

Ana Helena Casé

**Modelagem Molecular da Quinase Dependente de Ciclina 1
Complexada com Inibidores**



1910001429



Ana Helena Casé

MODELAGEM MOLECULAR DA QUINASE DEPENDENTE DE CICLINA 1 COMPLEXADA COM INIBIDORES

T. 1423/05



Dissertação apresentada como parte das exigências para obtenção do título de Mestre em Biofísica Molecular, área de concentração Biofísica Molecular do Departamento de Física do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista "Julio de Mesquita Filho" – UNESP.

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

São Jose do Rio Preto – SP
2005

+577.3
C337m

20901003

07/10/2005

Doação

R\$ 23,00

Autor(a)

UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de São José do Rio Preto

Ana Helena Casé

DEPENDENTE DE CICLINA I COMPLEXADA COM
MODELAGEM MOLECULAR DA QUINASE

Casé, Ana Helena.

Modelagem molecular da quinase dependente de ciclina I
complexada com inibidores / Ana Helena Casé – São José do Rio Preto :
[s.n.], 2005

88 f. : il. ; 30 cm.

Orientador: Walter Filgueira de Azevedo Júnior

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

1. Biofísica molecular. 2. Inibidores químicos. 3. Ciclo celular. 4.
Bioinformática. 5. Modelagem molecular. 6. Quinase - Inibidores. 7.
Ciclina. I. Azevedo Júnior, Walter Filgueira de. II. Universidade
Estadual Paulista. Instituto de Biociências, Letras e Ciências Exatas. III.
Título.

CDU – 577.32



AGRADECIMENTOS

Ao Prof. Dr. Walter Filgueira de Azevedo Junior, pela orientação, e pelo crescimento profissional e científico proporcionado durante a realização desse trabalho e por ter me recebido em seu grupo e depositado em mim um voto de confiança.

Ao Prof. Dr. Marcelo Andrés Fossey, pela amizade e compreensão durante minha busca por carimbo.

Ao Prof. Dr. José Ramon Beltrán Abrego, pela amizade e dedicação, colaboração e incentivo.

Ao Prof. Dr. Elói da Silva Fencosa, pela atenção, receptividade e incentivo.

Ao Prof. Dr. Jorge Chahine, pela receptividade quando procurei o Departamento de Física.

Ao Prof. José Euripedes, pelo incentivo e por ter colaborado com muita incessante busca pelos meus ideais.

Aos Demais Professores do Departamento de Física pela atenção, companheirismo e incentivo.

A TODOS os colegas e amigos do Laboratório de Sistemas Biomoleculares.

A TODOS os colegas e amigos do Departamento de Física.

A Minha Amiga Solange, pelo Companheirismo, amizade e toda dedicação. Pelos vários momentos que passamos juntas durante esses anos.

A Minha Amiga Irene, pela amizade que será eterna e por me ouvir e me orientar sempre que necessário.

Aos Demais colegas do Departamento de Física pela boa convivência e colaboração sempre os procurei.

Aos Funcionários de Departamento de Física pelo companheirismo que sempre nos recebe.

Aos Funcionários da Seção de Pós-Graduação pela atenção e eficiência no atendimento.

A Minha irmã Sandra, por todos esses anos de amor e compreensão e também pela paciência e incentivo.

A Minha Sobrinha Ana Paula, Pelo amor e dedicação e por sempre me proporcionar.

Ao Meu filho, *Vinicius* que é responsável por toda vontade que tenho de vencer.

AGRADECIMENTOS

Ao Prof. Dr. Walter Filgueira de Azevedo Júnior, pela orientação, e pelo crescimento profissional e científico proporcionado durante a realização desse trabalho e por ter me recebido em seu grupo e depositado em mim um voto de confiança.

Ao Prof. Dr. Marcelo Andrés Fossey, pela amizade e compreensão durante minha busca por caminhos.

Ao Prof. Dr. José Ramon Beltran Abrego, pela amizade e dedicação, colaboração e incentivo.

Ao Prof. Dr. Eloi da Silva Feitosa, pela atenção, receptividade e incentivo.

Ao Prof. Dr. Jorge Chahine, pela receptividade quando procurei o Departamento de Física.

Ao Prof. José Eurípedes, pelo incentivo e por ter colaborado com minha incessante busca pelos meus ideais.

Aos Demais Professores do Departamento de Física pela atenção, companheirismo e incentivo.

A TODOS os colegas e amigos do Laboratório de Sistemas Biomoleculares.

A TODOS os colegas e amigos do Departamento de Física.

À Minha Amiga Solange, pelo Companheirismo, amizade e toda dedicação. Pelos vários momentos que passamos juntas durante esses anos.

À Minha Amiga Irene, pela amizade que será eterna e por me ouvir e me orientar sempre que necessário.

Aos Demais colegas do Departamento de Física pela boa convivência e colaboração sempre os procurei.

Aos Funcionários do Departamento de Física pelo companheirismo que sempre nos recebe.

Aos Funcionários da Seção de Pós-Graduação pela atenção e eficiência no atendimento.

À Minha Irmã Sandra, por todos esses anos de amor e compreensão e também pela paciência e incentivo.

À Minha Sobrinha Ana Paula, Pelo amor e dedicação e por todos momentos de alegrias que proporciona.

Aos Meus Pais e Irmãos, pela a estrutura e pelo incentivo para que eu pudesse realizar meus sonhos.

À todos que direta ou indiretamente contribuíram para o andamento deste trabalho.

Ao meu grande amigo André Martins por todos esses anos de amizade e por me ouvir e me apoiar todas as vezes que precisei.

Ao Meu Filho Vinícius por compreender a ausência e a distância, por estar sempre comigo sendo companheiro, compartilhando toda alegria refletida em seu olhar que motiva vencer.

À CAPES pelo auxílio financeiro para que eu pudesse desenvolver minhas pesquisas.

E finalmente, à sua atenção, que esta lendo essa dissertação.

1.2 Quinase dependente de Ciclina 1 (CDK1)	12
1.3 Quinase dependente de Ciclina 2 (CDK2)	14
1.4 Inibidores	15
1.5 MODELLER	18
1.5.1 Identificação e seleção de proteínas modelo	19
1.5.2 Alinhamento das sequências de resíduos de aminoácidos	20
1.5.2.1 Alinhamento global	21
1.5.2.2 Alinhamento local	21
1.5.2.3 Alinhamento múltiplo	21
1.5.3 Construção das coordenadas do modelo	22
1.5.4 Modelagem das regiões de alça	23
1.5.5 Modelagem das cadeias laterais	24
1.5.6 Validação e otimização do modelo	25
CAPÍTULO 2	27
OBJETIVOS	27
CAPÍTULO 3	28
MÉTODOS	28
3.1 Modelagem dos complexos	28
3.2 Análises dos modelos	29
3.3 Análises das interações do inibidor com a proteína	34
CAPÍTULO 4	36
RESULTADOS E DISCUSSÕES	36

Índice Geral

INDÍCE Geral.....	i
INDÍCE FIGURAS.....	iii
INDÍCE TABELAS.....	vi
INDÍCE ABREVIACÕES.....	vii
GLOSSÁRIO DE TERMOS UTILIZADOS E ENDEREÇOS DA INTERNET.....	viii
RESUMO.....	1
ABSTRACT.....	3
CAPÍTULO 1.....	5
INTRODUÇÃO.....	5
1.1 Ciclo Celular e CDK.....	7
1.2 Quinase dependente de Ciclina 1 (CDK1).....	12
1.3 Quinase dependente de Ciclina 2 (CDK2).....	14
1.4 Inibidores.....	15
1.5 MODELLER.....	18
1.5.1 Identificação e seleção de proteínas molde.....	19
1.5.2 Alinhamento das seqüências de resíduos de aminoácidos.....	20
1.5.2.1 Alinhamento global.....	21
1.5.2.2 Alinhamento local.....	21
1.5.2.3 Alinhamento múltiplo.....	21
1.5.3 Construção das coordenadas do modelo.....	22
1.5.4 Modelagem das regiões de alças.....	23
1.5.5 Modelagem das cadeias laterais.....	24
1.5.6 Validação e otimização do modelo.....	25
CAPÍTULO 2.....	27
OBJETIVOS.....	27
CAPÍTULO 3.....	28
MÉTODOS.....	28
3.1 Modelagem dos complexos.....	28
3.2 Análises dos modelos.....	29
3.3 Análises das interações do inibidor com a proteína.....	34
CAPÍTULO 4.....	36
RESULTADOS E DISCUSSÕES.....	36



