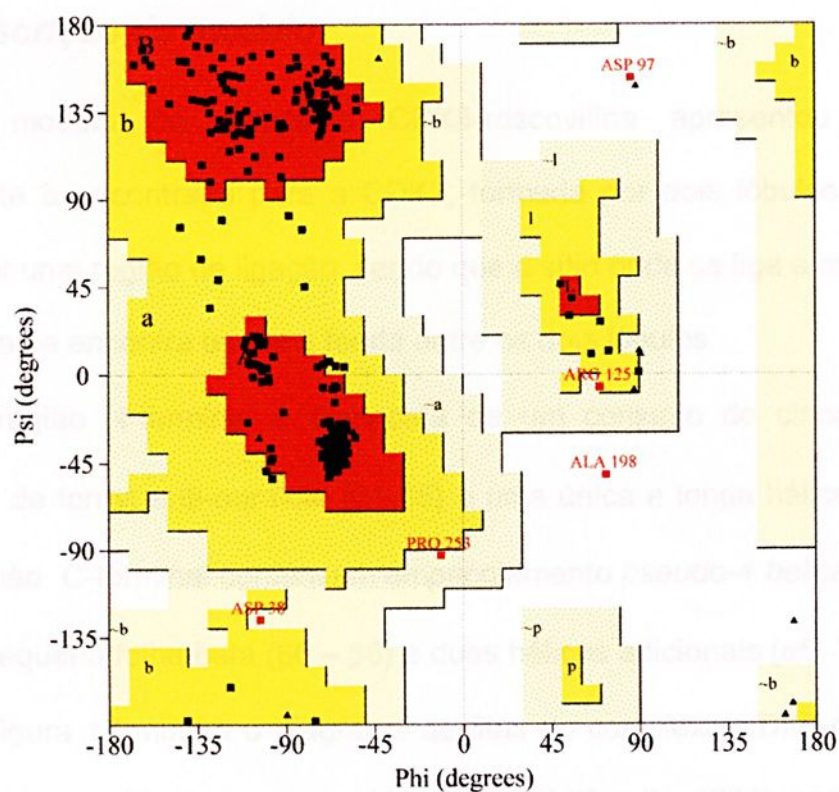


Figura 10: Gráfico de Ramachandran para o modelo da CDK5 complexada com roscovitina.

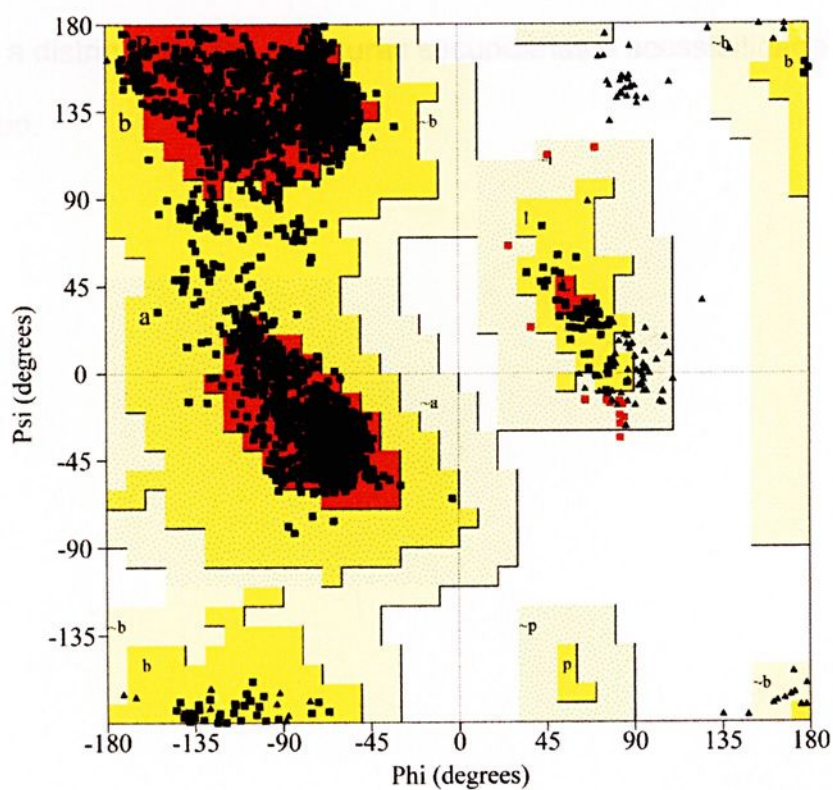


Figura 11: Gráfico de Ramachandran para as 8 CDKs usadas como padrão (templates)

3.2 - Descrição do modelo

O modelo do complexo CDK5-roscovitina apresentou estrutura semelhante à encontrada para a CDK2, formada por dois lóbulos principais unidos por uma região de ligação, sendo que o sítio onde se liga a molécula de roscovitina 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 ($\beta 1$ - $\beta 5$) e uma única e longa hélice alfa ($\alpha 1$). Já a região C-terminal contém um empacotamento *pseudo-4-helical* ($\alpha 2$, 3, 4, 6), uma pequena folha beta ($\beta 6$ – $\beta 8$) e duas hélices adicionais ($\alpha 5$, 7).

A figura 12 mostra o diagrama de fitas do complexo CDK5-roscovitina, gerado com o auxílio do programa MOLSCRIPT (Kraulis, 1991), onde vemos a molécula de roscovitina corretamente posicionada no sítio de ligação. Na figura 13 temos a distribuição das estruturas secundárias e acessibilidade ao solvente por resíduo.

Figura 12: Diagrama de fitas do complexo CDK5-roscovitina.

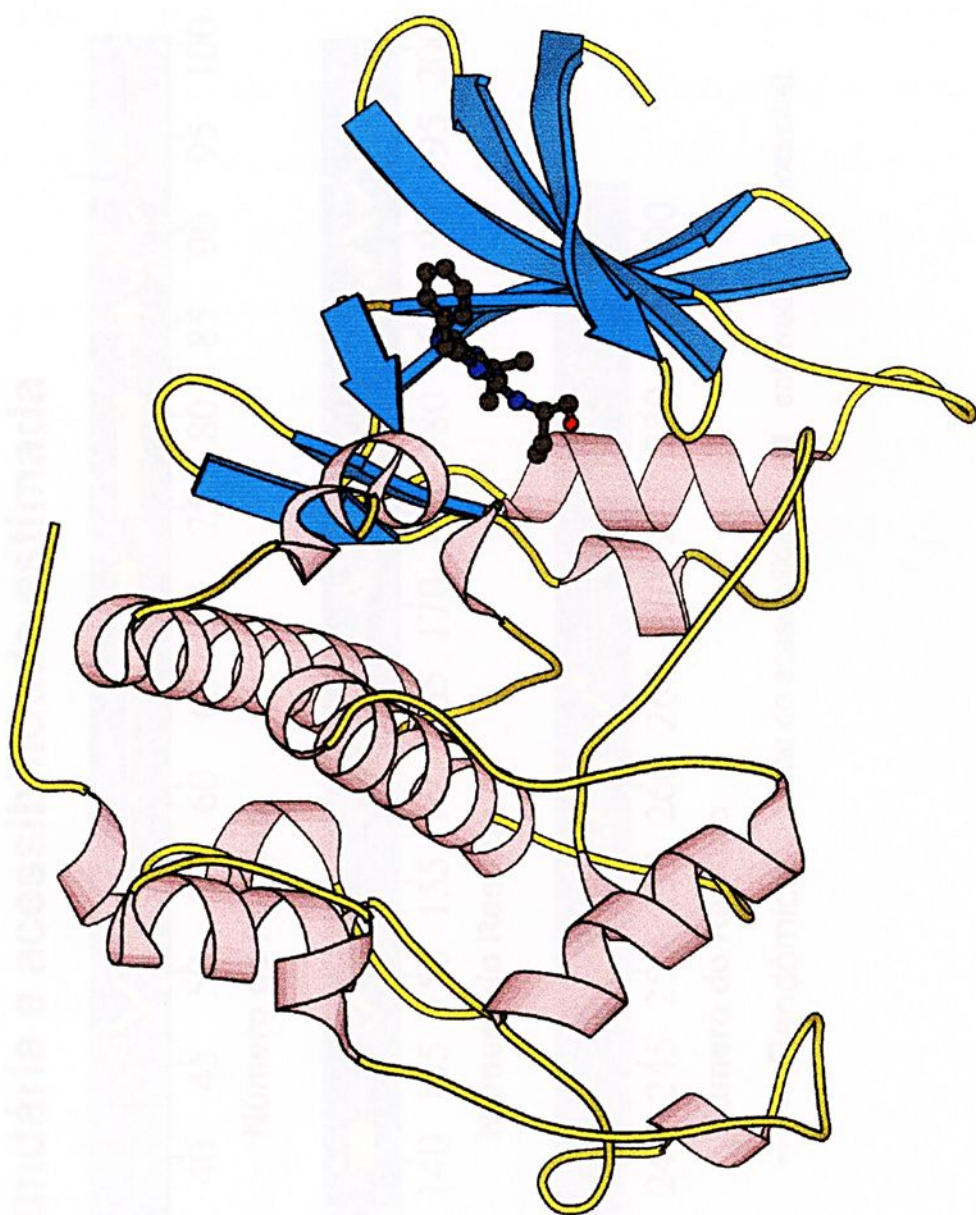


Figura 12: Diagrama de fitas do complexo CDK5-roscovitina

Estrutura secundária e acessibilidade estimada

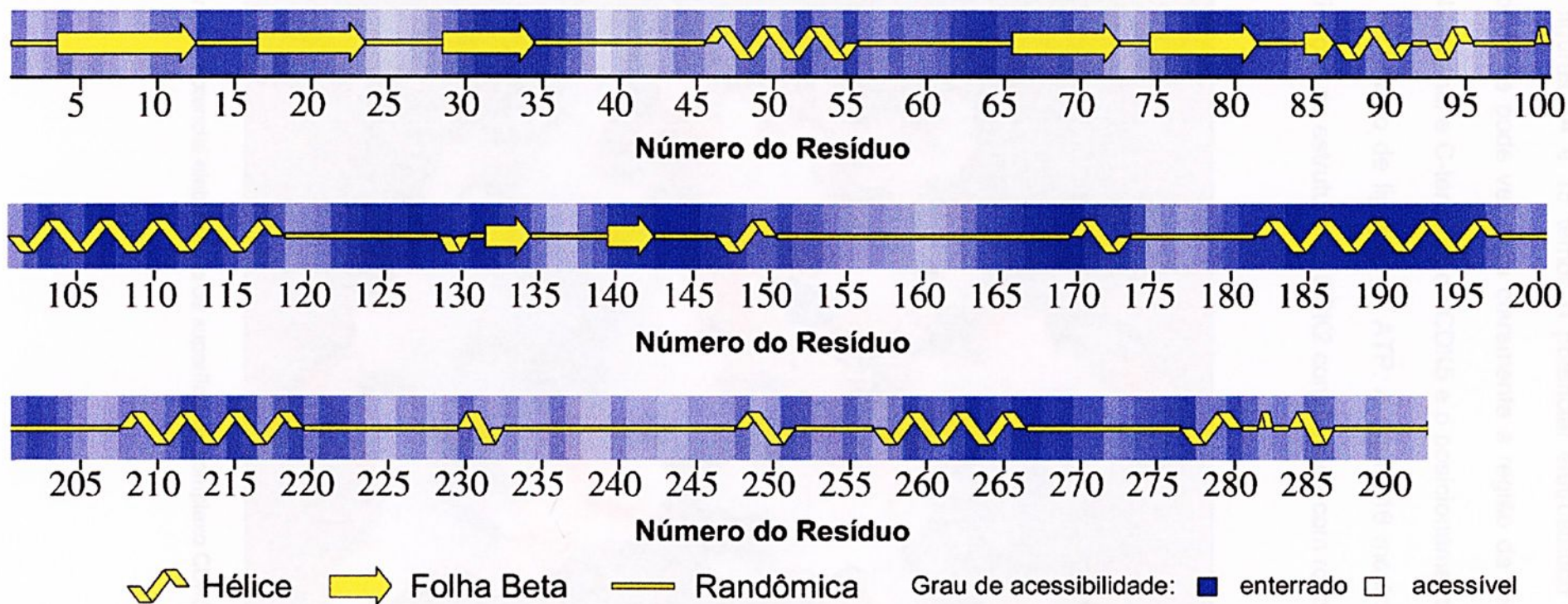


Figura 13: Análise da estrutura secundária e nível de acessibilidade ao solvente por resíduo (PROCHECK)

Nas figuras 14 e 15 temos o potencial eletrostático da superfície do complexo onde se pode ver mais claramente a região da rachadura entre os domínios N-terminal e C-terminal da CDK5 e o posicionamento da molécula de roscovitina no bolsão de ligação de ATP. A figura 16 mostra os detalhes do bolsão de ligação da estrutura da CDK2 complexada com roscovitina.