4.1 Descrição Estrutural.....	41
4.2 Interações	43
4.3 Análises da Qualidade do Modelo.....	57
CAPÍTULO 5	62
CONCLUSÃO	62
CAPÍTULO 6	63
REFERÊNCIAS BIBLIOGRÁFICAS	63
APÊNDICE A – MODELLER	71
Derivação Restrições Espaciais.....	72
Derivação de pdfs básicas	73
Banco de dados local	74
Composição do Banco de dados local	75
Derivação das pdfs de características a partir de pdfs básicas	76
Derivação da pdf molecular a partir das pdfs características.....	78
Otimização da pdf molecular	79
APÊNDICE B – PARMODEL	81
Modelagem.....	81
Avaliação	84
Otimização	86
Visualização	87

Figura 11 - Nos diagramas de Ramachandran as regiões vermelhas, amarelas, cinzas claro e brancas representam respectivamente as regiões permitida, adicionalmente permitida, geralmente permitida e não permitida (a) complexo CDK1/Isopenteniladenina (b) complexo CDK2/Isopenteniladenina (c) complexo CDK1/Oleocutina (d) complexo CDK2/Oleocutina (e) complexo CDK1/Oxidol-16 e (f) complexo CDK2/Oxidol-16

Figura 12 - Nos diagramas de Ramachandran as regiões vermelhas, amarelas, cinzas claro e brancas representam respectivamente as regiões permitida, adicionalmente permitida, geralmente permitida e não permitida (a) complexo CDK1/Purvalana (b) complexo CDK2/Purvalana (c) complexo CDK1/Pocovina (d) complexo CDK2/Pocovina (e) complexo CDK1/Estauripona e (f) complexo CDK2/Estauripona

Figura 13 - Alinhamento das sequências de CDK1 e CDK2. Há 67,34% de identidade entre as sequências. Os resíduos idênticos ou conservados são representados em vermelho, já os resíduos que não apresentaram nenhuma identidade estão em azul e preto.

Figura 14 - Elementos de estrutura secundária de CDK1.

Índice de Figuras

Figura 1 - Fases do ciclo celular e duplicação cromossômica.....	8
Figura 2 - Proteínas que regulam as fases do ciclo celular, positiva (ciclina) e/ou negativamente (CKIs) que são representados em vermelho.....	9
Figura 3 - Mecanismos de inibição dos CKIs durante o ciclo celular, inibição realizada pela subclasse INK que se liga a CDK impedido a formação do complexo CDK/ciclina ou pela KIP que se liga a complexo já formado impedindo que a CDK fosforile outros substratos.....	10
Figura 4 - Atividade da CDK1 durante o ciclo celular em uma célula animal.....	13
Figura 5 - Estrutura secundária da CDK2.....	14
Figura 6 - Fórmula estrutural de todos inibidores estudados.....	17
Figura 7 - Esquema geral dos principais passos da modelagem por homologia.....	19
Figura 8 - Definição dos ângulos de torção da cadeia principal (Φ e Ψ) e das cadeias laterais (χ_n).....	24
Figura 9 - Ângulos de torção da cadeia polipeptídica, dois grupos peptídicos planares são mostrados. Os únicos movimentos razoavelmente livres são a rotação em torno da ligação C_α -N (ϕ) e da ligação C- C_α (φ). Por convenção, ϕ e φ têm 180° na conformação apresentada e aumenta, como indicado, na direção horária quando visto a partir do C_α	31
Figura 10 - Nos diagramas de Ramachandran as regiões vermelhas, amarelas, amarelas claro e brancas, representam respectivamente as regiões permitida, adicionalmente permitida, generosamente permitida e não permitida. (a) complexo CDK1:flavopiridol (b) complexo CDK2:flavopiridol (c) complexo CDK1:h717 (d) complexo CDK2:h717 (e) complexo CDK1:indirubina-5-sulfonato (f) complexo CDK2:indirubina-5-sulfonato.....	38
Figura 11 - Nos diagramas de Ramachandran as regiões vermelhas, amarelas, amarelas claro e brancas, representam respectivamente as regiões permitida, adicionalmente permitida, generosamente permitida e não permitida. (a) complexo CDK1:Isopenteniladenina (b) complexo CDK2:Isopenteniladenina (c) complexo CDK1:Olomoucina (d) complexo CDK2:Olomoucina (e) complexo CDK1:Oxindol-16 e (f) complexo CDK2:Oxindol-16.....	39
Figura 12 - Nos diagramas de Ramachandran as regiões vermelhas, amarelas, amarelas claro e brancas, representam respectivamente as regiões permitida, adicionalmente permitida, generosamente permitida e não permitida. (a) complexo CDK1:Purvalanol (b) complexo CDK2:Purvalanol (c) complexo CDK1:Roscovitina (d) complexo CDK2:Roscovitina (e) complexo CDK1:Estaurosporina e (f) complexo CDK2:Estaurosporina.....	40
Figura 13 - Alinhamento das seqüências da CDK1 e CDK2. Há 67,34% de identidade entre as seqüências. Os resíduos idênticos ou conservados são representados em vermelho, já os resíduos que não apresentam nenhuma identidade estão em azul e pretos.....	41
Figura 14 - Elementos de estrutura secundária da CDK1.....	42



Figura 15 - Representação dos sítios de ligação e todos os resíduos que realizam as ligações de hidrogênio com o inibidor dos complexos estudados (a) CDK1:flavopiridol (b) CDK1:h717 (c) CDK1:indirubina-5-sulfonato (d) CDK1:isopenteniladenina (e) CDK1:olomoucina	43
Figura 16 - Superfície do potencial eletrostático. A região em azul representa os resíduos carregados positivamente, em vermelho são os carregados negativamente e em branco os resíduos neutros.....	44
Figura 17 - Superfície do potencial eletrostático do sítio de ligação (a) CDK1:flavopiridol (b) CDK1:h717 (c) CDK1:indirubina-5-sulfonato (d) CDK1:isopenteniladenina (e) CDK1:olomoucina (f) CDK1:oxindol-16 (g) CDK1:purvalanol (h) CDK1:roscovitina e (i) CDK1:estaurosporina.	45
Figura 18 - Doadores e aceitadores de hidrogênio da CDK.	47
Figura 19 - Sobreposição dos bolsões de ligação dos complexos, em azul é apresentado o complexo CDK1:flavopiridol e em vermelho o complexo CDK2:flavopiridol.....	52
Figura 20 - Sobreposição dos bolsões de ligação dos complexos, em azul é apresentado o complexo CDK1:h717 e em vermelho o complexo CDK2:h717.	53
Figura 21 - Sobreposição dos bolsões de ligação dos complexos, em azul é apresentado o complexo CDK1:indirubina-5-sulfonato e em vermelho o complexo CDK2:indirubina-5-sulfonato.....	53
Figura 22 - Sobreposição dos bolsões de ligação dos complexos, em azul é apresentado o complexo CDK1:isopenteniladenina e em vermelho o complexo CDK2:isopenteniladenina. ...	54
Figura 23 - Sobreposição dos bolsões de ligação dos complexos, em azul é apresentado o complexo CDK1:olomoucina e em vermelho o complexo CDK2:olomoucina.....	54
Figura 24 - Sobreposição dos bolsões de ligação dos complexos, em azul é apresentado o complexo CDK1:Oxindol-16 e em vermelho o complexo CDK2: Oxindol-16.	55
Figura 25 - Sobreposição dos bolsões de ligação dos complexos, em azul é apresentado o complexo CDK1:purvalanol e em vermelho o complexo CDK2:purvalanol.	55
Figura 26 - Sobreposição dos bolsões de ligação dos complexos, em azul é apresentado o complexo CDK1:roscovitina e em vermelho o complexo CDK2: roscovitina.....	56
Figura 27 - Sobreposição dos bolsões de ligação dos complexos, em azul é apresentado o complexo CDK1:estaurosporina e em vermelho o complexo CDK2:estaurosporina.	56
Figura 28 - Derivação de uma pdf característica 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) $W'(d/d', d'', s', s'')$, onde $0,2 < s' < 0,4$ e $0,2 < s'' < 0,4$. c) $W'(d/d', d'', s', s'')$, onde $0,2 < s' < 0,4$ e $0,4 < s'' < 0,6$. $W'(d'/d', d'', s', s'')$, onde $0,4 < s' < 0,6$ e $0,2 < s'' < 0,4$	77
Figura 29 - Interface do ParModel para modelagem.....	83
Figura 30 - Interface do ParModel para a avaliação realizada pelo PROCHECK.....	84
Figura 31 - Interface do ParModel para a avaliação realizada pelo 3DAnalysis.	86

Figura 32 - Interface do ParModel para otimização do modelo.....87

Figura 33 - Interface do ParModel para a visualização realizada ProtGif.....88

Tabela 1 - Nome completo, fórmula molecular e massa molecular de todos os inibidores 16

Tabela 2 - Resultados das análises de qualidade estereoquímica 37

Tabela 3 - Ligações de hidrogênio entre o Garfo Molecular das CDK1 e CDK2 e Inibidores 46

Tabela 4 - Ligações de hidrogênio formadas nos complexos CDK-flavopiridol 48

Tabela 5 - Ligações de hidrogênio formadas nos complexos CDK-IG17 48

Tabela 6 - Ligações de hidrogênio formadas nos complexos CDK-indinavira-5-sulfonato 49

Tabela 7 - Ligações de hidrogênio formadas nos complexos CDK-olomustina 49

Tabela 8 - Ligações de hidrogênio formadas nos complexos CDK-oxindol-16 50

Tabela 9 - Ligações de hidrogênio formadas nos complexos CDK-purvalanol 50

Tabela 10 - Ligações de hidrogênio formadas nos complexos CDK-roscovitina 50

Tabela 11 - Ligações de hidrogênio formadas nos complexos CDK-estaurosporina 51

Tabela 12 - Ligações de hidrogênio formadas nos complexos CDK-isopenteniladenina 51

Tabela 13 - Resultado de G-factor e 3D Profile do complexo CDK1-Inibidor e moléculas 58

Tabela 14 - Valores de RMSD para a Geometria Ideal 59

Tabela 15 - Valores de RMSD obtidos através da sobreposição dos complexos e para os 60

Tabela 16 - Código e resolução das estruturas cristalográficas de CDK2 usadas como modelos, valores do Contato de Superfícies e de IC₅₀ dos complexos CDK1-Inibidores e CDK2-Inibidores 61

Tabela 17 - Tamanho do tempo de dados local 75

Tabela 18 - Classes Estruturais de Proteínas 76

Índice de Tabela

Tabela 1 - Nome completo, fórmula molecular e massa molecular de todos os inibidores.	16
Tabela 2 - Resultados das análises de qualidade estereoquímica.....	37
Tabela 3 - ligações de hidrogênio entre o Garfo Molecular das CDK1 e CDK2 e Inibidores. ..	46
Tabela 4 - Ligações de hidrogênio formadas nos complexos CDK:flavopiridol.....	48
Tabela 5 - Ligações de hidrogênio formadas nos complexos CDK:h717.	48
Tabela 6 - Ligações de hidrogênio formadas nos complexos CDK:indirubina-5-sulfonato.	49
Tabela 7 - Ligações de hidrogênio formadas nos complexos CDK:olomoucina.....	49
Tabela 8 - Ligações de hidrogênio formadas nos complexos CDK:oxindol-16.	50
Tabela 9 - Ligações de hidrogênio formadas nos complexos CDK:purvalanol.....	50
Tabela 10 - Ligações de hidrogênio formadas nos complexos CDK:roscovitina.	50
Tabela 11 - Ligações de hidrogênio formadas nos complexos CDK:estaurosporina.....	51
Tabela 12 - Ligações de hidrogênio formadas nos complexos CDK:isopenteniladenina.	51
Tabela 13 - Resultado de G-fator e 3D Profile do complexo CDK1:Inibidor e moldes.	58
Tabela 14 - Valores de RMSD para a Geometria Ideal.	59
Tabela 15 - Valores de RMSD obtidos através da sobreposição dos complexos e para os.....	60
Tabela 16 - Código e resolução das estruturas cristalográficas de CDK2 usadas como moldes, valores de contato de superfícies e de IC ₅₀ dos complexos CDK1:Inibidores e CDK2:Inibidores.	61
Tabela 17 - Tamanho do banco de dados local.	75
Tabela 18 - Classes Estruturais de Proteínas.	76



Índice de Abreviações

CDK	Cyclin-Dependent Kinase, Quinase dependente de Ciclina
ATP	Adenosina Tri-fosfato
DNA	Deoxyribonucleic acid, ácido desoxirribonucleico
Rb	Retinoblastoma
PIR	Protein Identification Resource, Busca para identificação de proteína
3D	Tridimensional
RMN	Nuclear Magnetic Resonance, Ressonância Magnética Nuclear
RMSD	Root mean Squared Deviations, desvio quadrático médio
PDB	Protein Data Bank, Banco de dados de proteínas
PDF	Probability density function, função de densidade de probabilidade
IC50	Inhibitory Concentration 50%, Concentração inibitória mínima 50%

Banco de dados

GenBank www.ncbi.nlm.nih.gov/genbank/

PDB www.rcsb.org/pdb/

SWISS-PROT www.ebi.ac.uk/swiss-prot/

Alinhamento de seqüências

CLUSTAL www.ch.roc.uk/~clustal/

FastA www1.ch.roc.uk/~fasta/

MULTALIN <http://prodes.torvalds.org/ftp/multi/multi.html>



Glossário de termos utilizados e endereços da Internet

MODELLER: programa para modelagem molecular por homologia.

<http://salilab.org/modeller/>

ParModel: programa que realiza a automatização e paralelização do programa MODELLER.

<http://www.biocristalografia.df.ibilce.unesp.br/tools/parmodel/>

PROCHEK: programa para análise da qualidade estereoquímica dos modelos.

http://laboheme.df.ibilce.unesp.br/cluster/parmodel_mpi/

Verify 3D: programa que mede a compatibilidade entre uma determinada sequência de aminoácido e a estrutura 3D de uma proteína.

http://www.doe-mbi.ucla.edu/Services/Verify_3D/

X-Plor: programa para refinamento de estrutura tridimensional de macromoléculas, usando dados de difração por raios X ou ressonância magnética.

<http://xplor.csb.yale.edu/xplor/download/distributions/>

Banco de dados

GenBank www.ncbi.nlm.nih.gov/GenBank

PDB www.rcsb.org/pdb/

SWISS-PROT www.ebi.ac.uk/swissprot

Alinhamento de seqüências

CLUSTAL www.ebi.ac.uk/clustalw/

FastA www2.ebi.ac.uk/fasta3

MULTALIN <http://prodes.toulouse.inra.fr/multalin/multalin.html>



RESUMO

Quinases dependentes de ciclinas (CDK) controlam a progressão do ciclo celular. As CDKs são inativas como monômeros, para sua ativação é necessária sua ligação às ciclinas, uma família de proteínas, cuja concentração oscila durante o ciclo celular, e a fosforilação pela quinase ativadora de CDK (CAK), em um resíduo específico de treonina. Além da função regulatória positiva das ciclinas e CAK, existem muitas proteínas regulatórias negativas, os inibidores intrínsecos de CDKs, as CKIs. Alterações ou ausências de CKIs estão relacionados com muitos tipos de cânceres. O potencial terapêutico de inibidores naturais encorajam a busca por inibidores químicos de CDKs. Estes inibidores bloqueiam a progressão do ciclo celular e apresentam potencial atividades anti-tumoral. Existem outras aplicações relacionadas com inibidores de CDKs, como por exemplo, o controle de algumas viroses, parasitas protozoários e doenças neurodegenerativas. Análises cristalográficas em conjunto com estudos em bioinformática de CDKs geram novas informações sobre a base estrutural para a inibição de CDKs.

Com o objetivo de estudar a interação da proteína CDK1 com vários inibidores e estabelecer as bases estruturais para a inibição da CDK1, foram realizadas as modelagens moleculares dos complexos binários CDK1:Inibidores, utilizando como molde a estrutura CDK2 complexada com os mesmos inibidores. Para a realização dessas modelagens foi utilizado o programa PARMODEL, que é uma implementação paralela do programa MODELLER. Os inibidores estudados foram flavopiridol, h717, purvalanol, estaurosporina, indirubina-5-sulfonato, oxindol 16 (composto3), roscovitina, isopenteniladenina e olomoucina.

O modelo da CDK1 apresenta o garfo molecular no sítio ativo, como observado nas estruturas da CDK2, CDK5 e CDK9, e apresenta as mesmas características estruturais: dois



lóbulo principais, com o lóbulo maior, C-Terminal formado principalmente de hélices α e o lóbulo menor, N-Terminal, formado principalmente de folhas β .

Os resultados obtidos mostram que esses inibidores estão fortemente ligados ao sítio de ligação do ATP, sugerindo que os inibidores propostos podem inibir a CDK1 com a mesma eficiência apresentada para a CDK2.



Cyclin-dependent kinases (CDKs) control the progression of the cell division cycle. CDKs are inactive as monomers and their activation require binding to cyclins, a diverse family of proteins whose levels oscillate during the cell cycle, and the phosphorylation by CDK-activating kinase (CAK) on a specific threonine residue. In addition to the positive regulatory role of cyclins and CAK, there are many negative regulatory proteins (CDK inhibitors, CKIs). Deregulation or absence of CKIs has been associated with many cancers. The therapeutic potential of natural inhibitors have encouraged an active search for chemical inhibitors of CDKs. These inhibitors block cell cycle progression and display potentially interesting antitumor activities. There are other application related with inhibitors of CDKs, for example, the control of some viruses, parasitic protozoa and neurodegeneration diseases. The crystallographic analysis together with bioinformatics studies of CDKs are generating new information about the structural basis for inhibition of CDKs.