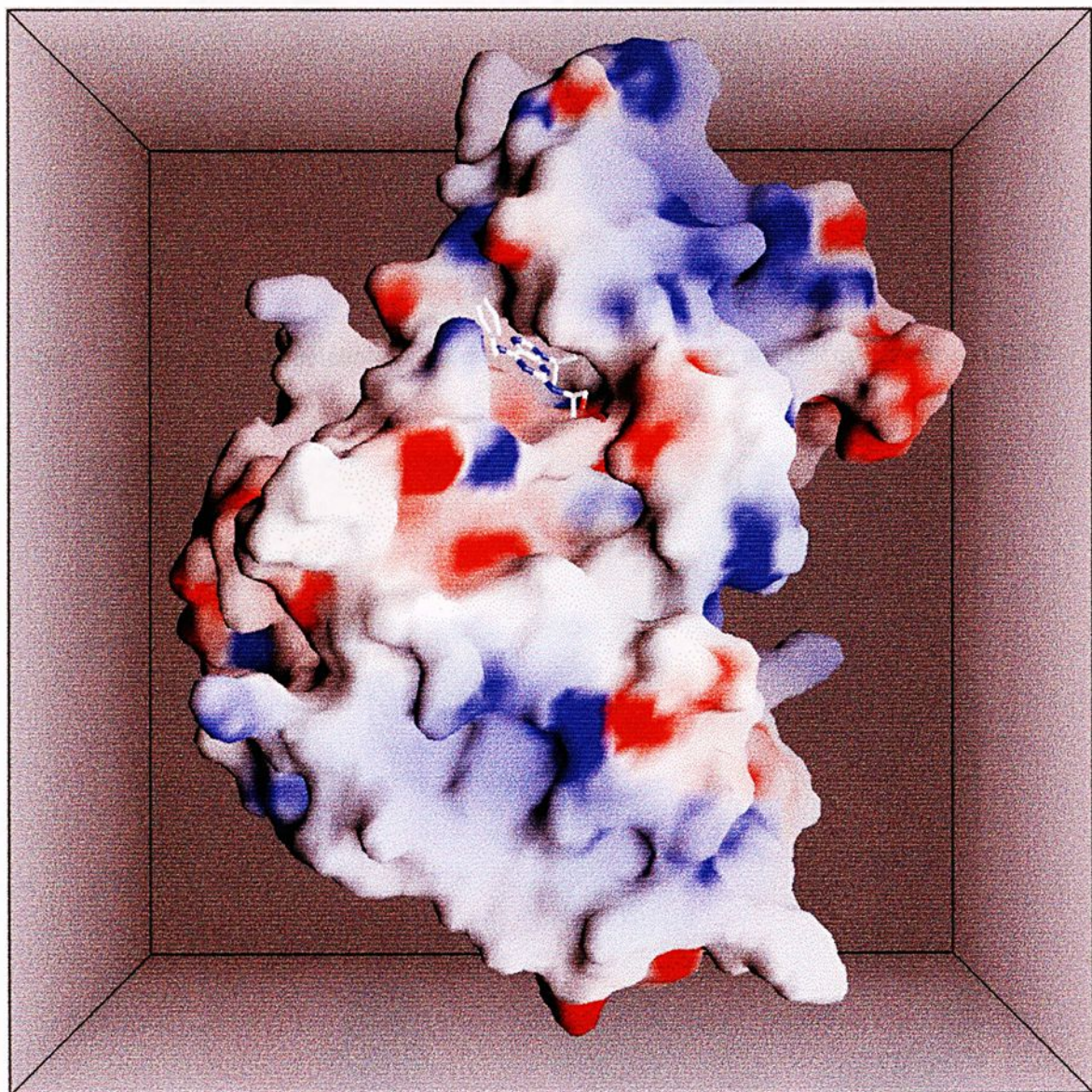


Figura 14: Potencial eletrostático da superfície do complexo CDK5-roscovitina

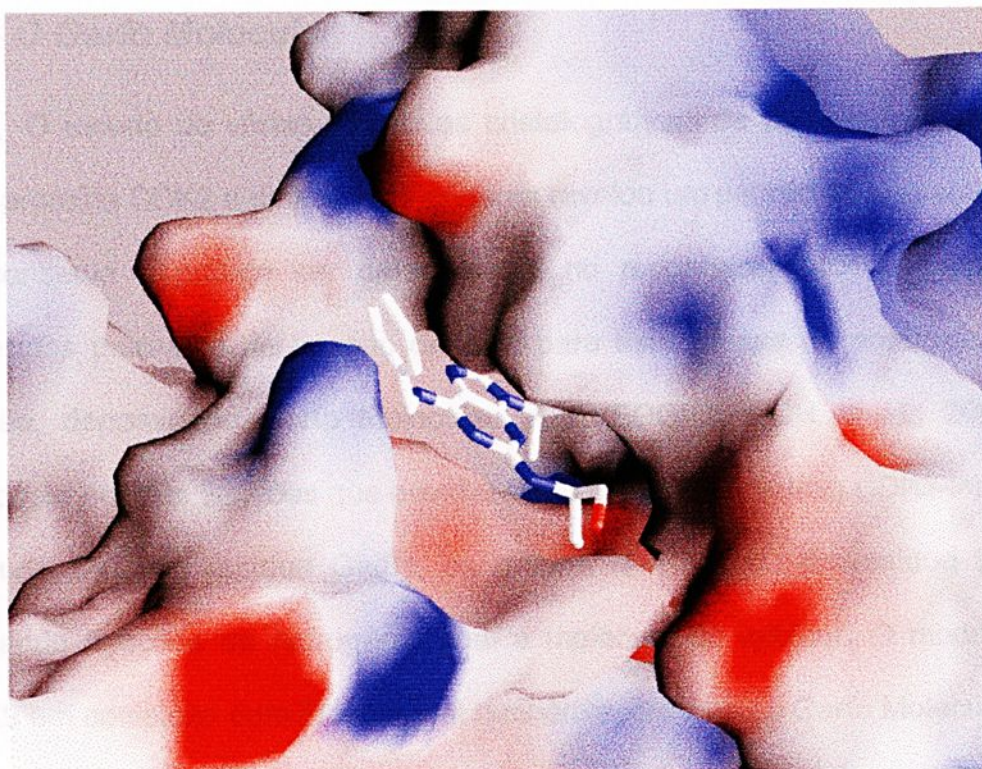


Figura 15: Potencial eletrostático da superfície do sítio de ligação de ATP do complexo CDK5-roscovitina

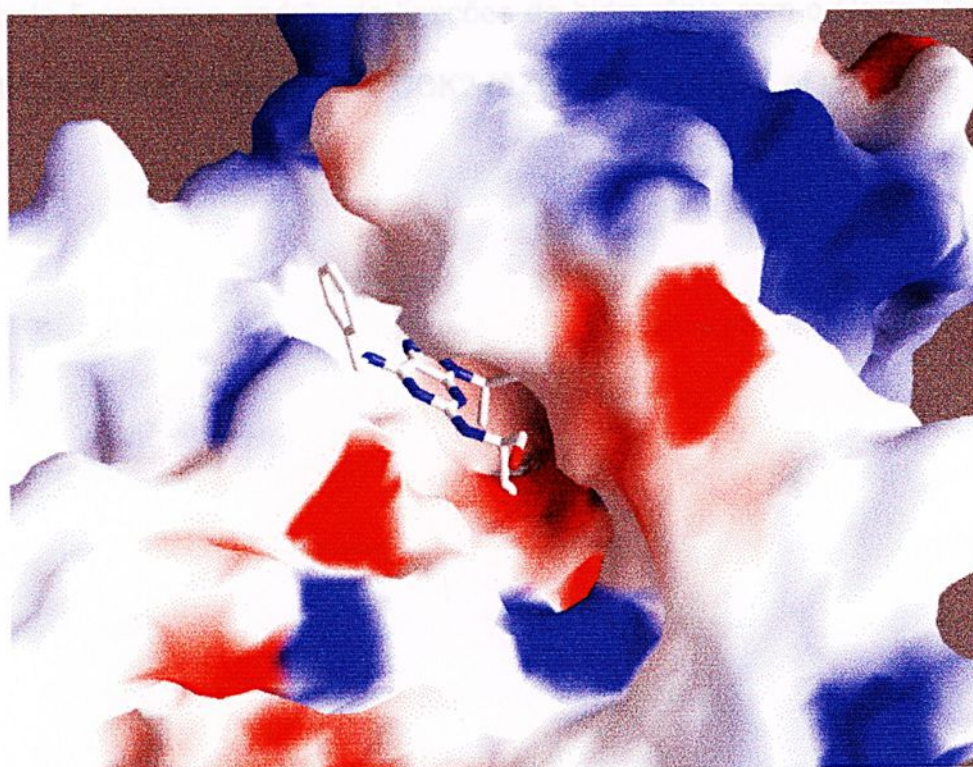


Figura 16: Potencial eletrostático da superfície do sítio de ligação de ATP do complexo CDK2-roscovitina

3.3 - O Garfo Molecular

1 O estudo de várias estruturas cristalográficas de complexos formados pela proteína CDK2 e diferentes inibidores revelou 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 (AZEVEDO JUNIOR et al., 2002) é composto por três ligações que envolvem, no caso do complexo CDK2-inibidor, o grupo C=O do resíduo Glu81, os grupos N-H e C=O do resíduo Leu83 formando ligações de hidrogênio com a molécula do inibidor. O modelo da CDK9 em complexo com Flavopiridol também apresentou o Garfo Molecular na análise das interações existentes entre a proteína e o inibidor (AZEVEDO JUNIOR et al., 2002).

A tabela 5 mostra o padrão de ligações de hidrogênio com o Garfo Molecular identificadas nos complexos de CDK2 (AZEVEDO JUNIOR et al., 2002).



Tabela 5: Ligações de hidrogênio entre o Garfo Molecular da CDK2 e diferentes inibidores.

Inibidor	Distância entre o inibidor e C=O em Glu81 (Å)	Distância entre o inibidor e N-H em Leu83 (Å)	Distância entre o inibidor e C=O em Leu83 (Å)
Staurosporina	2,75	2,63	NO
Purvalanol B	3,73	3,18	2,54
Himentialdisina	2,76	2,74	3,17
NU2058	2,57	3,11	2,34
NU6027	2,84	2,97	2,49
U55	NO	3,30	2,74
PKFO 49-38	2,71	3,13	3,27
Descloflavopiridol	2,86	2,91	NO
Roscovitina	NO	3,38	2,82
Olomoucina	NO	2,94	2,65
Isopenteniladenina	NO	3,39	NO
Indirubina	3,04	2,72	3,10

(NO – Não observado)

Um diagrama esquemático das interações que compõem o Garfo Molecular é apresentado na figura 17 onde vemos os átomos que compõem a cadeia principal da proteína, na região do Garfo Molecular e os grupos que contêm os doadores e aceptores de prótons que formam as ligações de hidrogênio que caracterizam este tipo de interação.

No modelo da CDK5-roscovitina encontramos também o Garfo Molecular composto por interações dos grupos C=O do resíduo Glu81, N-H e C=O do

resíduo Cys83, os quais formam ligações de hidrogênio com a molécula de roscovitina.

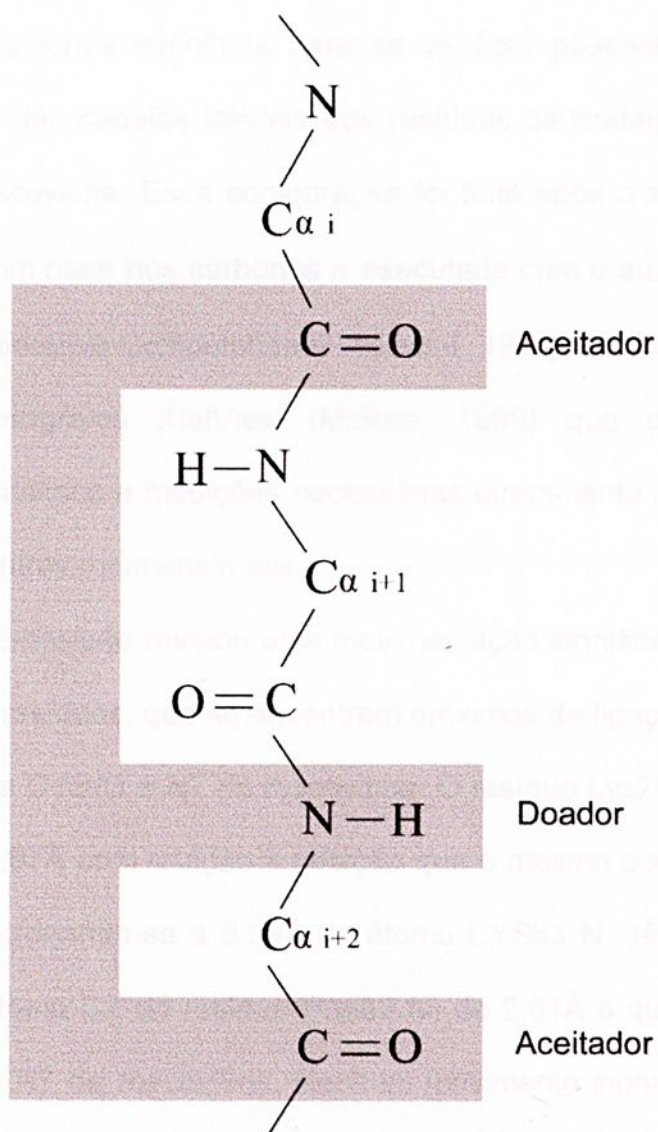


Figura 17: Diagrama esquemático do Garfo Molecular

3.4 - Posições das cadeias laterais

A comparação dos sítios de ligação dos complexos CDK5-roscovitina e CDK2-ATP foi a forma escolhida para se verificar possíveis alterações no posicionamento das cadeias laterais dos resíduos causadas pela ligação da molécula de roscovitina. Essa comparação foi feita após a superposição dos dois modelos com base nos carbonos α executada com o auxílio do programa LSQKAB (Collaborative Computational Project 4, 1994). Para esta comparação utilizou-se o programa XtalView (McRee, 1999) que permite executar visualizações, análises e medições necessárias diretamente dos arquivos que contém as estruturas tridimensionais.

Esta superposição revelou uma movimentação significativa das cadeias laterais de dois resíduos, que se encontram próximos da ligação de hidrogênio dos átomos N da CYS83 e N7 da roscovitina. O resíduo Lys20 teve seu átomo NZ deslocado 1,97Å com relação à posição que o mesmo ocupa no complexo CDK2-ATP posicionando-se a 6,59Å do átomo CYS83 N. Já o deslocamento observado no átomo CZ do resíduo Phe82 foi de 2,91Å o qual se encontra a 4,29Å do átomo N7 da roscovitina. Nenhum movimento significativo pode ser observado nas cadeias laterais dos resíduos na região do sítio de ligação de ATP do complexo CDK5-roscovitina o que indica que este inibidor se ajusta muito bem ao sítio de ligação causando apenas pequenos ajustes nas cadeias laterais da região por ele ocupada.



3.5 - Interações da roscovitina com a CDK5

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., 2001; AZEVEDO JUNIOR et al., 2001; Kim et al., 1996). No caso do modelo apresentado para o complexo CDK5-roscovitina um total de 5 ligações de hidrogênio foram identificadas nas interações entre as duas moléculas (Figura 18). Tais interações envolvem os resíduos Cys83, Asp86 e Gln130. Já na estrutura do complexo CDK2-roscovitina apenas 2 ligações de hidrogênio foram observadas (Figura 19). A tabela 6 descreve as interações citadas para os 2 complexos.