In order to study the interaction of the CDK1 protein with several inhibitors and to establish structural bases for the inhibition of the CDK1, were carried out the molecular model of the binary complexes CDK1:Inhibitors, utilizing as template the CDK2 structure complexed with the same inhibitors. For the achievement of those modeling was utilized the PARMODEL program, a parallel implementation of the MODELLER program. The inhibitors studied were flavopiridol, h717, purvalanol, staurosporine, indirubin, oxindole 16 (compound3), roscovitine, isopentenyladenine and oloumocine.

CDK1 presents the molecular fork in the active site, as observed in CDK2, CDK5 and CDK9 structures, and presents the same structural characteristics: two main lobes, with the larger C-terminal lobe consisting predominantly of α -helices and the smaller N-terminal lobe consisting predominantly of β -sheet.



The results of models CDK1:inhibitors point out that these inhibitors binds strongly and with the same efficiency to ATP binding site as observed for CDK2.

INTRODUÇÃO

As maiores perspectivas de novas descobertas no campo das ciências médicas nas próximas décadas, estão centradas no Projeto Genoma. Os pares de bases sequenciados recentemente no genoma humano constituem importantes informações para a elucidação do genoma completo (VENTER *et al.*, 2001, LANDER *et al.*, 2001). A compreensão de fenômenos moleculares exerce uma função cada vez mais relevante, o que certamente implicará em avanços significativos na medicina, principalmente, em técnicas de diagnóstico e tratamento. Com isso, espera-se que seja possível iniciar os tratamentos clínicos no feto ou na primeira infância, muito antes do surgimento dos primeiros sintomas (SANTOS FILHO & DE ALENCASTRO, 2001).

Neste contexto, a determinação do maior número possível de estruturas tridimensionais (estruturas terciárias e quaternárias) de proteínas é fundamental para novos avanços no projeto genoma e na medicina (NGUYEN *et al.*, 2000; TERWILLIGER, 2000; YOKOYAMA *et al.*, 2000; HEINEMANN, 2000). Apesar das consideráveis inovações técnicas, sobretudo nas áreas de cristalografia por raios X e ressonância magnética nuclear (RMN), muitos problemas básicos persistem. A obtenção de amostras em quantidade suficiente para os ensaios necessários e, em muitos casos, difícil e os cristais obtidos nem sempre têm a qualidade necessária para as experiências (somente 10% das proteínas, aproximadamente, produzem cristais adequados) (MAGGIO & RAMNARAYAN, 2001). Além disso, em certas classes de proteínas, como por exemplo, as proteínas de membrana celular, que possuem baixa solubilidade em meios aquosos, a determinação estrutural é um desafio. Essas proteínas



INTRODUÇÃO

As maiores perspectivas de novas descobertas no campo das ciências médicas nas próximas décadas, estão centradas no Projeto Genoma. Os pares de bases seqüenciados recentemente no genoma humano constituem importantes informações para a elucidação do genoma completo (VENTER *et al.*, 2001; LANDER *et al.*, 2001). A compreensão de fenômenos moleculares exerce uma função cada vez mais relevante, o que certamente implicará em avanços significativos na medicina, principalmente, em técnicas de diagnóstico e tratamento. Com isso, espera-se que seja possível iniciar os tratamentos clínicos no feto ou na primeira infância, muito antes do surgimento dos primeiros sintomas (SANTOS FILHO & DE ALENCASTRO, 2001).

Neste contexto, a determinação do maior número possível de estruturas tridimensionais (estruturas terciárias e quaternárias) de proteínas é fundamental para novos avanços no projeto genoma e na medicina (NORVELL *et al.*, 2000; TERWILLIGER, 2000; YOKOYAMA *et al.*, 2000; HEINEMANN, 2000). Apesar das consideráveis inovações técnicas, sobretudo nas áreas de cristalografia por raios X e ressonância magnética nuclear (RMN), muitos problemas básicos persistem. A obtenção de amostras em quantidade suficiente para os ensaios necessários é, em muitos casos, difícil e os cristais obtidos nem sempre têm a qualidade necessária para as experiências (somente 10% das proteínas, aproximadamente, produz cristais adequados) (MAGGIO & RAMNARAYAN, 2001). Além disso, em certas classes de proteínas, como por exemplo, as proteínas de membrana celular, que possuem baixa solubilidade em meios aquosos, a determinação estrutural é um desafio. Essas proteínas



raramente cristalizam e dificilmente podem ser tratadas de modo satisfatório pela técnica de RMN.

Por outro lado, a elucidação das seqüências de aminoácidos (estruturas primárias) é uma tarefa relativamente mais simples. Por isto, é visto uma grande diferença entre o número de estruturas primárias e tridimensionais disponíveis. Para se ter uma idéia do problema, em fevereiro de 2005 o SWISS-PROT (<http://bo.expasy.org/sprot/>), o mais importante banco de dados de estruturas primárias, incluía 170.140 seqüências primárias, neste mesmo período, somente 29.040 estruturas protéicas estavam disponíveis no PDB (*PROTEIN DATA BANK*) (<http://www.rcsb.org/pdb/> - BERMAN *et al.*, 2000), o principal banco de dados de estruturas 3D de proteínas. Em consequência, foram desenvolvidos outros métodos para a determinação da estrutura 3D de proteínas. Como por exemplo, o método *ab initio* que parte do princípio que é possível determinar a estrutura tridimensional de proteínas a partir de sua estrutura primária (ORTIZ *et al.*, 1998; OSGUTHORPE, 1997; SIMONS *et al.*, 1997; SIMONS *et al.*, 1999).

A ferramenta mais bem sucedida para a determinação de estruturas tridimensionais de proteínas é a modelagem por homologia, também conhecida como modelagem comparativa (*comparative protein modeling*) (SANTOS FILHO & DE ALENCASTRO, 2001; D'ALFONSO *et al.*, 2001; PEITSCH e CHARIFSON, 1997; ROST *et al.*, 1996; JOHNSON *et al.*, 1994). Esta abordagem é fundamentada em alguns padrões gerais que foram observados, em nível molecular, no processo de evolução biológica:

- a) Homologia entre seqüências de aminoácidos implica em semelhança estrutural e funcional;
- b) Proteínas homólogas apresentam regiões internas conservadas (constituídas principalmente de elementos de estrutura secundária: hélices- α e fitas- β);



c) As principais diferenças estruturais entre proteínas homólogas ocorrem nas regiões externas, constituídas principalmente por alças (*loops*) (SANTOS FILHO & DE ALENCASTRO, 2001).

Outro fato importante é que as proteínas se unem em um número limitado de famílias tridimensionais. Existem cerca de 40.000 famílias protéicas (ROBERT, 2005). Conseqüentemente, quando se conhece a estrutura de pelo menos um representante de uma família, é geralmente possível modelar, por homologia, os demais membros da família (SANTOS FILHO & DE ALENCASTRO, 2003). Nesta dissertação é apresentado a modelagem comparativa ou por homologia, de nove complexos CDK1:Inibidor utilizando como molde os complexos CDK2:Inibidor. Dos complexos utilizados como molde, cinco (CDK2:h717, CDK2:purvalanol, CDK2:estaurosporina, CDK2:indirubina-5-sulfonato, CDK2:oxindol 16 (composto3), possuem estruturas tridimensionais depositadas no PDB. Os quatro complexos restantes (CDK2:roscovitina, CDK2:isopenteniladenina, CDK2:olomoucina e CDK2:flavopiridol) tem estruturas tridimensionais resolvidas mas ainda não foram depositadas no PDB (KIM *et al.*, 1996; SCHULZE-GAHMEN *et al.*, 1995; DE AZEVEDO *et al.*, 1996).

1.1 Ciclo Celular e CDK

Todos os organismos são constituídos por células, sendo que a maioria se multiplicam por divisão celular. Um indivíduo adulto possui aproximadamente 100 bilhões de células, todas derivadas de uma única célula. O ciclo celular é dividido em quatro fases distintas (M, G1 (e G0), G2 e S) (figura 1). Durante a fase M (mitose), que é relativamente curta (cerca de uma hora), ocorre a divisão celular. A fase G1 que utiliza o maior período do ciclo celular (6 a 12 horas) é caracterizada por expressão de genes e síntese de proteínas. Isto permite a célula crescer e produzir todas as proteínas necessárias para a síntese de DNA. Durante a fase S (6 a 8 horas), a



célula replica seu DNA, dando origem a uma cópia de cada cromossomo. Na fase G2 (3 a 4 horas), a célula novamente cresce e prepara-se para a mitose. Ao completar esta fase a célula entra na fase M novamente, e que é a divisão celular propriamente dita, podendo em seguida iniciar um novo ciclo. Muitas células terminalmente diferenciadas, como os neurônios ou células musculares que não se dividem entram em um estado quiescentes conhecido como G0 (VOET & VOET, 1995) (Alterações nestas fases do ciclo de divisão celular, resultar em alterações na duplicação dos cromossomos). Cromossomos ou partes de cromossomos podem ser perdidos, rearranjados ou distribuídos desigualmente entre as células duplicadas. Estas alterações são observadas com frequência em células cancerosas (WARD, 2002).