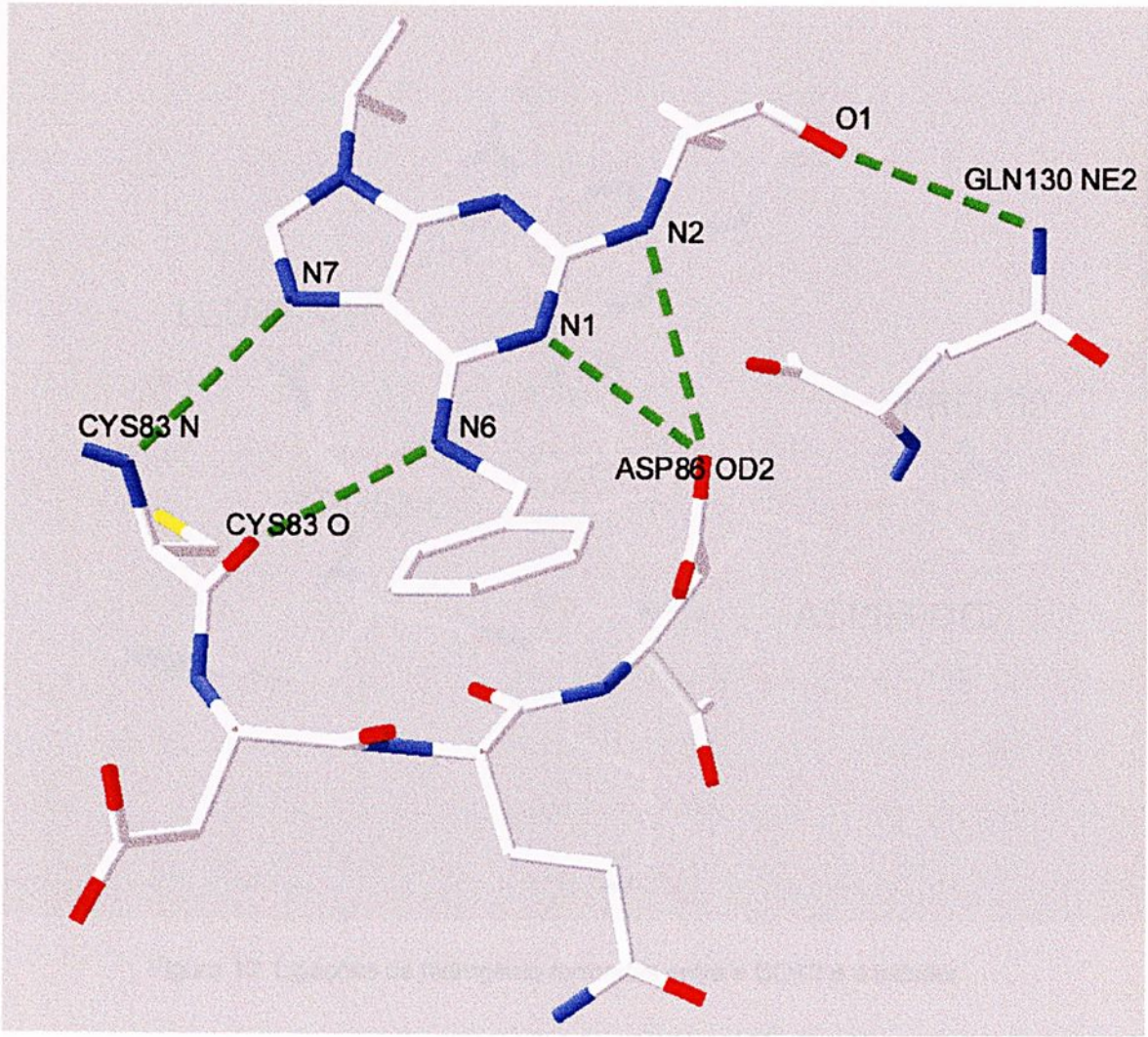


Figura 18: Ligações de hidrogênio formadas entre a CDK5 e o inibidor
Imagem gerada com o programa:
Swiss-PdbViewer V3.51 (Glaxo Wellcome Experimental Research)

Resíduo	CDK5	Inibidor (A)	Distância (Å)
N7	CDK5	CYS83 O	2,82
N7	CDK5	CYS83 N	3,38
N6	CDK5	CYS83 O	3,10
N7	CDK5	CYS83 N	3,83
O1	CDK5	GLN130 NE2	2,25
N2	CDK5	ASP86 OD2	3,54
N1	CDK5	ASP86 OD2	3,15

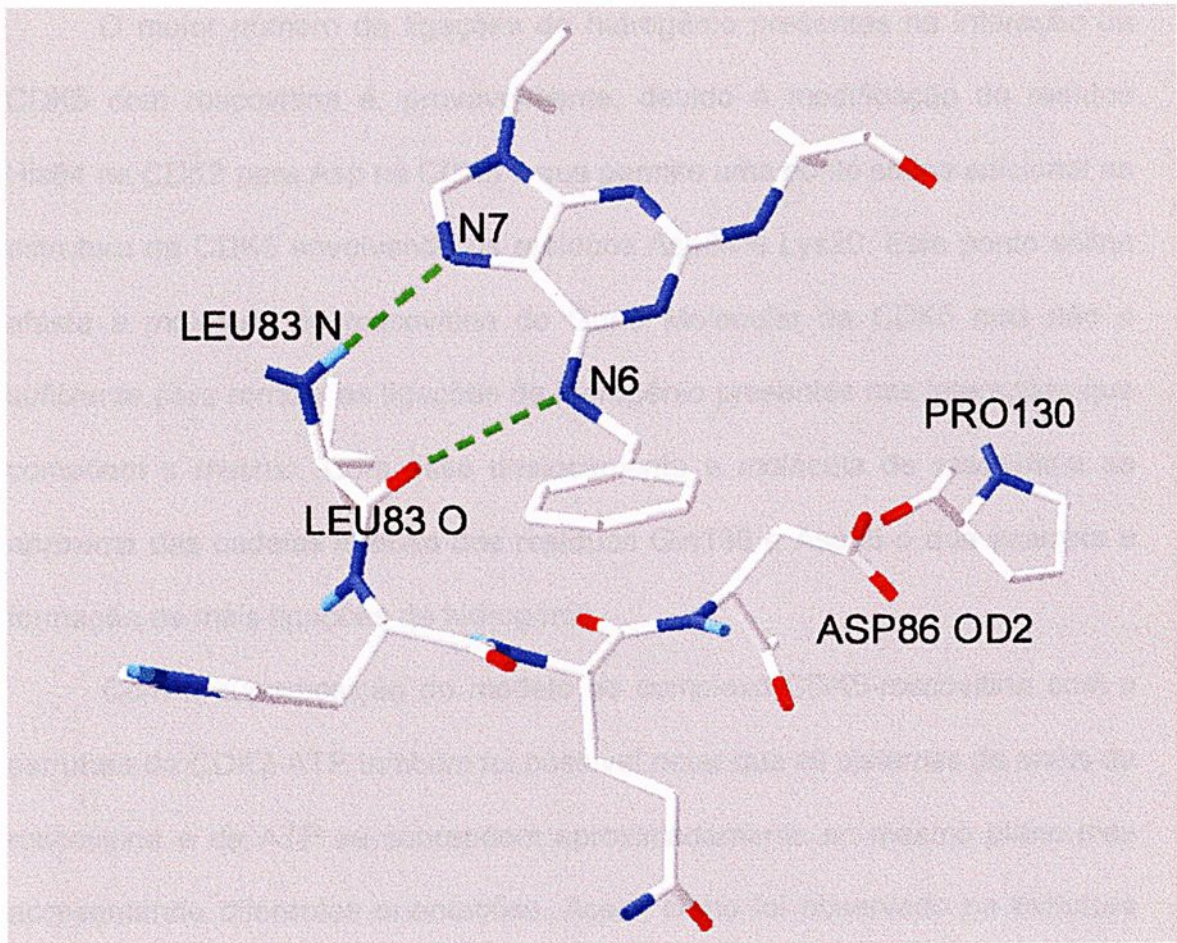


Figura 19: Ligações de hidrogênio formadas entre a CDK2 e o inibidor

Tabela 6: Ligações de hidrogênio formadas nos complexos CDK-roscovitina

Roscovitina	CDK	Resíduo, átomo	Distância (Å)
N6	CDK2	Leu83 O	2,82
N7	CDK2	Leu83 N	3,38
N6	CDK5	Cys83 O	3,10
N7	CDK5	Cys83 N	3,53
O1	CDK5	Gln130 NE2	3,29
N2	CDK5	Asp86 OD2	3,54
N1	CDK5	Asp86 OD2	3,16

O maior número de ligações de hidrogênio presentes na interação da CDK5 com roscovitina é, provavelmente, devido à modificação do resíduo His84 na CDK2 para Asp na CDK5 o que permite uma ponte salina adicional na estrutura da CDK5 envolvendo os resíduos Asp84 e Lys20. Esta ponte salina afasta a molécula de roscovitina do Garfo Molecular da CDK5 mas não o suficiente para romper as ligações de hidrogênio presentes nas interações que compõem o mesmo. Com esse deslocamento a molécula de roscovitina se aproxima das cadeias laterais dos resíduos Gln130 e Asp86 o que acarreta a formação de mais ligações de hidrogênio.

Com a superposição do modelo do complexo CDK5-roscovitina com a estrutura da CDK2-ATP também foi possível notar que os sistemas de anéis da roscovitina e da ATP se sobrepõem aproximadamente no mesmo plano mas apresentando diferentes orientações. Assim como foi observado na estrutura cristalográfica da CDK2 em complexo com roscovitina, (AZEVEDO JUNIOR et al., 1997) na CDK5, o anel fenil da molécula de roscovitina aponta para fora do sítio de ligação de ATP e se encontra parcialmente exposto ao solvente. A figura 20 mostra o sítio de ligação de ATP dos complexos CDK2-ATP e CDK5-roscovitina.

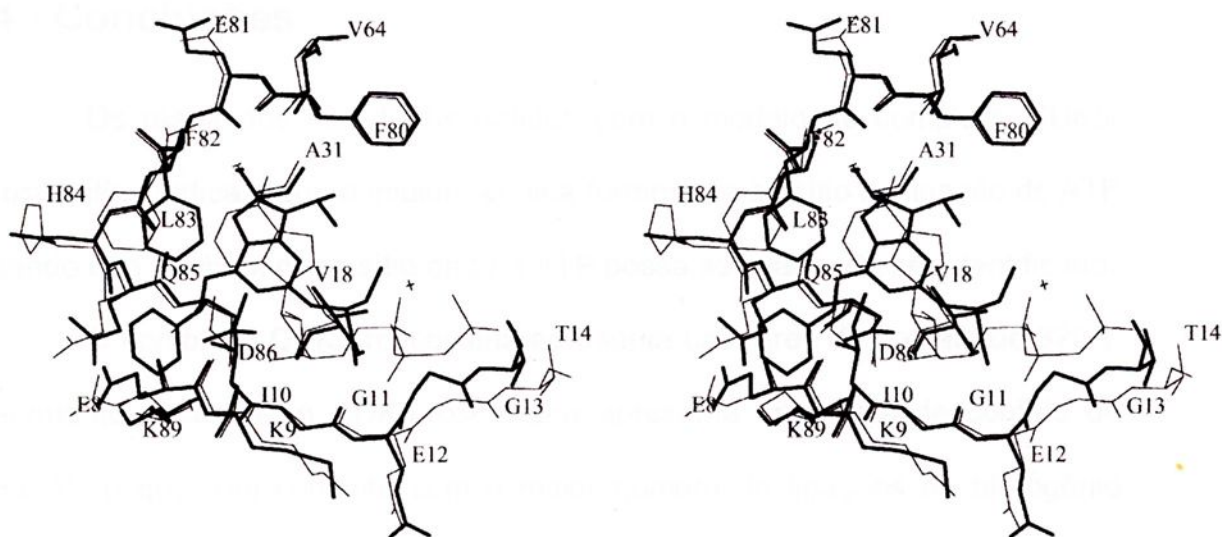


Figura 20: Superposição dos sítios de ligação dos complexos CDK5-roscovitina (linha grossa) e CDK2-ATP (linha fina)

A análise do potencial eletrostático das superfícies foi feita com o auxílio do programa GRASP (Nicholls et al., 1991) que gerou as figuras para as moléculas de CDK2 e CDK5 em complexo com roscovitina. Tal análise 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. Estas mesmas características de hidrofobicidade do sítio de ligação e complementaridade de cargas com o inibidor também foram encontradas na análise da modelagem molecular do complexo CDK9-flavopiridol (AZEVEDO JUNIOR et al., 2002).

4 - Conclusões

Os resultados estruturais obtidos com o modelo do complexo CDK5-roscovitina indicam que o inibidor se liga fortemente ao sítio de ligação de ATP sendo que nenhum outro sítio onde a ATP possa se ligar pode ser identificado.

O complexo CDK2-roscovitina apresenta uma área de contato de $320\text{\AA}^2$ enquanto o complexo CDK5-roscovitina apresenta uma área de contato de $334\text{\AA}^2$, o que, em conjunto com o maior número de ligações de hidrogênio observadas, pode parcialmente explicar a maior afinidade do inibidor pela molécula de CDK5. O valor de IC_{50} para inibição da atividade da CDK2 por roscovitina é de $0,70\mu\text{M}$ enquanto que para a CDK5 este valor é de apenas $0,16\mu\text{M}$ (Meijer et al., 1997).

A análise das informações estruturais fornecidas pelos modelos em conjunto com estudos de atividade indica 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.

O Garfo Molecular identificado na estrutura da CDK2 e no modelo de estrutura da CDK9 parece estar presente na estrutura da CDK5 e participa na formação do conjunto de ligações de hidrogênio na interação com o inibidor, o que sugere que a CDK5 pode também ser fortemente inibida por outros inibidores da CDK2. Tal fato abre a possibilidade para a utilização deste método nos testes de novos inibidores.

Devido ao sucesso do protocolo de modelagem usado para CDK5 pretendemos usá-lo na modelagem de outras CDKs, o que permitirá entender

as bases estruturais para os diferentes perfis de inibição apresentados por diversos inibidores de CDKs.



Referências:

ARRIS, C. E. et al. Identification of novel purine and pyrimidine cyclin-dependent kinase inhibitors with distinct molecular interactions and tumor cell growth inhibition profiles. *J. Med. Chem.*, v.43 p. 2797, 2000.

BERMAN, H. M. et al. The Protein data bank. *Nucl. Acids Res.*, v.28, p. 235-242, 2000.