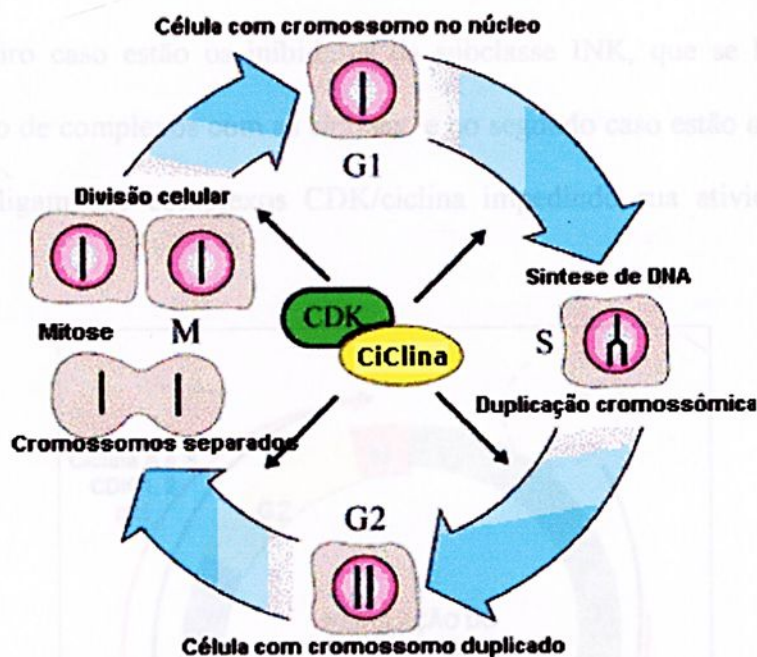


Figura 1 - Fases do ciclo celular e duplicação cromossômica.

A progressão pelo ciclo celular é, em parte, controlada, por uma série de proteínas chamadas quinases dependentes de ciclinas (CDKs), particularmente nas transições de fases, tanto de G1 para S quanto de G2 para M. As CDKs são proteínas com massa molecular entre 33-45 kDa, são inativas como monômeros necessitando se ligar a uma ciclina, que são proteínas com a massa molecular entre 30-45 kDa, para a sua ativação (JEFFREY *et al.*, 1995).

Os níveis de ciclinas oscilam durante as fases do ciclo, determinando o momento apropriado de sua ligação às CDKs (figura 2). As ciclinas se associam às CDKs, através de um domínio específico conhecido como a caixa de ciclina (*cyclin box*) (HUNT, 1991). Após a associação da ciclina à CDK esta sofre uma ativação pela fosforilação de resíduos de treonina, serina e tirosina por proteínas ativadoras de CDKs (CAK). Através do domínio catalítico da CDK, conhecido como núcleo (*core*) catalítico, complexos CDK/ciclina fosforilam substratos específicos, ativando-os em diferentes fases do ciclo celular (JEFFREY *et al.*, 1995).

Ao contrário das ciclinas, que tem a função de promover à progressão do ciclo celular, os inibidores do ciclo celular, ou seja, inibidores de CDKs (CKIs), que são proteínas geralmente de baixa massa molecular que se ligam a CDKs ou a complexos CDK/ciclina (figura 3). No primeiro caso estão os inibidores da subclasse INK, que se ligam a CDKs impedindo a formação de complexos com as ciclinas, e no segundo caso estão os inibidores da subclasse KIP, que ligam aos complexos CDK/ciclina impedindo sua atividade quinásica (HUNT, 1991).

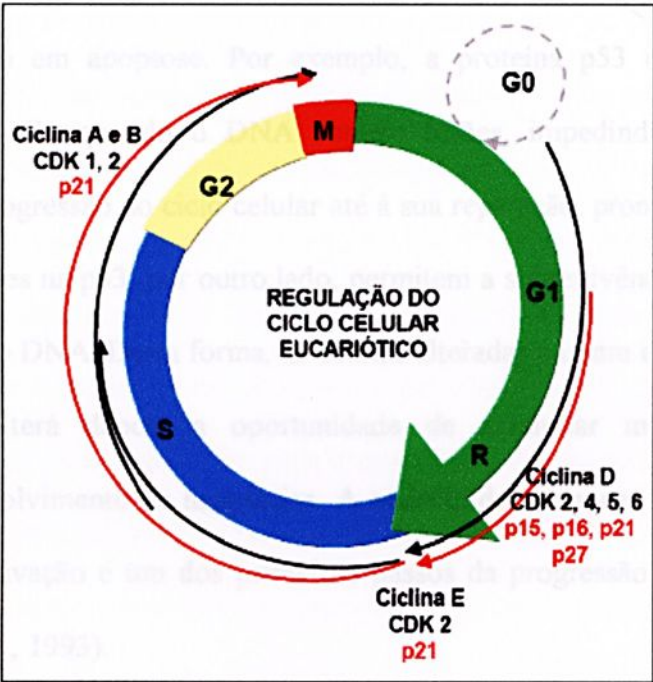


Figura 2 - Proteínas que regulam as fases do ciclo celular, positiva (ciclinas) e/ou negativamente (CKIs) que são representados em vermelho.

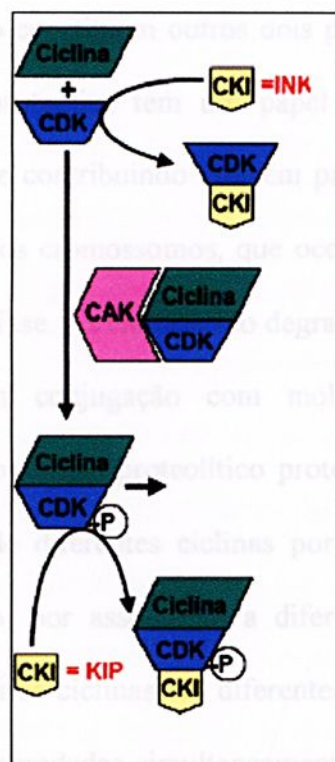


Figura 3 - Mecanismos de inibição dos CKIs durante o ciclo celular, inibição realizada pela subclasse INK que se liga a CDK impedindo a formação do complexo CDK/ciclina ou pela KIP que se liga a complexo já formado impedindo que a CDK fosforile outros substratos

As proteínas p53 e a retinoblastoma (Rb) descritas como supressoras de tumores estão envolvidas no controle do ciclo celular, sendo centrais na progressão e/ou interrupção do ciclo celular e entrada ou não em apoptose. Por exemplo, a proteína p53 é responsável pela interrupção do ciclo em G1 quando o DNA contém lesões, impedindo a ocorrência da replicação, atrasando a progressão do ciclo celular até à sua reparação, promovendo a apoptose (suicídio celular). Mutações na p53, por outro lado, permitem a sobrevivência e reprodução de células contendo lesões no DNA. Desta forma, as células alteradas passam essas mutações para sua descendência, que terá depois a oportunidade de acumular mutações adicionais responsáveis pelo desenvolvimento de neoplasias. A maioria dos tumores possuem alterações nesta proteína, a sua inativação é um dos primeiros passos da progressão da célula para um estado maligno (CAI *et al.*, 1993).

A proteólise e a transcrição constituem outros dois processos reguladores importantes para a progressão do ciclo. A proteólise tem um papel direto na regulação das CDKs controlando os níveis de ciclinas, e contribuindo também para outros passos do ciclo celular tais como alterações nas cópias dos cromossomos, que ocorrem quando os cromossomos se separam, na transição metafase/anafase. As ciclinas são degradadas de um modo seletivo sendo “marcadas” para destruição pela conjugação com moléculas de ubiquitina, e assim reconhecidas e degradadas pelo complexo proteolítico proteossoma. Por exemplo, durante a mitose ocorrem a ubiquitinação de diferentes ciclinas por ação de um complexo protéico denominado APC. Este complexo, por associação a diferentes proteínas reguladoras, tem especificidade diferente para as várias ciclinas em diferentes fases da mitose. Deste modo as ciclinas mitóticas não são todas degradadas simultaneamente, mas sim em fases específicas. Assim, aumento e diminuição sucessivas na concentração de cada ciclina são obtidos pela coordenação entre a sua transcrição e a sua degradação pelo proteossoma. Este duplo controle assegura que a concentração de cada ciclina atinja o nível necessário para ativar o seu parceiro catalítico (CDK), na fase correta do ciclo celular. A destruição programada das células (apoptose) é, por outro lado, um passo importante que permite à célula passar à próxima fase do ciclo. Por esta razão, células mutantes que não conseguem degradar as ciclinas mitóticas ficam paradas em mitose e não entram na fase G1 (VALENTINI *et al.*, 2004).

Recentemente estudos realizados com células eucarióticas onde genes de algumas CDKs foram suprimidos resultaram na progressão de diferentes fases do ciclo celular, indicando que as CDKs não são essenciais para a progressão do ciclo (ALEEM *et al.*, 2004; BERTHET *et al.*, 2003; MALUMBRES *et al.*, 2004). Mesmo assim o interesse na descoberta de inibidores de CDKs como possíveis drogas anti-tumorais continua grande, visto que diversos inibidores de CDKs são eficientes como drogas anti-tumorais, entre eles o flavopiridol e a roscovitina (CANDURI & DE AZEVEDO, 2005).



1.2 Quinase dependente de Ciclina 1 (CDK1)

Em meados da década de 70, Paul Nurse seguindo a abordagem de Hartwell – que usou métodos genéticos e um sistema modelo de levedura *Saccharomyces cerevisiae* (DORER *et al.*, 1995) para estudar ciclo celular, e entre 70 e 71 descobriu centenas de genes envolvidos na progressão do ciclo celular - descobriu o gene *cdc2* utilizando outro tipo de levedura como sistema modelo: a *Schizosaccharomyces pombe*. Nurse demonstrou que este gene tem um papel chave no controle da divisão celular (transição da fase G2 para a mitose, M). Mais tarde descobriu que este gene tinha uma função mais geral. Era idêntico ao gene *cdc28 "start"*, que Hartwell identificara anteriormente na levedura *S. cerevisiae*, controlando a transição de G1 para S. a Cdc2 estaria então envolvido na regulação de diferentes passos do ciclo celular. Em 1987, Nurse isolou o gene correspondente em humanos, ao qual foi posteriormente dado o nome de CDK1 (quinase dependente de ciclina 1). O gene codifica uma proteína que é membro da família das quinases dependentes de ciclinas (CDK). Nurse demonstrou que a ativação das CDKs é dependente de fosforilação ou desfosforilação. Com base nestas descobertas, várias CDKs foram encontradas em humanos (LEE e NURSE, 1987).

A CDK1 é composta por 297 resíduos, sua massa molecular é de 33 kDa. Ela está envolvida na progressão do ciclo celular. Na fase G1 a CDK1 sofre a fosforilação do resíduo de treonina 161 pela quinase ativadora de CDK, a CAK, que nesta fase é exercida pela CDK7. Na fase seguinte (Fase S), a CDK1 se associa a ciclina B formando o complexo Ciclina B/CDK e duas outras proteínas a *weel* e a *myt1* são responsáveis pela fosforilação dos resíduos de treonina 14 e tirosina 15, inativando a CDK. Ao se aproximar da transição G2/M a CAK *cdc25*, que é a proteína responsável pela ativação da CDK1, desfosforila os resíduos treonina 14 e tirosina 15, permitindo a ligação de uma molécula de ATP ao sítio de ligação da CDK1, já na fase M. a CDK1 forma o complexo CDK1/ciclina B, e em seguida fosforila seu substrato alvo.



A fosforilação do “complexo promotor da anáfase” é necessária para a transição para a anáfase, que é a fase na qual ocorre a separação dos cromossomos, e a finalização da mitose, a ciclina é liberada e degradada, e a CDK1 volta à fase G1 (figura 4) (CAI *et al.*, 1993). Esses eventos sucessivos de ativação e inibição dos complexos CDK/ciclina são fortemente regulados por modificações pós-traducionais e translocações intracelular (BRUCE *et al.*, 1994).

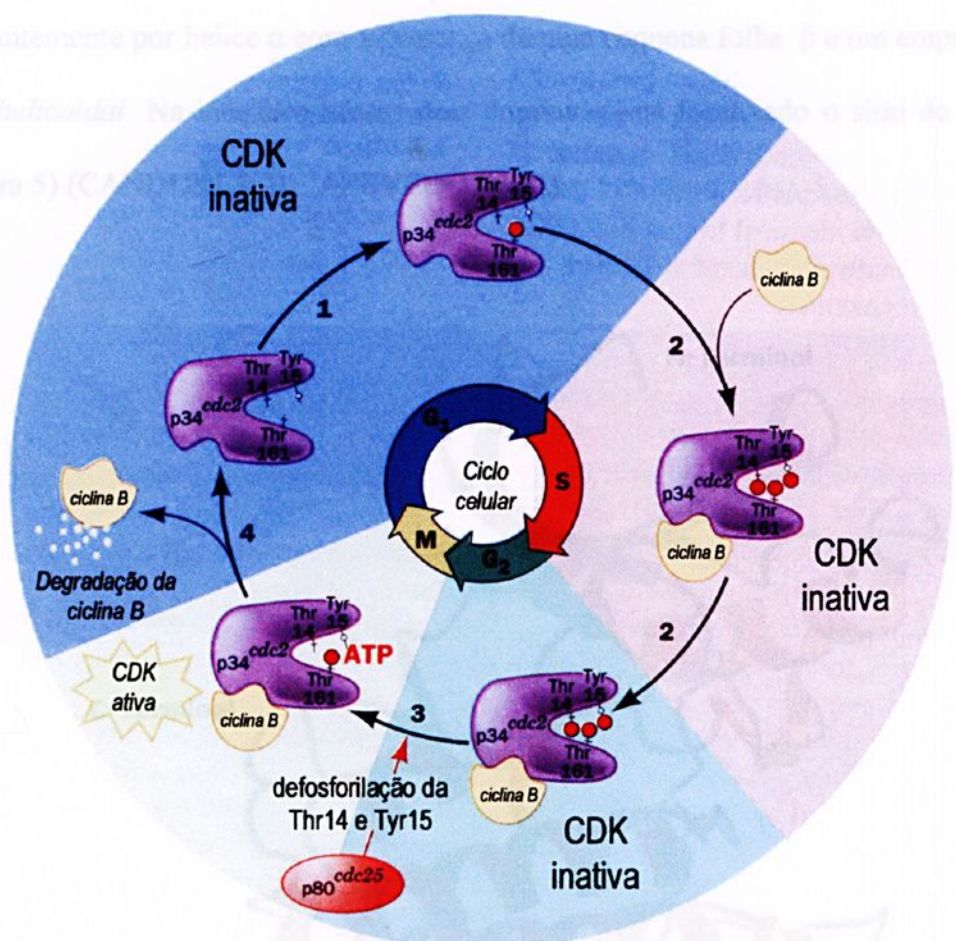


Figura 4 - Atividade da CDK1 durante o ciclo celular em uma célula animal.
(VOET & VOET, 1995)

Figura 5 - Estrutura secundária da CDK2

1.3 Quinase dependente de Ciclina 2 (CDK2)

A CDK2 participa da regulação do ciclo celular, nas fases G1 formando complexos com as ciclinas D (1-3) e na fase G2 formando complexo com a ciclina A, é uma proteína formada por 298 resíduos, sua massa molecular é 34 kDa. A CDK2 apresenta uma estrutura bilobular, com um pequeno domínio N-terminal formado por cinco fitas β antiparalelas que formam uma folha β e a presença de uma única hélice α . O domínio maior C-terminal é formado predominantemente por hélice α com a presença de uma pequena folha β e um empacotamento *pseudo-4-helicoidal*. Na interface desses dois domínios está localizado o sítio de ligação do ATP (figura 5) (CANDURI & DE AZEVEDO, 2005).