BLUM, G. et al. Substrate competitive inhibitors of igf-1 receptor kinase. *Biochemistry*, v.39, p.15705-15712, 2000.

BRANDEN, C.; TOOZE, J. *Introduction to protein structure*. 2nd ed. New York : Garland, 1998.

BRAUN, W.; GO, N. Calculation of protein conformations by proton-proton distance constraints: a new efficient algorithm. *J. Mol. Biol.*, v.186, n.3, p.611-626, 1985.

BROADBRIDGE, R. J.; SHARMA, R. P. The src homology-2 domains (sh2 domains) of the protein tyrosine kinase p56lck: structure, mechanism and drug design. *Curr. Drug Targets*, v.1, p.365-386, 2000.

BROOKS, B. R. et al. CHARMM: A program for macromolecular energy minimization and dynamics calculations. *J. Comput. Chem.*, v.4, p.187-217, 1983.

CANDURI, F. et al. Crystal structure of human uropepsin at 2.45Å resolution. *Acta Crystallogr.*, v.57, p.1560-1570, 2001.

CLARE, P. M.; POORMAN, R. A.; KELLEY, L. C. The cyclin-dependent kinases cdk2 and cdk5 Act by a random, anticooperative kinetic mechanism. *J. Biol. Chem.*, v.276, p.48292-48299, 2001.

CODY, W. L. et al. Progress in the development of inhibitors of sh2 domains. *Curr. Pharm. Design*, v.6, p.59-98, 2000.



Collaborative Computational 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. *Nucl. Acids Res.*, v.16, n.22, p.10881-10890, 1988.

AZEVEDO JUNIOR, W. F.; LECLERC, S.; MEIJER, L. Inhibition of cyclin-dependent kinases by purine analogues: crystal structure of human cdk2 complexed with roscovitine. *Eur. J. Biochem.*, v.243, p.518-526, 1997.

AZEVEDO JUNIOR, W. F.; CANDURI, F.; FADEL, V. Molecular model for the binary complex of uropepsin and pepstatin. *Biochem. Biophys. Res. Comun.*, v.287, n.1, p.277-281, 2001.

AZEVEDO JUNIOR, W. F.; CANDURI, F.; SILVEIRA, N. J. F. Structural basis for inhibition of cyclin-dependent kinase 9 by Flavopiridol. *Biochem. Biophys. Res. Comun.*, v.293, n.1, p.566-571, 2002.

GRAY, N. S.; WODICKA, L.; THUNNISSEN, A. M. Exploiting chemical libraries, structure, and genomics in the search for kinase inhibitors. *Science*, v.281, p.533, 1998.

HANKS, S. K.; HUNTER, T. Protein kinases 6: The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB*, v.9, p.576-596, 1995.

HIGGINS, D.G.; BLEASBY, A.J.; FUCHS, R. Clustal v: improved software for multiple sequence alignment. *Comput. Appl. Biosci.*, v.8, n.2, p.189 -191, 1992.

HUNTER, T. Protein kinases and phosphatases: the yin and yang of protein phosphorylation and signaling. *Cell*, v.80, p.225-236, 1995.

JOHNSON, L. N.; LEWIS, R. J. Structural basis for control by phosphorylation. *Chem. Rev.*, v.101, p.2209-2242, 2001.

KABSCH, W. A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr.*, v.32, p.922-923, 1976.



KIM, S. H.; SCHULZE-GAHMEN, U.; BRANDSEN, J. Structural basis for chemical inhibitor of cdk2. *Prog. Cell Cycle Res.*, v.2, p.137-145, 1996.

KRAULIS, P. J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, v.24, p.946-950, 1991.

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

LAWRENCE, D.S.; NIU, J. Protein kinase inhibitors: the tyrosine-specific protein kinases. *Pharmacol. Ther.*, v.77, p.81-114, 1998.

LAWRIE, A. M.; NOBLE, M. E.; TUNNAH, P. Protein kinase inhibition by staurosporine revealed in details of the molecular interaction with cdk2. *Nat. Struct. Biol.*, v.4, p.796, 1997.

LEE, M. S. et al. Neurotoxicity induces cleavage of p35 to p25 by calpain. *Nature*, v.405, p.360-364, 2000.

MACCIONI, R. B. et al. The protein kinase cdk5 structural aspects, roles in neurogenesis and involvement in alzheimer's pathology. *Eur. J. Biochem.*, v.268, p.1518-1527, 2001.

MARTI-RENOM, M. A.; YERKOVICH, B.; SALI, A. Modeling protein structure from its sequence. *Curr. Prot. Bioinf.*, 2003. No prelo

McREE, D. E. XtalView/Xfit: a versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.*, v.125, p.156-165, 1999.

MEIJER, L.; BORGNE, A.; MULNER, O. Biochemical and cellular effects of roscovitine, a potent and selective inhibitor of the cyclin-dependent kinases cdk2 and cdk5. *Eur. J. Biochem.*, v.243, p.527-536, 1997.

MEIJER, L.; THUNNISSEN, A. M.; WHITE, A. W. Inhibition of cyclin-dependent kinases, gsk-3beta and ck1 by Hymenialdisine, a marine sponge constituent. *Chem. Biol.*, v.7, p.51, 2000.



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

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

NIKOLIC, M. et al. The cdk5/p35 kinase is essential for neurite outgrowth during neuronal differentiation. *Genes Dev.*, v.10, p.816-825, 1996.

OHSHIMA, T. et al. Targeted disruption of the cyclin-dependent kinase 5 gene results in abnormal corticogenesis, neuronal pathology and perinatal death. *Proc. Natl. Acad. Sci. U.S.A.*, v.93, n.20, p.11173-11178, out.1996.

PARANG, K. et al. Mechanism-based design of a protein kinase inhibitor. *Nat. Struct. Biol.*, v.8, p.37-41, 2001.

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

SANCHEZ R; SALI A. Comparative protein structure modeling. Introduction and practical examples with modeller. *Methods Mol. Biol.*, v.143, p.97-129, 2000.

SCAPIN G. Structural biology in drug design: selective protein kinase inhibitors. *Drug Discov. Today*, v.7, n.11, p.601-611, 2002.

SCHULZE-GAHMEN, U.; BRANDSEN, J.; JONES, H.D. Multiple modes of ligand recognition: crystal structures of cyclin-dependent protein kinase 2 in complex with ATP and two inhibitors, Olomoucine and Isopentenyladenine. *Proteins*, v.22, p.378, 1995.

SCHULZE-GAHMEN, U.; DE BONDT, H. L.; KIM, S. H. High-resolution crystal structures of human cyclin-dependent kinase 2 with and without ATP: bound waters and natural ligand as guides for inhibitor design. *J Med Chem.*, v.39, p.4540, 1996.

TAYLOR, S. S.; RADZIO-ANDZELM, E. Protein kinase inhibition: natural and synthetic variations on a theme. *Curr. Opin. Chem. Biol.*, v.1, p.219-226, 1997.

VOET, D.; VOET, J. G. Biochemistry, 2nd ed., New York: John Wiley, 1995.

VU, C. B. Recent advances in the design and synthesis of sh2 inhibitors of src, grb2 and zap-70. *Curr. Med. Chem.*, v.7, p.1081-1100, 2000.

YEH, R. H. et al. From consensus sequence peptide to high affinity ligand, a "library scan" strategy. *J. Biol. Chem.*, v.276, p.12235-12240, 2001.



Apêndice A – Modelagem comparativa de estruturas

Abordagens para a determinação e predição da estrutura tridimensional de proteínas são classificadas com base no tipo predominante de informações utilizadas para se calcular os modelos.

A modelagem comparativa usa estruturas de proteínas obtidas experimentalmente para prever a conformação de outras proteínas com sequências de aminoácidos similares. Isso é possível porque uma pequena mudança na sequência normalmente resulta em uma pequena mudança na estrutura (Hubbard & Blundell, 1987; Lesk & Chothia, 1986).

A modelagem comparativa apresenta melhores resultados que outros métodos teóricos *ab initio* e produz modelos com valores de RMSD tão baixos quanto 1Å para sequências que tenham homologia suficiente [acima de 40%] com as estruturas 3D conhecidas, (Topham et al., 1991). Já os métodos de predição física apresentam valores de RMSD de aproximadamente 3,5Å para pequenas proteínas (Cohen & Kuntz, 1989; Wilson & Doniach, 1989). Por outro lado, a modelagem comparativa não atinge os níveis de exatidão obtidos com as técnicas de cristalografia por difração de raios X e NMR (Ressonância Magnética Nuclear) com as quais se obtêm estruturas com valores de RMSD de aproximadamente 0,3 e 0,5Å, respectivamente.

MODELLER é um programa de computador que modela estruturas 3D de proteínas utilizando a técnica de satisfação de restrições espaciais.

O MODELLER é mais freqüentemente utilizado para modelagem comparativa de proteínas ou modelagem por homologia. O usuário fornece um alinhamento da sequência de aminoácidos a ser modelada com estruturas



relacionadas conhecidas e o MODELLER automaticamente calcula o modelo com todos os átomos que não sejam hidrogênio.

De uma forma geral, as entradas para o programa são restrições na estrutura espacial da sequência de aminoácidos e ligantes a ser modelada. A saída é uma estrutura tridimensional que satisfaz essas restrições da melhor forma possível. Restrições podem, em princípio, ser derivadas de um grande número de diferentes fontes. Isso inclui estruturas de proteínas relacionadas (modelagem comparativa), experimentos de NMR (refinamento de NMR), experimentos de *cross-linking*, espectroscopia de fluorescência, reconstrução de imagens em microscopia eletrônica, mutagênese sítio dirigida, potenciais de força média (*potentials of mean force*) de resíduo-resíduo ou átomo-átomo. As restrições podem operar em distâncias, ângulos, ângulos diedrais, pares de ângulos diedrais e algumas outras características espaciais definidas para átomos e pseudo-átomos. Atualmente, o MODELLER somente deriva de forma automática as restrições a partir de estruturas tridimensionais relacionadas conhecidas e dos alinhamentos dessas estruturas com a sequência alvo.

O modelo tridimensional é obtido com a otimização da função de densidade de probabilidade (pdf) molecular. A pdf molecular para a modelagem comparativa é otimizada pelo procedimento de função de alvo variável no espaço Cartesiano e emprega os métodos de gradientes conjugados e dinâmica molecular com recozimento simulado.

O programa pode também executar múltiplas comparações de sequências de proteínas e/ou estruturas, agrupamentos de proteínas e busca de sequências em bancos de dados. A utilização do programa é feita através de uma linguagem de script e não inclui nenhuma interface gráfica. A versão



atual foi escrita em FORTRAN 90 padrão com a intenção de ser executável em computadores com sistema operacional UNIX.

O método

O Método de modelagem comparativa foi desenvolvido de forma a utilizar a maior quantidade de informações possível sobre a estrutura a ser modelada e consiste em três estágios (Figura 21):

- (1) alinhamento da sequência a ser modelada com as estruturas e segmentos relacionados,
- (2) determinação das restrições espaciais na sequência utilizando os alinhamentos,
- (3) satisfação das restrições espaciais e obtenção dos modelos 3D.

As restrições espaciais na sequência a ser modelada são obtidas a partir da análise estatística das relações entre várias características das estruturas de proteínas. Um banco de dados contendo alinhamentos de 17 famílias de proteínas em um total de 80 estruturas foi construído de forma a se obter tabelas quantificando os relacionamentos, tais como as relações entre as distâncias entre dois C^α - C^α equivalentes ou entre ângulos diedrais equivalentes em duas proteínas relacionadas. Esses relacionamentos foram descritos como funções de densidade de probabilidades (pdfs) para as características a serem preditas. Por exemplo, a probabilidade para diferentes valores dos ângulos diedrais da cadeia principal é calculada com base no tipo de resíduo, na conformação da cadeia principal de um resíduo equivalente e na similaridade entre as sequências das duas proteínas.



O cálculo do modelo tridimensional a partir da satisfação de restrições espaciais é feito com a otimização da chamada pdf molecular que é uma combinação de várias pdfs as quais definem as restrições de características espaciais individuais para a molécula. O método de otimização é função alvo variável (Braun & Go, 1985) que aplica o método de gradientes conjugados para as posições de todos os átomos que não sejam hidrogênio.



1. Alinhamento da sequência com as estruturas

3D

GRISFFEDAGF-GHCYECSSDC-NI

3D

GKISFYEDRGFGHCYECSSDC-NI

SEQ

GKITFYEDRG---RCYECSSDCPNI

2. Extração das restrições espaciais

GKITFYEDRGRCYECSSDCPNLDP

3. Satisfação das restrições espaciais

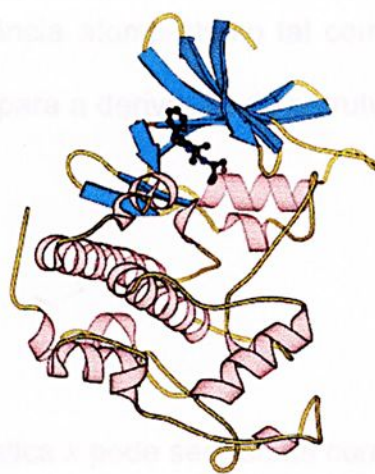


Figura 21: Principais passos do processo de modelagem.

Definição das restrições espaciais

A estrutura tridimensional de uma proteína pode ser bem determinada caso se obtenha um número suficiente de restrições espaciais fortes.

Uma restrição é mais precisamente definida em termos de uma função de densidade de probabilidade, $p(x)$, para uma característica x a se restringir. Tal função pode ser de qualquer forma, desde que não seja negativa, e que resulte 1 quando integrada sobre todos os valores possíveis de x . A probabilidade de um evento $x_1 \leq x \leq x_2$ é obtida pela integração de p :

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

A expressão de uma restrição em termos de uma pdf fornece mais informações sobre os possíveis valores da característica a ser restrita que somente a média dos valores medidos. Uma pdf é também mais completa que os limites superior e inferior de uma certa distância átomo-átomo tal como é utilizada na abordagem de distância geométrica para a derivação de estruturas 3D por dados de NMR multidimensional.

Derivação de pdfs básicas

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. Por exemplo, $p(\chi_1/\text{tipo de resíduo}, \Phi, \Psi)$ pode ser usado para prever o ângulo

diedral χ_1 da cadeia lateral com base no tipo de resíduo e nos ângulos Φ e Ψ da cadeia principal.