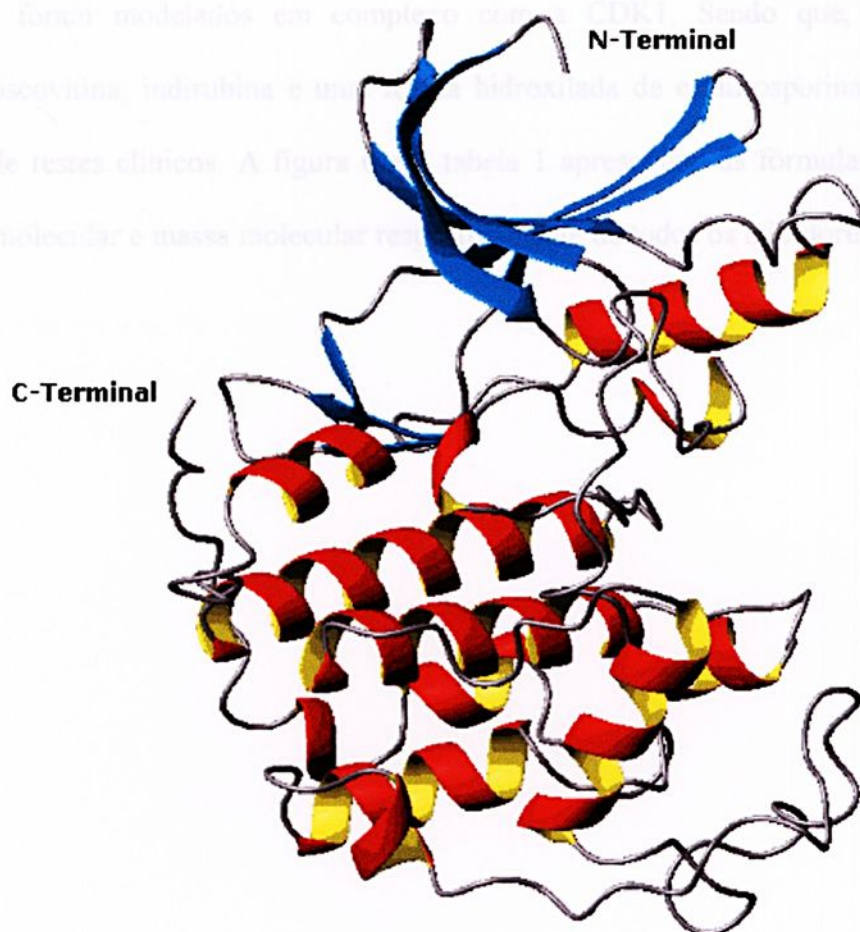


Figura 5 - Estrutura secundária da CDK2

1.4 Inibidores

Os inibidores de CDK2 apresentam algumas características similares que facilitam sua ligação ao sítio ativo da CDK2: 1) possuem baixa massa molecular ($< 600\text{Da}$); 2) eles são planos, heterocíclicos e hidrofóbicos; 3) competem com o ATP para se ligar ao sítio ativo da quinase; 4) sua ligação a quinase é feita principalmente por interações hidrofóbicas e por ligação de hidrogênio e 5) o carbono da cadeia principal e amina da cadeia lateral da Leu83, agem respectivamente, como um aceitador e um doador de hidrogênio aos inibidores, ao passo que o carbono da cadeia principal do Glu81 age frequentemente como um aceitador de hidrogênio (DE AZEVEDO *et al.*, 1997). Existem mais de 50 moléculas descritas como possíveis inibidores da CDK2. Nesta dissertação são apresentadas informações de nove desses inibidores, que foram modelados em complexo com a CDK1. Sendo que, cinco deles (flavopiridol, roscovitina, indirubina e uma forma hidroxilada de estaurosporina o UCN-01) estão em fase de testes clínicos. A figura 6 e a tabela 1 apresentam as fórmulas estruturais, nome, fórmula molecular e massa molecular respectivamente de todos os inibidores estudados.



Tabela 1 - Nome completo, fórmula molecular e massa molecular de todos os inibidores.

Nome do Inibidor	Fórmula Molecular	Massa Molecular (Da)
Flavopiridol 4H-1-Benzopyran-4-one, 2-(2-chlorophenyl)-5,7-dihydroxy-8-(3-hydroxy-1-methyl-4-piperidinyl)-, hydrochloride, (-)-cis-	$C_{21}H_{20}ClNO_5HCl$	438,00
H717 (2-[Trans-(4-Aminocyclohexyl)Amino]-6-(Benzyl-Amino)-9-Cyclopentylpurine)	$C_{23}H_{31}N_7$	405,5
Purvalanol 2-(1r-Isopropyl-2-Hydroxyethylamino)-6-(3-Chloroanilino)-9-Isopropyl-Purine	$C_{19}H_{25}ClN_6O_3Cl$	388,9
Staurosporine Antibiotic AM-2282	$C_{28}H_{26}N_4O_3$	466,5
Indirubina-5-Sulfonato (2',3-Dioxo-1,1',2',3-Tetrahydro-2,3'-Biindole-5'-Sulfonic Acid)	$(C_{16}H_9N_2O_5S)$	262,1
Oxindole 4-[(7-Oxo-7h-Thiazolo[5,4-E]Indol-8-Ylmethyl)- Amino]-N-Pyridin-2-Yl-Benzenesulfonamide	$C_{21}H_{15}N_5O_3S_2$	133,2
Roscovitina 2-(S)-(1-Ethyl-2-hydroxyethylamino)-6-benzylamino-9-isopropylpurine	$C_{19}H_{26}N_6O$	354,5
Isopenteniladenina ($N^6 - (\Delta^2 - \text{isopentenyl})\text{adenine}$ (iP))	$C_9H_{13}N_5$	174,2
Olomoucina 2-(Hydroxyethylamino)-6-Benzylamino-9-Methylpurine	$C_{15}H_{18}N_6O$	298,3

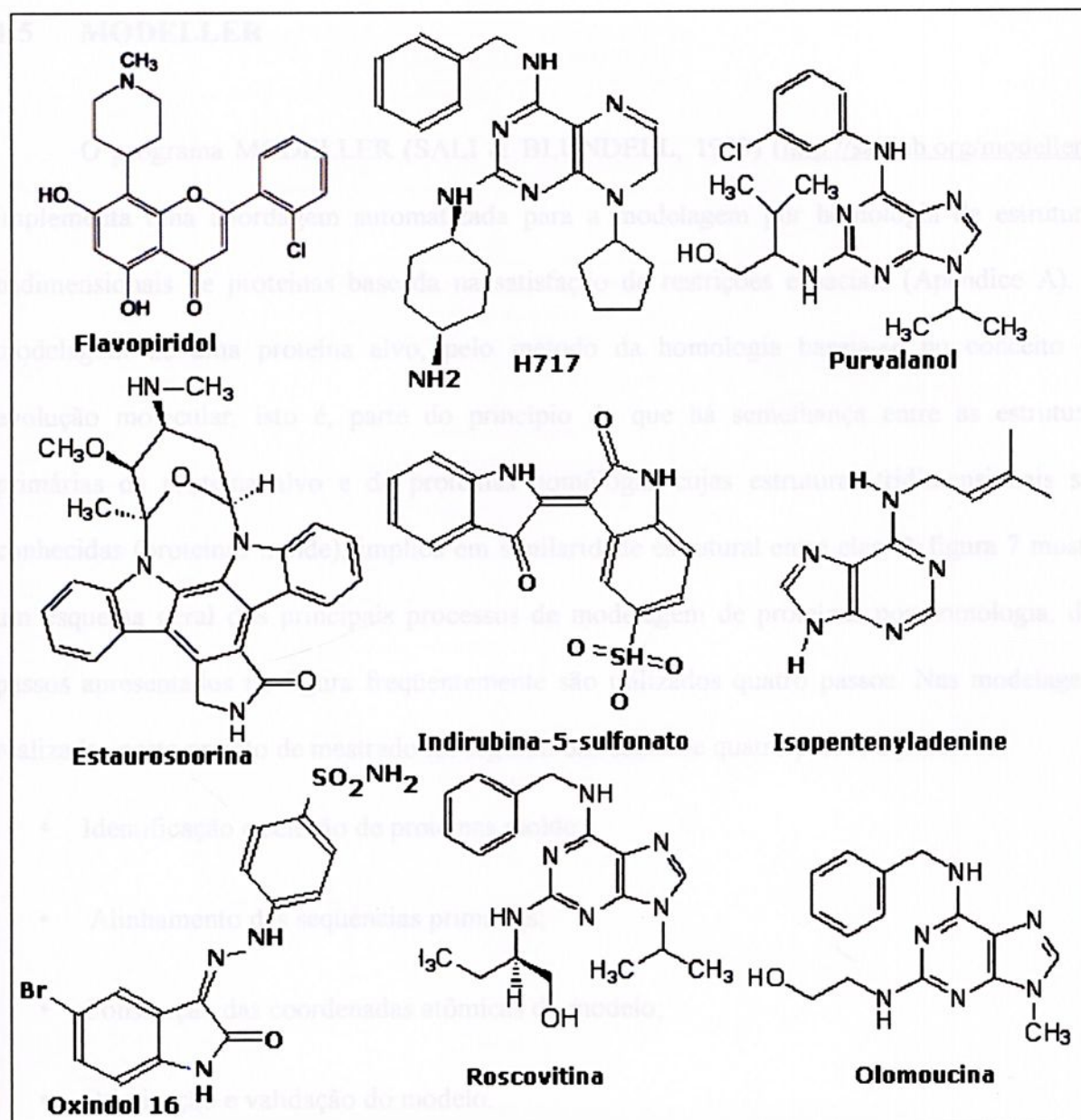


Figura 6 - Fórmula estrutural de todos inibidores estudados

1.5 MODELLER

O programa MODELLER (SALI & BLUNDELL, 1993) (<http://salilab.org/modeller/>), implementa uma abordagem automatizada para a modelagem por homologia de estruturas tridimensionais de proteínas baseada na satisfação de restrições espaciais (Apêndice A). A modelagem de uma proteína alvo, pelo método da homologia baseia-se no conceito de evolução molecular, isto é, parte do princípio de que há semelhança entre as estruturas primárias da proteína alvo e de proteínas homólogas cujas estruturas tridimensionais são conhecidas (proteínas molde), implica em similaridade estrutural entre elas. A figura 7 mostra um esquema geral dos principais processos de modelagem de proteínas por homologia, dos passos apresentados na figura frequentemente são utilizados quatro passos. Nas modelagens realizadas neste projeto de mestrado foi seguido basicamente quatro passos sucessivos:

- Identificação e seleção de proteínas molde;
- Alinhamento das seqüências primárias;
- Construção das coordenadas atômicas do modelo;
- Otimização e validação do modelo.