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 da sequência a ser modelada. A característica conhecida mais importante é aquela do mesmo tipo de x e associada com posições equivalentes nas estruturas conhecidas relacionadas. Por exemplo, para restringir uma certa distância $C^\alpha-C^\alpha$, a informação mais útil é a distância equivalente nas estruturas 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, como seus elementos, as frequências 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 frequências relativas W é calculada com base nas frequências absolutas W' , usando-se a seguinte equação:

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

As frequências absolutas W' são obtidas por contagem direta do número de ocorrências de cada combinação de valores $(x,a,b,...,c)$ presentes na



amostra, e tal amostra é derivada do banco de dados local de estruturas de proteínas conhecidas e seus alinhamentos.

O banco de dados local

Para a implementação do processo de modelagem comparativa foi criado, dentro do próprio MODELLER, um banco de dados local, o qual contém os dados estruturais de membros de 17 famílias de proteínas retirados do Brookhaven Protein Databank (Abola et al., 1987; Bernstein et al., 1977).

As famílias de proteínas escolhidas para compor o banco de dados local foram as seguintes: Proteinases Aspárticas, Serino Proteinases, Azurinas, Imunoglobulinas, Gama-Cristalinas, Cisteino Proteinases, Lisozimas, Globinas, Fosfolipases, 2 famílias de Citocromos, Centros de Reação Fotosintética, Ferredoxinas, Flavodoxinas, Dehidrofolato Redutases, Dehidrogenases e Oxidoredutases.

A tabela 7 apresenta mais detalhes sobre a composição deste banco de dados.



Tabela 7: Composição do banco de dados local

Alinhamentos	17
Estruturas de proteínas	80
Pares de proteínas homólogas	488
Resíduos	11.352
Pares de resíduos equivalentes	92.780
Pares de resíduos intramoleculares	1.888.881
Pares de resíduos intramoleculares equivalentes	20.270.936

Esse banco de dados local contém membros representativos de 4 classes estruturais de proteínas: α , β , $\alpha+\beta$ e α/β sendo a classe α formada predominantemente de hélices α , a classe β formada predominantemente por folhas β , a classe $\alpha+\beta$ contendo os dois tipos de estruturas secundárias e a classe α/β que apresenta um padrão característico de composição de estruturas α e β como visto na figura 22. A frequência de ocorrência de resíduos na conformação de folhas- β é similar à frequência de resíduos na conformação de hélice. A percentagem de identidade para as sequências das estruturas que compõem o banco de dados varia de 6% a 98%, o que também se reflete na diferença de vizinhança local de resíduos, para os pares de resíduos equivalentes. A maior parte das estruturas de proteínas que compõem o banco de dados local foi resolvida a média ou alta resolução, ou seja, acima de 2,5Å, sendo que também existem algumas estruturas resolvidas a baixa resolução.



Figura 22: Configuração típica das estruturas da classe α/β

Algumas características da composição do banco de dados local, como frequência de tipos de aminoácidos e a distribuição fracional de identidade de sequência, são mostradas na figura 23. A definição das classes de conformações da cadeia principal é mostrada na figura 24.

Por suas características, esse banco de dados de estruturas de proteínas e seus alinhamentos são considerados como amostras representativas das estruturas globulares de proteínas, sendo então adequados para se estabelecer os relacionamentos gerais entre as mais variadas características das estruturas de proteínas.

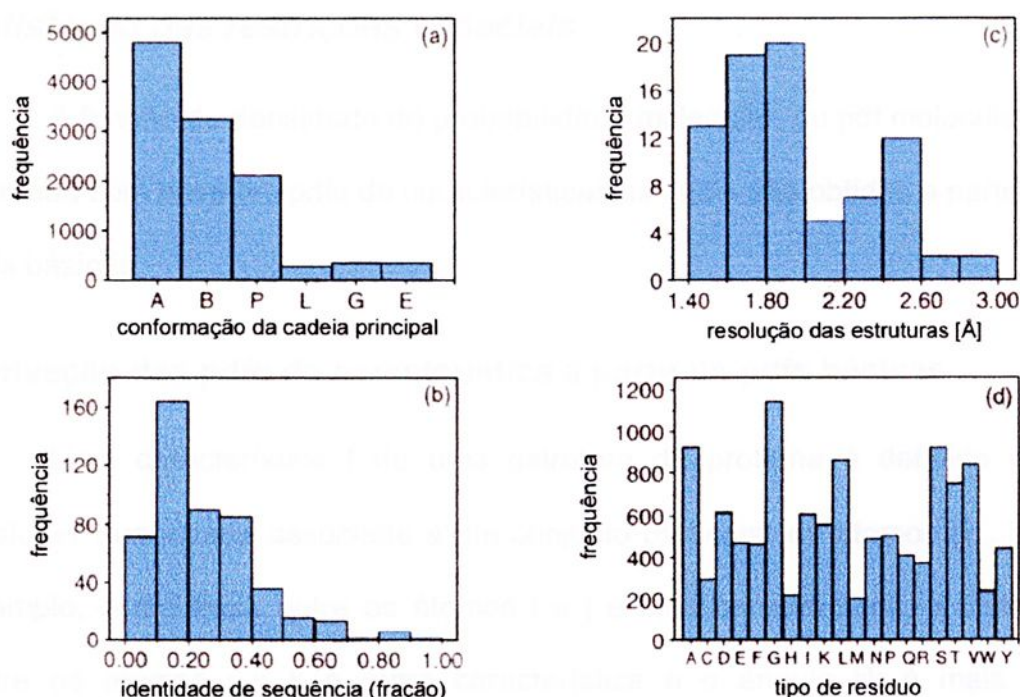


Figura 23: Composição do banco de dados local. (a) Classes de conformações da cadeia lateral para todos os resíduos. (b) Escore fracional de identidade de sequência para todos os pares de proteínas relacionadas. (c) Resolução da análise de Raios X para todas as proteínas. (d) Frequência de ocorrência por tipo de resíduos.

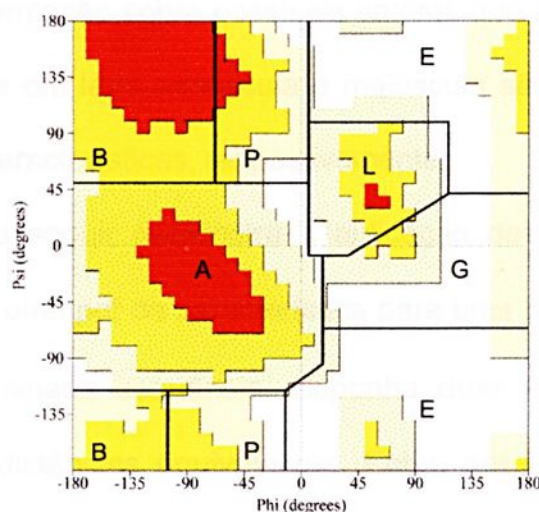


Figura 24: Definição das classes de conformações da cadeia principal.

Satisfação das restrições espaciais

A função de densidade de probabilidade molecular, ou pdf molecular, é montada com base em pdfs de características as quais são obtidas a partir das pdfs básicas.

Derivação das pdfs de característica a partir de pdfs básicas

Uma característica f de uma estrutura de proteína é definida como qualquer quantidade associada a um conjunto particular de átomos $ijk...l$. Por exemplo, a distância entre os átomos i e j é uma característica, a distância entre os átomos j e k é outra característica e o ângulo ijk é mais uma característica. As pdfs básicas $p^f(f)$ são aquelas já descritas anteriormente. De uma forma geral, cada característica estrutural f pode ser restrita por várias pdfs básicas $p_k^f(f)$ para $k=1,2,\dots$. Uma pdf de característica $p^F(f)$ é uma pdf que combina toda a informação sobre possíveis valores que a característica f pode assumir. Os índices em letra minúscula e maiúscula são usados para indicar pdfs básicas e de características, respectivamente.

O exemplo a seguir demonstra a aplicação das definições acima. O objetivo é construir uma pdf de característica para uma distância particular $C^\alpha-C^\alpha$ em uma determinada sequência. Suponha duas estruturas conhecidas relacionadas com distâncias equivalentes. Então nós temos 2 pdfs básicas correspondentes para a distância $C^\alpha-C^\alpha$ na sequência alvo. Além disso, nós também sabemos que cada uma das restrições deve satisfazer o critério de van der Waals, isto é, a distância tem que ser maior que a soma dos dois raios de van der Waals. De forma a poder combinar todas essas informações nós



temos que combinar as três pdfs básicas em uma única pdf de característica. Para fazer isso nós podemos utilizar todos os possíveis alinhamentos destas três proteínas.

Um exemplo da dependência da distância $C^\alpha - C^\alpha$ para duas distâncias equivalentes em duas estruturas, $p(d/d', d'')$, é mostrado na figura 25a.

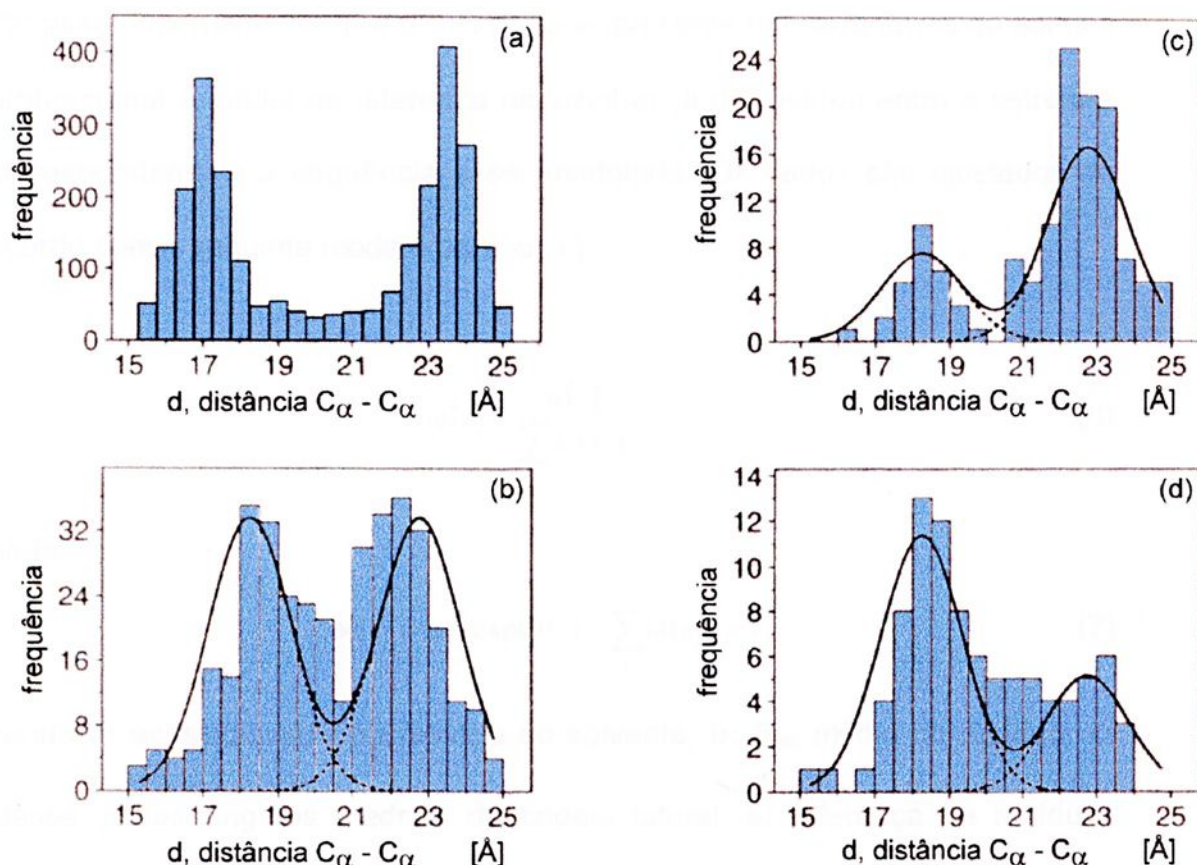


Figura 25: Derivação de uma pdf de características a partir de pdfs básicas.

Em todos os gráficos, $18,0 < d' < 18,5$ e $22,5 < d'' < 23,0$. (a) $W'(d/d', d'')$. (b)

$W'(d/d', d'', \bar{s}', \bar{s}'')$, onde $0,2 < \bar{s}' < 0,4$ e $0,2 < \bar{s}'' < 0,4$ (c) $W'(d/d', d'', \bar{s}', \bar{s}'')$, onde $0,2 < \bar{s}' < 0,4$ e $0,4 < \bar{s}'' < 0,6$ (d) $W'(d/d', d'', \bar{s}', \bar{s}'')$, onde $0,4 < \bar{s}' < 0,6$ e $0,2 < \bar{s}'' < 0,4$

O histograma sugere que $p(d/d', d'')$ pode ser calculada como uma soma ponderada das pdfs individuais $p(d/d')$ e $p(d/d'')$:

$$p(d/d', d'', \bar{s}', \bar{s}'') = \omega(\bar{s}') \cdot p(d/d') + \omega(\bar{s}'') \cdot p(d/d'') \quad (5)$$

Esta relação gera um ajuste dos dados claramente melhor que um simples produto das pdfs individuais o que acabaria fazendo com que a distância d se tornasse uma média de d' e d'' . O peso ω presente em cada termo da soma é proporcional à média da diferença na vizinhança do resíduo entre a estrutura correspondente e a sequência a ser modelada. Os dados são ajustados de acordo com o seguinte modelo para $\omega(s)$:

$$\omega(s) = \frac{w(s)}{\sum_j w(s_j)} \quad (6)$$

onde

$$w(s) = a + \exp(bs^c); \sum_j \omega(s_j) = 1 \quad (7)$$

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.

Exemplos de histogramas e curvas analíticas para esta pdf de características correspondentes a diferentes pesos são mostradas na figura 25 (b até d).

O último passo na derivação da pdf de característica é incluir a restrição de van der Waals. Como todas as restrições estereoquímicas tem que ser

satisfeitas em todas as estruturas, estas restrições são multiplicadas na pdf de característica, e assim se obtêm a pdf de característica final:

$$p^D(d) = [\omega_1 p_1^d(d) + \omega_2 p_2^d(d)] p^v(d) \quad (8)$$

Esta abordagem simples para se combinar duas pdfs básicas é utilizada para qualquer número de pdfs básicas que foram derivadas de estruturas relacionadas.

Derivação da pdf molecular a partir das pdfs de características

O último estágio na derivação de uma pdf molecular é a combinação de todas as pdfs de características em uma única pdf. A estrutura tridimensional de uma proteína somente pode ser determinada se um número suficientemente grande de características espaciais f_i for especificado. O objetivo é encontrar a estrutura 3D que seja consistente com os valores mais prováveis para cada característica individual f_i . A pdf molecular deve fornecer a probabilidade de ocorrência de cada combinação de tais características simultaneamente. Então o modelo para a estrutura 3D da sequência a ser modelada deve corresponder ao máximo da pdf molecular. Assumindo que as pdfs de características são independentes uma das outras, a pdf molecular é simplesmente um produto das pdfs de características.

$$P = \prod_i p^F(f_i) \quad (9)$$

Sendo assim, maximizando a função P nós podemos encontrar o modelo mais provável para a estrutura 3D da sequência, dado seu alinhamento com as estruturas conhecidas.

Otimização da pdf molecular

O último passo para a obtenção do modelo de estrutura tridimensional é a maximização da função P que representa a pdf molecular obtida.

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

$$F = -\ln(P) \quad (10)$$

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 , entretanto F é computacionalmente mais indicada para ser otimizada que P 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 padrão 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 (Braun & Go, 1985). A principal diferença entre o método original e a atual implementação é que a otimização atual é feita no espaço Cartesiano e o procedimento original otimizava os ângulos diedrais.



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 sequência 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 sequência, 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 26).



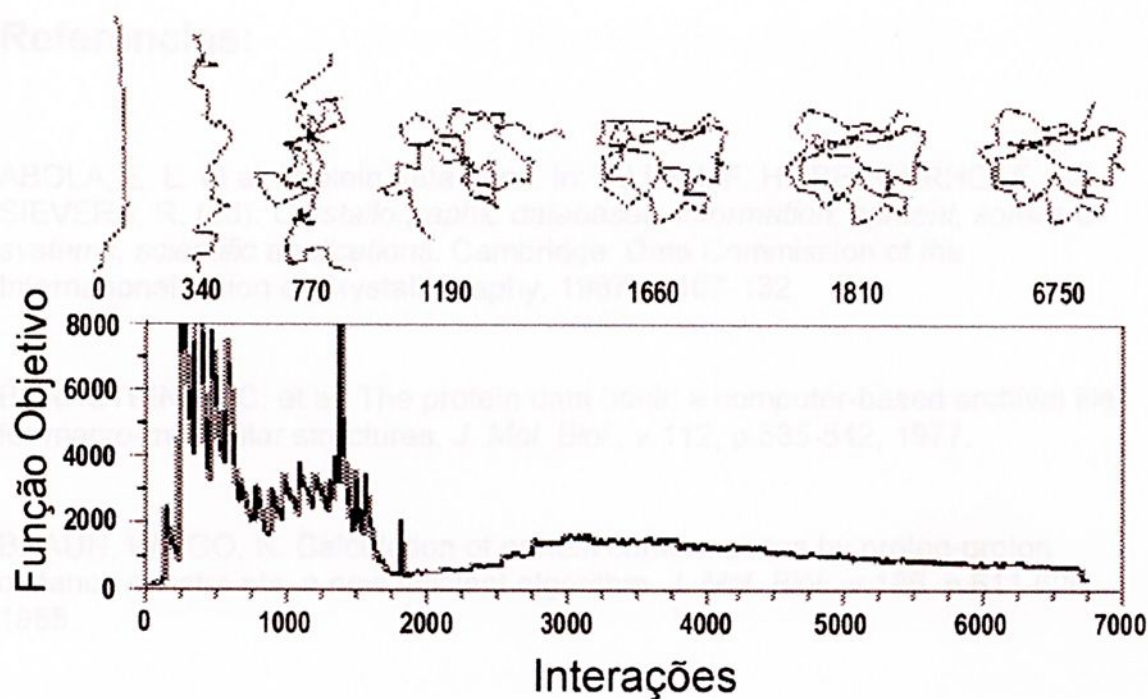


Figura 26: Mudança de conformação pela otimização da função objetivo

Referências:

ABOLA, E. E. et al. Protein data bank. In: ALLEN, F. H.; BERGERHOFF, G.; SIEVERS, R. (ed). *Crystallographic databases: information, content, software systems, scientific applications*. Cambridge: Data Commission of the International Union of Crystallography, 1987. p.107-132.

BERNSTEIN, F. C. et al. The protein data bank: a computer-based archival file for macro-molecular structures. *J. Mol. Biol.*, v.112, p.535-542, 1977.

BRAUN, W.; GO, N. Calculation of protein conformations by proton-proton distance constraints: a new efficient algorithm. *J. Mol. Biol.*, v.186, p.611-626, 1985.

COHEN, F. E.; KUNTZ, I. D. Tertiary structure prediction. In: FASMAN, G. D. (ed.) *Prediction of protein structure and the principles of protein conformation*. New York : Plenum, 1989. p.647-705.

HUBBARD, T. J. P.; BLUNDELL, T. L. Comparison of solvent inaccessible cores of homologous proteins: definitions useful for protein modelling. *Protein Eng.*, v.1, p.159-171, 1987.

LESK, A. M.; CHOTHIA, C. H. The response of protein structures to amino-acid sequence changes. *Philos. Trans. R. Soc. Lond.*, v.317, p.345-356, 1986.

TOPHAM, C. M. et al. An assessment of COMPOSER: a rule-based approach to modelling protein structure. *Biochem. Soc. Symp.*, v.57, p.1-9, 1991.

WILSON, C.; DONIACH, S. A computer model to dynamically simulate protein folding: studies with Crambim. *Proteins*, v.6, p.193-209, 1989.



Apêndice B – Cálculo da acessibilidade estática

Em 1971, Lee e Richards publicaram as bases do algoritmo utilizado para se calcular a superfície acessível ao solvente de moléculas de proteínas, os principais passos deste algoritmo são descritos abaixo.

O primeiro passo é a construção da superfície de van der Waals da molécula analisada, o que é feito atribuindo-se o raio de van der Waals a cada átomo ou grupo de átomos.

A estrutura é então representada por um conjunto de esferas, cada uma com o raio adequado e centralizadas em seus respectivos átomos. Esta estrutura é seccionada por um conjunto de planos paralelos com espaçamento pré-determinado. A interseção entre cada esfera e um determinado plano aparece na forma de um círculo e os arcos de círculos que se interceptam são então eliminados, desde que os átomos representados por tais círculos pertençam a uma mesma molécula (figura 27).

Desta forma, temos em cada plano o desenho do contorno da superfície de van der Waals da molécula.

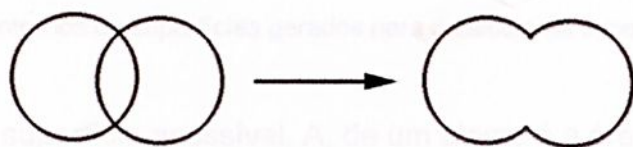


Figura 27: Contornos da superfície de van der Waals no plano

Um átomo ou grupo de átomos é definido como acessível se uma molécula de solvente, de um determinado tamanho, puder ser colocada em contato com a sua superfície de van der Waals.

Para o cálculo, uma esfera é criada e centralizada em cada coordenada atômica, com um raio igual à soma dos raios de van der Waals do átomo e da molécula de solvente. A superfície desta nova esfera indica os pontos por onde passa o centro de uma molécula de solvente enquanto rola sobre a superfície da proteína fazendo o máximo contato permitido (figura 28). Caso exista alguma parte desta nova superfície desenhada em algum dos planos, estando fora dos contornos da superfície de van der Waals previamente calculada, então, esse átomo é acessível. O tamanho do arco desenhado é a medida da acessibilidade em cada plano e a acessibilidade total de cada átomo será proporcional à soma dos comprimentos de todos os arcos gerados por ele.

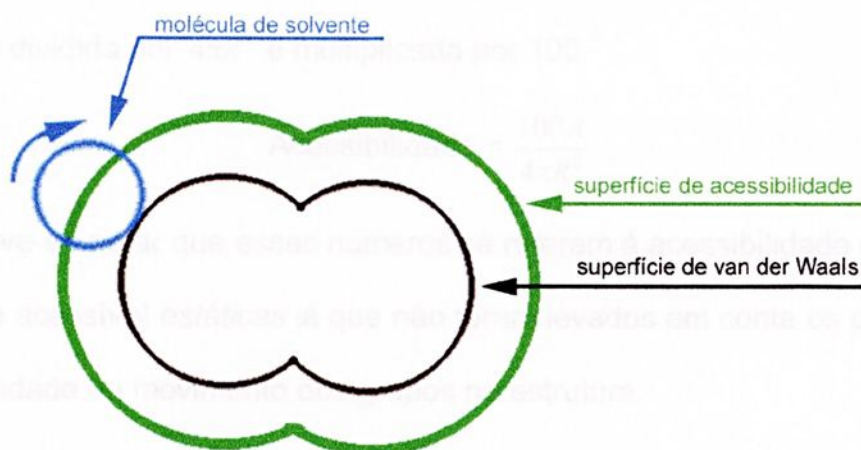


Figura 28: Contornos de superfícies gerados para o cálculo da superfície acessível

A área de superfície acessível, A , de um átomo é a área da superfície de uma esfera de raio R , em cada ponto onde o centro de uma molécula de solvente puder ser colocado em contato com essa esfera sem penetrar nenhuma outra esfera equivalente da molécula. O raio R é dado pela soma do raio de van der Waals do átomo e do raio escolhido para a molécula de

solvente. Uma aproximação desta área é calculada pelo programa usando a equação abaixo:

$$\text{Área da superfície acessível} = A = \sum \left(\frac{R}{\sqrt{R^2 - Z_i^2}} \right) \cdot D \cdot L_i$$

sendo $D = \frac{\Delta Z}{2 + \Delta'Z}$ onde L_i é o comprimento do arco desenhado em um dado plano i , Z_i é a distância perpendicular do centro da esfera até o plano i , ΔZ é o espaçamento entre os planos, e $\Delta'Z$ é $\frac{\Delta Z}{2}$ ou $R - Z_i$, o que for menor.

A somatória é sobre todos os arcos desenhados para um determinado átomo. A acessibilidade é definida simplesmente como a área de superfície acessível dividida por $4\pi R^2$ e multiplicada por 100.

$$\text{Acessibilidade} = \frac{100A}{4\pi R^2}$$

Deve-se notar que esses números se referem à acessibilidade e área de superfície acessível *estáticas* já que não foram levados em conta os potenciais de flexibilidade ou movimento dos grupos na estrutura.

Identificação de cavidades

O programa é capaz de identificar cavidades dentro da molécula de proteína que sejam grandes o suficiente para acomodar ao menos uma molécula de solvente. Isso é feito isolando-se todos os contornos de acessibilidade de curvatura côncava em cada plano e eliminando aqueles que tiverem um canal que leve do interior do contorno côncavo ao exterior da molécula.

O "volume" da cavidade é calculado multiplicando a área interna de cada contorno da cavidade pelo espaçamento entre os planos e considerando todos os planos onde aparecem os contornos desta cavidade.

Esse número representa o volume que pode ser ocupado pelo centro de uma molécula de solvente. Assim sendo, uma cavidade grande o suficiente para acomodar uma molécula de solvente, mas sem restar nenhum espaço para deslocamento da mesma, apresenta um volume igual a zero por essa definição.

Referências:

LEE, B.; RICHARDS, F. M. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.*, v.55, n.3, p.379-400, 1971.



Apêndice C – Estruturas de CDKs depositadas no Protein Data Bank

Código PDB: 1AQ1

Título: Human Cyclin Dependent Kinase 2 Complexed With The Inhibitor Staurosporine

Código PDB: 1B39

Título: Human Cyclin-Dependent Kinase 2 Phosphorylated On Thr 160

Código PDB: 1BI7

Título: Mechanism Of G1 Cyclin Dependent Kinase Inhibition From The Structure Of The Cdk6-P16Ink4A Tumor Suppressor Complex

Código PDB: 1BI8

Título: Mechanism Of G1 Cyclin Dependent Kinase Inhibition From The Structures Cdk6-P19Ink4D Inhibitor Complex

Código PDB: 1BLX

Título: P19Ink4D/Cdk6 Complex

Código PDB: 1BUH

Título: Crystal Structure Of The Human Cdk2 Kinase Complex With Cell Cycle-Regulatory Protein Cks1



Código PDB: 1CDK

Título: Camp-Dependent Protein Kinase Catalytic Subunit (E.C. 2.7.1.37) (Protein Kinase A) Complexed With Protein Kinase Inhibitor Peptide Fragment 5-24 (Pki(5-24) Isoelectric Variant Ca) and Mn^{2+} Adenylyl Imidodiphosphate (Mnamp-Pnp) At pH 5.6 and 7C and 4C

Código PDB: 1CKP

Título: Human Cyclin Dependent Kinase 2 Complexed With The Inhibitor Purvalanol B

Código PDB: 1DI8

Título: The Structure Of Cyclin-Dependent Kinase 2 (Cdk2) In Complex With 4-[3-Hydroxyanilino]-6,7-Dimethoxyquinazoline

Código PDB: 1DM2

Título: Human Cyclin-Dependent Kinase 2 Complexed With The Inhibitor Hymenialdisine

Código PDB: 1E1V

Título: Human Cyclin Dependent Kinase 2 Complexed With The Inhibitor 2058

Código PDB: 1E1X

Título: Human Cyclin Dependent Kinase 2 Complexed With The Inhibitor 6027

Código PDB: 1E9H

Título: Thr 160 Phosphorylated Cdk2 - Human Cyclin A3 Complex With The Inhibitor Indirubin-5-Sulphonate Bound

Código PDB: 1F5Q

Título: Crystal Structure Of Murine Herpesvirus Cyclin Complexed To Human Cyclin Dependent Kinase 2

Código PDB: 1G2

Código PDB: 1FIN

Título: Cyclin A - Cyclin-Dependent Kinase 2 Complex

Código PDB: 1GU

Código PDB: 1FQ1

Título: Crystal Structure Of Kinase Associated Phosphatase (Kap) In Complex With Phospho-Cdk2

Código PDB: 1FV

Código PDB: 1FVT

Título: The Structure Of Cyclin-Dependent Kinase 2 (Cdk2) In Complex With An Oxindole Inhibitor

Código PDB: 1FV

Código PDB: 1FVW

Título: The Structure Of Cdk2/Cyclin A In Complex With An Oxindole Inhibitor

Código PDB: 1G3N

Código PDB: 1G3N

Título: Structure Of A P18(Ink4C)-Cdk6-K-Cyclin Ternary Complex

Código PDB: 1G5S

Título: Crystal Structure Of Human Cyclin Dependent Kinase 2 (Cdk2) In Complex With The Inhibitor H717

Código PDB: 1GIH

Título: Human Cyclin Dependent Kinase 2 Complexed With The Cdk4 Inhibitor

The Inhibitor 0102

Código PDB: 1GII

Título: Human Cyclin Dependent Kinase 2 Complexed With The Cdk4 Inhibitor

Título: Cdk2/Cyclin A In Complex With A 1-Residue Acetylated Peptide From

Código PDB: 1GIJ

Título: Human Cyclin Dependent Kinase 2 Complexed With The Cdk4 Inhibitor

Código PDB: 1G25

Código PDB: 1GY3

Título: Pcdk2/Cyclin A In Complex With Mgadp, Nitrate and Peptide Substrate

Código PDB: 1H1P

Título: Structure Of Human Thr160-Phospho Cdk2/Cyclin A Complexed With The Inhibitor 2058

Código PDB: 1H1Q

Título: Structure Of Human Thr160-Phospho Cdk2/Cyclin A Complexed With The Inhibitor 6094



Código PDB: 1H1R

Título: Structure Of Human Thr160-Phospho Cdk2/Cyclin A Complexed With The Inhibitor 6086

Código PDB: 1H1S

Título: Structure Of Human Thr160-Phospho Cdk2/Cyclin A Complexed With The Inhibitor 6102

Código PDB: 1H24

Título: Cdk2/Cyclin A In Complex With A 9 Residue Recruitment Peptide From E2F

Título: Human Cyclin-Dependent Kinase 2

Código PDB: 1H25

Título: Cdk2/Cyclin A In Complex With An 11-Residue Recruitment Peptide From Retinoblastoma-Associated Protein

Código PDB: 1H26

Título: Cdk2/Cyclin A In Complex With An 11-Residue Recruitment Peptide From P53

Código PDB: 1H27

Título: Cdk2/Cyclin A In Complex With An 11-Residue Recruitment Peptide From P27

Título: The Structure Of Cyclin-Dependent Kinase 2 (Cdk2) In Complex With A [(S)-Ammonio-4-Pyridinyl]-Acetophenonesulfonamide

Código PDB: 1H28

Título: Cdk2/Cyclin A In Complex With An 11-Residue Recruitment Peptide
From P107

Código PDB: 1H4L

Título: Title Structure and Regulation Of The Cdk5-P25(Nck5A) Complex

Código PDB: 1HCK

Título: Human Cyclin-Dependent Kinase 2

Código PDB: 1HCL

Título: Human Cyclin-Dependent Kinase 2

Código PDB: 1JOW

Título: Crystal Structure Of A Complex Of Human Cdk6 and A Viral Cyclin

Código PDB: 1JST

Título: Phosphorylated Cyclin-Dependent Kinase-2 Bound To Cyclin A

Código PDB: 1JSU

Título: P27(Kip1)/Cyclin A/Cdk2 Complex

Código PDB: 1JSV

Título: The Structure Of Cyclin-Dependent Kinase 2 (Cdk2) In Complex With 4-
[(6-Amino-4-Pyrimidinyl) Amino]Benzenesulfonamide



Código PDB: 1JVP

Título: Crystal Structure Of Human Cdk2 (Unphosphorylated) In Complex With Pkf049-365

Código PDB: 1KE5

Título: Cdk2 Complexed With N-Methyl-4-[(2-Oxo-1,2-Dihydro-3H-Indol-3-Ylidene)Methyl]Amino}Benzenesulfonamide

Código PDB: 1KE6

Título: Cyclin-Dependent Kinase 2 (Cdk2) Complexed With N-Methyl-{4- [2-(7-Oxo-6,7-Dihydro-8H-[1,3]Thiazolo[5,4-E]Indol-8-Ylidene)Hydrazino]Phenyl}Methanesulfonamide

Código PDB: 1KE7

Título: Cyclin-Dependent Kinase 2 (Cdk2) Complexed With 3-[(2,2-Dioxido-1,3-Dihydro-2-Benzothien-5-Yl)Amino]Methylene}-5-(1,3-Oxazol-5-Yl)-1,3-Dihydro-2H-Indol-2-One

Código PDB: 1KE8

Título: Cyclin-Dependent Kinase 2 (Cdk2) Complexed With 4-[(2-Oxo-1,2-Dihydro-3H-Indol-3-Ylidene)Methyl]Amino}-N-(1,3-Thiazol-2-Yl)Benzenesulfonamide

Código PDB: 1KE9

Título: Cyclin-Dependent Kinase 2 (Cdk2) Complexed With 3-{[4-
({[Amino(Imino)Methyl]Aminosulfonyl)Anilino]Methylene}-2- Oxo-2,3-Dihydro-
1H-Indole

Código PDB: 1QMZ W. F. et al. Molecular crystal of cyclin-dependent kinase 2

Título: Phosphorylated Cdk2-Cyclin A-Substrate Peptide Complex

1154-1158, 2002





Molecular model of cyclin-dependent kinase 5 complexed with roscovitine

Walter Filgueira de Azevedo Jr.,^{a,b,*} Renato Tadeu Gaspar,^a Fernanda Canduri,^{a,b}
João Carlos Camera Jr.,^a and Nelson José Freitas da Silveira^a

^a Departamento de Física, IBILCE, UNESP São José do Rio Preto, SP 15054-000, Brazil

^b Center for Applied Toxinology, Instituto Butantan, Av. Vital Brasil 1500, São Paulo, SP 05503-900, Brazil

Received 19 August 2002

Abstract

Here is described a structural model for the binary complex CDK5–roscovitine. Roscovitine has been shown to potently inhibit cyclin-dependent kinases 1, 2 and 5 (CDK1, 2, and 5), and the structure of CDK2 complexed with roscovitine has been reported; however, no structural data are available for complexes of CDK5 with inhibitors. The structural model indicates that roscovitine strongly binds to the ATP-binding pocket of CDK5 and structural comparison of the CDK2–roscovitine complex correlates the structural differences with differences in inhibition of these CDKs by this inhibitor. This structure opens the possibility of testing new inhibitor families, in addition to new substituents for the already known lead structures of adenine derivatives.

© 2002 Elsevier Science (USA). All rights reserved.

Keywords: CDK; Roscovitine; Bioinformatics; Structure; Drug design

Cyclin-dependent kinases (CDKs) control the progression of cell cycle [1,2]. CDKs are constituted of a catalytic subunit and a regulatory subunit (cyclin) [3]. Each step of the cell cycle is thought to be regulated by such CDK–cyclin complexes. However, certain CDKs do not participate in controlling cell cycle, but they are rather involved in controlling cell differentiation in neuronal and muscle cells [4,5]. In addition, CDK also play a role in apoptosis (CDK2) and in the control of transcription (CDKs7, 8, 9) [6,7].

CDK5 is a unique member of the CDK family of kinases. Although the CDK5 is widely expressed in many tissues and cells, CDK5 kinase activity is restricted to neuronal cells [8]. This specificity for neuronal tissue is the result of the CDK5 activator proteins p35, p25, and p39. CDK5 is a multifunctional kinase that associates with other cell proteins to interact with the cytoskeleton and form supramolecular complexes. Recent investigations have revealed that most of the CDK5 in cells forms large multimeric

complexes of high molecular weight, ranging from 60 to 670 kDa, in which CDK5 is associated with p25, p35, synapsin, τ -, β -catenins, and *N*-cadherins [4]. Increased CDK5 kinase activity has been implicated in Alzheimer's disease. Furthermore, pretreatment of cells with CDK5 inhibitors protected cells against neuronal death [9]. Since deregulation of CDK5 has been implicated in Alzheimer's disease, there is strong interest in chemical inhibitors of CDK5 that could play an important role in the discovery of anti-Alzheimer's disease agents.

Roscovitine [2-(*R*)-(1-ethyl-2-hydroxy-ethylamino)-6-benzylamino-9-isopropylpurine] is a purine analog that has been shown to potently inhibit CDK, with IC₅₀ values ranging from 0.16 μ M (for CDK5) to over 100 μ M (for CDK4 and CDK6) [10]. The crystallographic structure of the binary complex of CDK2–roscovitine indicates that the inhibitor is strongly bound in the ATP-binding pocket of CDK2. This article describes the modeling of human CDK5 complexed with roscovitine. The investigation was made in order to gain further insight into the structural basis for chemical inhibition of CDK5.

* Corresponding author. Fax: +55-17-221-2247.

E-mail address: walterfa@df.ibilce.unesp.br (W.F. de Azevedo Jr.).



Methods

Molecular modeling. The structure of CDK5 complexed with p25 has been recently determined [11] and a molecular model for CDK5 complexed with the peptide PKTPKKAKKL has also been reported [12]; however, the atomic coordinates for CDK5 are not available. Therefore, for modeling of the CDK5 we used a protocol previously described for modeling of CDK9 [13]. In brief, model building of CDK5 was carried out using the program MODELLER [14]. The atomic coordinates of eight crystallographic CDK2 structures solved to resolution better than 2.1 Å were used to build up an ensemble of CDK2 structures to be used as starting models for modeling of the CDK5. The alignment of CDK2 (template) and CDK5 (target) is shown in Fig. 1. Next, the spatial restraints and CHARMM energy terms enforcing proper stereochemistry [15] were combined into an objective function. Finally, the model is obtained by optimizing the objective function in Cartesian space. The optimization is carried out by the use of the variable target function method [16] employing methods of conjugate gradients and molecular dynamics with simulated annealing. Several slightly different models can be calculated by varying the initial structure. We generated 500 CDK5 models and the final model was selected based on stereochemical quality. The roscovitine molecule from the CDK2–roscovitine structure was superposed onto the best CDK5 model. Further optimization of the complex CDK5–roscovitine was performed following the procedure described for CDK5 model. All optimization process was performed on SGI Octane, R12000.

Analysis of the model. The overall stereochemical quality of the final model for CDK5 was assessed by the program PROCHECK [17]. Atomic models were superposed using the program LSQKAB from CCP4 [18]. The cutoff for hydrogen bonds and salt bridges was 3.6 Å. The contact surfaces for the binary complexes were calculated using AREAIMOL and RESAREA [18].

Results and discussion

Quality of the model

Fig. 2 shows the Ramachandran diagram ϕ - ψ plots for the binary complex of CDK5–roscovitine. The Ramachandran plot for the eight CDK2 structures was also generated (figure not shown) to compare the overall stereochemical quality of CDK5 model against CDK2 structures solved by biocrystallography. Analysis of the Ramachandran plot of the best CDK5 model shows that 91.0% of the residues lies in the most favorable regions, 8.2% in the additional allowed regions, and 0.8% in the disallowed regions (Asp97 and Ala198). The same analysis for eight crystallographic CDK2 structures

presents 89.8% of residues in the most favorable, 9.7% in additional allowed regions, and 0.5% in generously allowed regions. The overall rating for the model is better than the one obtained for the eight structures of CDK2.

Overall description

The model of the kinase in the complex CDK5–roscovitine is folded into the typical bilobal structure as was observed for CDK2 structure [19,20], with the smaller N-terminal lobe consisting predominantly of β -sheet structure and the larger C-terminal lobe consisting primarily of α -helices. Fig. 3 shows a schematic drawing of the complex CDK5–roscovitine. The N-terminal lobe of CDK5 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). The roscovitine molecule is found in the cleft between the two lobes.

Molecular fork

Analysis of the many available crystal structures of CDK2–inhibitor complexes reveals the existence of conserved interactions that appear to be determinant for recognition of the small molecules by the ATP-binding pocket of this class of enzymes. The participation of a molecular fork composed of C=O group on Glu81, N–H group, and C=O group in Leu83 in hydrogen bonds between CDK2 and inhibitor has been observed on several CDK2–inhibitor structures. Furthermore, the molecular model for CDK9 complexed with flavopiridol also presented the molecular fork [13]. In the CDK5–roscovitine model this molecular fork is composed of C=O on Glu81, N–H, and C=O on Cys83 and participates in hydrogen bonds between CDK5 and roscovitine.

Side chain positions

To find out whether roscovitine binding to CDK5 induces changes in side chain conformation in the binding pocket, we compared the binding pockets of CDK5 complexed with roscovitine and CDK2 complexed with ATP (PDB access code: 1HCK) after a

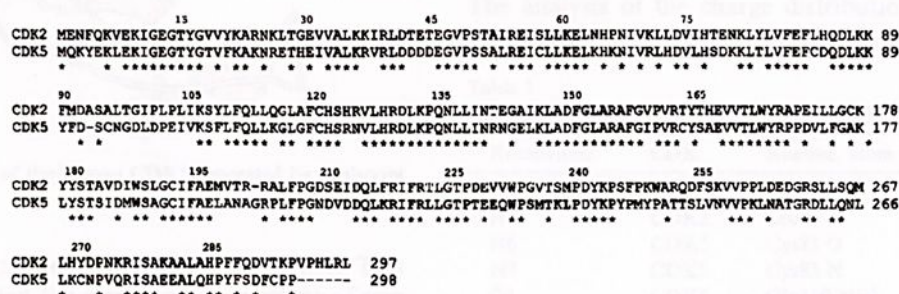


Fig. 1. Sequence alignment of human CDK2 and CDK5. There is 40.4% identity between CDK2 and CDK5 sequences in the modelled region. The alignment was performed with the program CLUSTAL V [25].

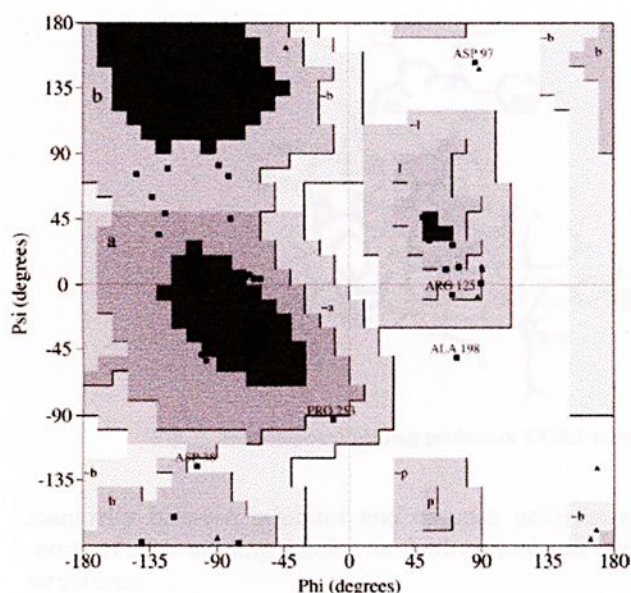


Fig. 2. Ramachandran plot for the CDK5. The regions A, B, and L are more favorable, the regions a, b, l, and p are allowed and ~a, ~b, ~l, and ~p are generously allowed regions. Glycine residues are shown as triangles.

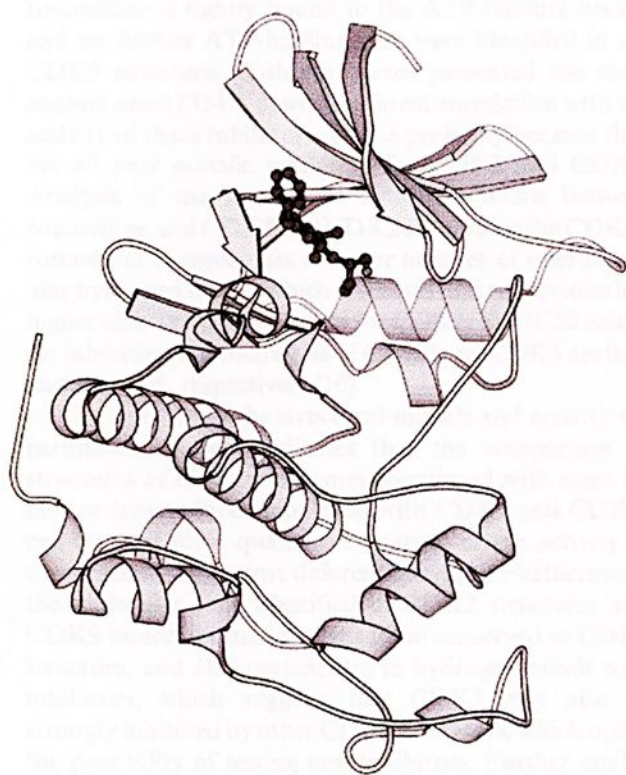


Fig. 3. Ribbon diagram of the human CDK5 generated by Molscript [26].

superposition of the models based on their C α atoms. This superposition revealed the significant movement of two side chains. For the roscovitine complex, it was observed that Lys20 changed its position, with NZ moving by

1.97 Å away from its position in the ATP complex and Phe82 changed its position, with CZ moving by 2.91 Å away from its position in the ATP complex. No significant movements were observed on the side chains in the ATP-binding pocket of the complex CDK5–roscovitine. In summary, roscovitine fits very well to the binding pocket with small adjustments in the side chains.

Interactions of roscovitine with CDK5

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 [21–23]. A total of five hydrogen bonds were observed between CDK5 and roscovitine, in binary model, involving the residues Cys83, Asp86, and Gln130. For the CDK2–roscovitine model two hydrogen bonds involving the residue Leu83 were observed. Table 1 shows the intermolecular hydrogen bonds for both complexes. The higher number of intermolecular hydrogen bonds, observed in the CDK5–roscovitine complex, is probably due to the modification of His84 (CDK2) to Asp in the CDK5 sequence, which allows an additional salt bridge in the CDK5 structure, involving Asp84 and Lys20. This salt bridge moves the inhibitor away from the molecular fork of CDK5, however, keeping the intermolecular hydrogen bonds between the molecular fork and roscovitine. Nevertheless, this movement is enough to bring roscovitine closer to the side chains of Gln130 and Asp86, which allows additional intermolecular hydrogen bonds.

Superposition of the CDK2–ATP onto CDK5–roscovitine structure indicates that the two ring systems of roscovitine and ATP overlap approximately in the same plane, however, with different orientations. As observed for the crystallographic structure of roscovitine bound to CDK2 [3], the region of CDK5 occupied by the phenyl ring of roscovitine is pointing away from the ATP-binding pocket and partially exposed to solvent. Fig. 4 shows the ATP-binding pocket for the complexes CDK2–ATP and CDK5–roscovitine.

The electrostatic potential surface of the CDK2–roscovitine and the model of CDK5 complexed with the same inhibitor were calculated with GRASP [24]. The analysis of the charge distribution of the binding pockets indicates the presence of some charge comple-

Table 1
Intermolecular hydrogen bonds between CDK5 and roscovitine

Roscovitine	CDK	Residue, atom	Distance (Å)
N6	CDK2	Leu83 O	2.82
N7	CDK2	Leu83 N	3.38
N6	CDK5	Cys83 O	3.10
N7	CDK5	Cys83 N	3.53
O1	CDK5	Gln130 NE2	3.29
N2	CDK5	Asp86 OD2	3.54
N1	CDK5	Asp86 OD2	3.16

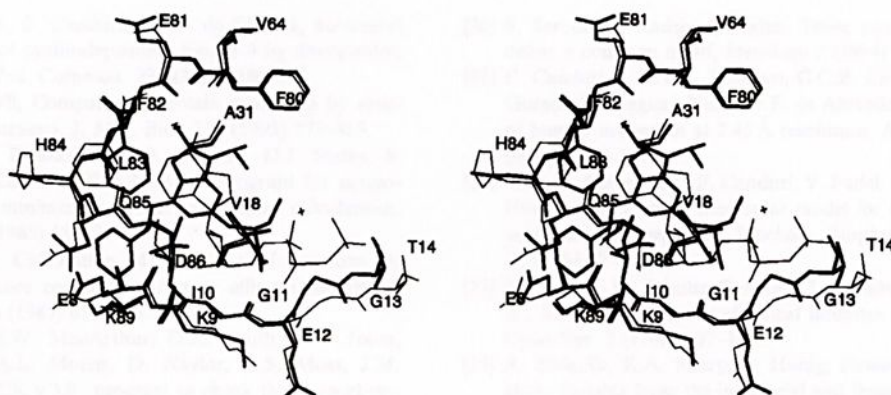


Fig. 4. Superimposed binding pockets of CDK5-roscovitine complex (thick line) and CDK2-ATP complex (thin line).

mentarity between inhibitor and enzyme; nevertheless, most of the binding pocket is hydrophobic in both structures.

Conclusions

The overall structure of the complex indicates that the roscovitine is tightly bound to the ATP-binding pocket and no further ATP-binding sites were identified in the CDK5 structure. Both complexes presented the same contact areas ($334 \text{ \AA}^2$), with no direct correlation with the activity of these inhibitors. This is probably because they are all very specific inhibitors for CDK2 and CDK5. Analysis of the number of hydrogen bonds between roscovitine and CDK5 and CDK2 shows that the CDK5-roscovitine complex has a higher number of intermolecular hydrogen bonds, which indicates that roscovitine has higher affinity for CDK5 than for CDK2. The IC_{50} values for inhibition by roscovitine of CDK2 and CDK5 are 0.70 and 0.16 μM , respectively [10].

The analysis of the structural models and activity experiments strongly indicates that the comparison of structures of different enzymes complexed with same inhibitor (roscovitine complexed with CDK5 and CDK2) can be used for a qualitative analysis of the activity of these inhibitors against different enzymes. Furthermore, the molecular fork identified in CDK2 structures and CDK9 molecular model seems to be conserved in CDK5 structure, and also participates in hydrogen bonds with inhibitors, which suggests that CDK5 may also be strongly inhibited by other CDK2 inhibitors, which opens the possibility of testing new inhibitors. Further inhibition experiments may confirm this prediction.

Acknowledgments

We thank Andressa Salb  dos Santos Oliveira for English revision. This work was supported by grants from FAPESP (SMOLBNet), CNPq, CAPES, and Fundo Bunka de Pesquisa (Banco Sumitomo).

W.F.A. Jr. is a researcher for the Brazilian Council for Scientific and Technological Development (CNPq, 300851/98-7).

References

- [1] C. Norbury, P. Nurse, Animal cell cycles and their control, *Annu. Rev. Biochem.* 61 (1992) 441–470.
- [2] W.F. de Azevedo Jr., H.J. Mueller-Dieckmann, U. Schulze-Gahmen, P.J. Worland, E. Sausville, S.-H. Kim, Structural basis for specificity and potency of a flavonoid inhibitor of human CDK2, a cell cycle kinase, *Proc. Natl. Acad. Sci. USA* 93 (7) (1996) 2735–2740.
- [3] W.F. de Azevedo Jr., S. Leclerc, L. Meijer, L. Havlicek, M. Strnad, S.-H. Kim, Inhibition of cyclin-dependent kinases by purine analogues: crystal structure of human CDK2 complexed with roscovitine, *Eur. J. Biochem.* 243 (1997) 518–526.
- [4] G.N. Patrick, L. Zuckerberg, M. Nikolic, S. de la Monte, P. Dikkes, L.H. Tsai, Conversion of p35 to p25 deregulates Cdk5 activity and promotes neurodegeneration, *Nature* 402 (6762) (1999) 615–622.
- [5] R.B. Maccioni, C. Otth, I.I. Concha, J.P. Mu oz, The protein kinase Cdk5. Structural aspects, roles in neurogenesis and involvement in Alzheimer's pathology, *Eur. J. Biochem.* 268 (6) (2001) 1518–1527.
- [6] S.H. Chao, K. Fujinaga, J.E. Marion, R. Taube, E.A. Sausville, A.M. Senderowicz, B.M. Peterlin, D.H.J. Price, Flavopiridol inhibits P-TEFb and blocks HIV-1 replication, *J. Biol. Chem.* 275 (37) (2000) 28345–28348.
- [7] N. Gray, L. Detivaud, C. Doerig, L.C. Meijer, ATP-site directed inhibitors of cyclin-dependent kinases, *Curr. Med. Chem.* 6 (9) (1999) 859–875.
- [8] P.M. Clare, R.A. Poorman, L.C. Kelley, K.D. Watenpaugh, C.A. Bannow, K.L. Leach, The cyclin-dependent kinases CDK2 and CDK5 act by a random, anticooperative kinetic mechanism, *J. Biol. Chem.* 276 (51) (2001) 48292–48299.
- [9] A. Alvarez, R. Toro, A. C rceres, R.B. Maccioni, Inhibition of tau phosphorylating protein kinase CDK5 prevents β -amyloid-induced neuronal death, *FEBS Lett.* 459 (3) (1999) 421–426.
- [10] L. Meijer, A. Borgne, O. Mulner, J.P. Chong, J.J. Blow, N. Inagaki, Biochemical and cellular effects of roscovitine, a patent and selective inhibitor of the cyclin-dependent kinases CDK2 and CDK5, *Eur. J. Biochem.* 243 (1997) 527–536.
- [11] C. Tarricone, R. Dhavan, J. Peng, L.B. Areces, L.H. Tsai, A. Musacchio, Structure and regulation of the CDK5-p25(nck5a) complex, *Mol. Cell* 8 (3) (2001) 657–669.
- [12] P. Sharma, P.J. Steinbach, M. Sharma, N.D. Amin, J.J. Barchi Jr., H.C. Pant, Identification of substrate binding site of cyclin-dependent kinase 5, *J. Biol. Chem.* 274 (1999) 9600–9606.

- [13] W.F. de Azevedo Jr., F. Canduri, N.J.F. da Silveira, Structural basis for inhibition of cyclin-dependent kinase 9 by flavopiridol, *Biochem. Biophys. Res. Commun.* 293 (2002) 566–571.
- [14] A. Sali, T.L. Blundell, Comparative protein modelling by satisfaction of spatial restraints, *J. Mol. Biol.* 234 (1993) 779–815.
- [15] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus, CHARMM: a program for macromolecular energy minimization and dynamics calculations, *J. Comp. Chem.* 4 (1983) 187–217.
- [16] W. Braun, N. Go, Calculation of protein conformations by proton–proton distance constraints. A new efficient algorithm, *J. Mol. Biol.* 186 (3) (1985) 611–626.
- [17] R.A. Laskowski, M.W. MacArthur, D.K. Smith, D.T. Jones, E.G. Hutchinson, A.L. Morris, D. Naylor, D.S. Moss, J.M. Thornton, PROCHECK v.3.0 program to check the stereochemistry quality of protein structures operating instructions, 1994.
- [18] Collaborative Computational Project No. 4. The CCP4 Suite: programs for protein crystallography, *Acta Crystallogr. D* 50 (1994) 760–763.
- [19] H.L. DeBonds, J. Rosenblatt, J. Jancarik, H.D. Jones, D.O. Morgan, S.-H. Kim, Crystal structural of cyclin-dependent kinase 2, *Nature (Lond.)* 363 (1993) 595–602.
- [20] S. Serota, E. Radsio-Andzelm, Three protein kinase structures define a common motif, *Structure* 2 (1994) 345–355.
- [21] F. Canduri, L.G.V.L. Teodoro, C.C.B. Lorenzi, V. Hial, R.A.S. Gomes, J. Ruggiero Neto, W.F. de Azevedo Jr., Crystal structure of human uropepsin at 2.45 Å resolution, *Acta Crystallogr. D* 57 (2001) 1560–1570.
- [22] W.F. de Azevedo Jr., F. Canduri, V. Fadel, L.G.V.L. Teodoro, V. Hial, R.A.S. Gomes, Molecular model for the binary complex of uropepsin and pepstatin, *Biochem. Biophys. Res. Commun.* 287 (1) (2001) 277–281.
- [23] S.-H. Kim, U. Schulze-Gahmen, J. Brandsen, W.F. de Azevedo Jr., Structural basis for chemical inhibitor of CDK2, *Prog. Cell Cycle Res.* 2 (1996) 137–145.
- [24] A. Nicholls, K.A. Sharp, B. Honig, Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons, *Proteins* 11 (4) (1991) 281–296.
- [25] D.G. Higgins, A.J. Bleasby, R. Fuchs, CLUSTAL V: Improved software for multiple sequence alignment, *Comput. Appl. Biosci.* 8 (2) (1992) 189–191.
- [26] P. Kraulis, MOLSCRIPT: A program to produce both detailed and schematic plots of proteins, *J. Appl. Cryst.* 24 (1991) 946–950.