1.5.1 Identificação e seleção de proteínas molde

A primeira etapa do método é a identificação de pelo menos uma proteína cuja estrutura tridimensional seja conhecida, que servirá de molde para a determinação da estrutura da proteína alvo. Inicialmente é feita uma busca por molde em um banco de dados de estruturas tridimensionais (PDB) (<http://www.rcsb.org/pdb/>). Duas situações são, em geral, possíveis:

a) É conhecida a família proteica à que a proteína alvo pertence;



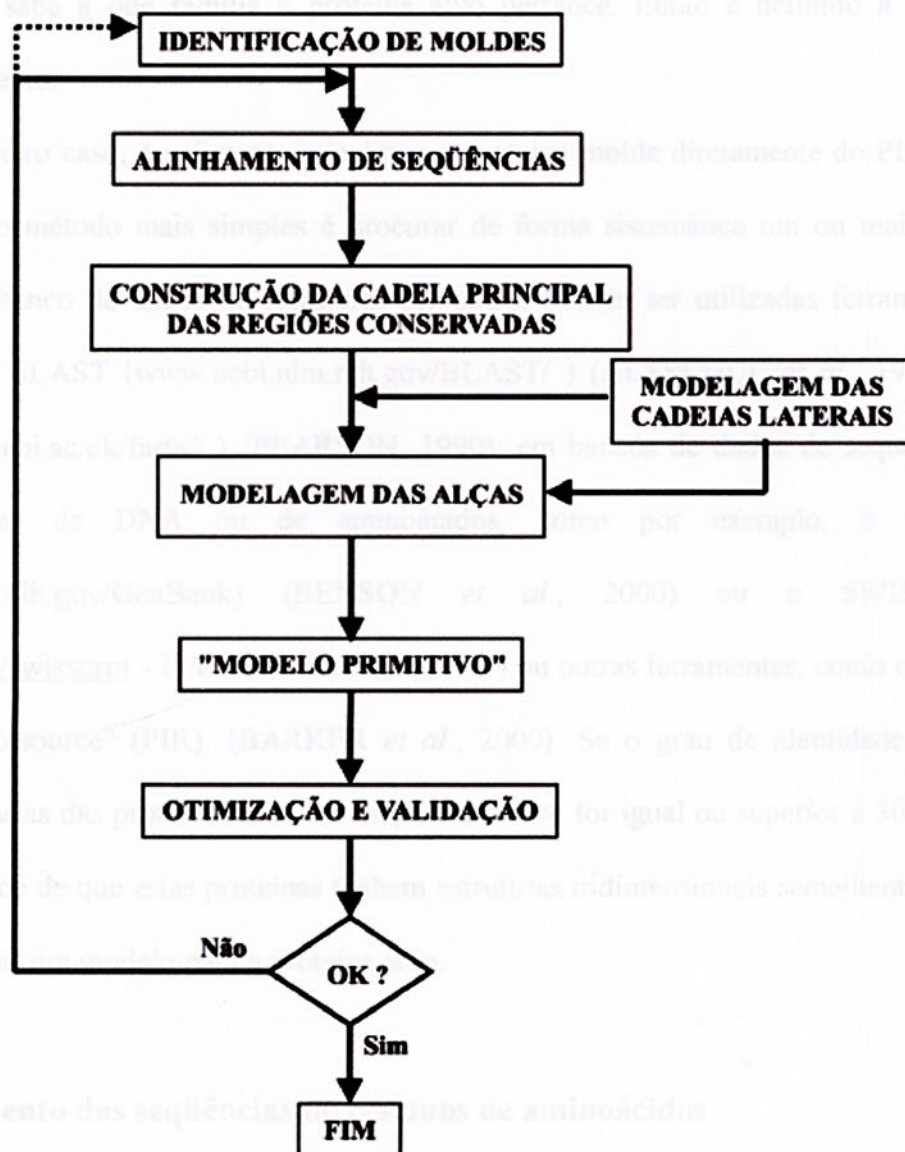


Figura 7 - Esquema geral dos principais passos da modelagem por homologia.

1.5.1 Identificação e seleção de proteínas molde

A primeira etapa do método é a identificação de pelo menos uma proteína cuja estrutura tridimensional seja conhecida, que servirá de molde para a determinação da estrutura da proteína alvo. Inicialmente é feita uma busca por moldes em um banco de dados de estruturas tridimensionais (PDB) (<http://www.rcsb.org/pdb/>). Duas situações são, em geral, possíveis:

- a) É conhecida a família protéica a que a proteína alvo pertence;

b) Não se sabe a que família a proteína alvo pertence. Então é definido a partir do alinhamento.

No primeiro caso, é suficiente selecionar a proteína molde diretamente do PDB. Já no segundo caso, o método mais simples é procurar de forma sistemática um ou mais moldes adequados em banco de dados de estruturas primárias. Podem ser utilizadas ferramentas de busca, como o BLAST (www.ncbi.nlm.nih.gov/BLAST/) (ALTSCHUL *et al.*, 1990) ou o FastA (www2.ebi.ac.uk/fasta3) (PEARSON, 1990), em bancos de dados de seqüências de pares de bases de DNA ou de aminoácidos, como por exemplo, o GenBank (www.ncbi.nlm.nih.gov/GenBank) (BENSON *et al.*, 2000) ou o SWISS-PROT (www.ebi.ac.uk/swissprot) - GASTEIGER *et al.*, 2001) ou outras ferramentas, como o "Protein Identification Resource" (PIR) (BARKER *et al.*, 2000). Se o grau de identidade entre as estruturas primárias das proteínas molde e da proteína alvo for igual ou superior a 30%, existe alta probabilidade de que estas proteínas tenham estruturas tridimensionais semelhantes, sendo possível construir um modelo para a proteína alvo.

1.5.2 Alinhamento das seqüências de resíduos de aminoácidos

O método para alinhamento de seqüências determina tipicamente a similaridade entre as seqüências utilizando uma matriz de valores para cada substituição possível durante o alinhamento. As matrizes mais utilizadas são baseadas no modelo de Dayhoff (DAYHOFF e ECK, 1968) de taxas evolucionárias e no modelo de Needleman e Wunsch de probabilidade (NEEDLEMAN e WUNSCH, 1970). O alinhamento pode ocorrer em pares seqüenciais ou entre três ou mais seqüências. A comparação entre duas seqüências recebe o nome de "alinhamento de pares", já o alinhamento entre três ou mais seqüências recebe o nome de "alinhamento múltiplo". Esses alinhamentos podem ser globais, locais ou múltiplos.



1.5.2.1 Alinhamento global

O alinhamento global considera a sequência completa de resíduos. Nesse tipo de alinhamento, as penalidades tanto para os intervalos de abertura como para os intervalos na extensão são bastante altos. Portanto, formações de blocos durante os alinhamentos são inexistentes e o que se observa são pequenas regiões ou alguns poucos intervalos espalhados ao longo da sequência, preservando o maior número possível de resíduos alinhados. Esse tipo de alinhamento é apropriado para sequências que possuem grande similaridade em todo o seu comprimento, já que o alinhamento é otimizado em toda a sua extensão (FITCH, 1966; GIBBS e MCLNTYRE, 1970).

1.5.2.2 Alinhamento local

Os alinhamentos locais podem ser representados como blocos com ausência de intervalos. A formação de blocos é facilitada pela baixa penalidade imposta aos intervalos de abertura e para os intervalos de extensão. Logo, uma sequência de resíduos poderá ter maior “mobilidade” e deslocar um grande número de resíduos através da inserção ou deleção de intervalos. Esse tipo de alinhamento é apropriado quando as sequências mostram regiões isoladas de similaridade, por exemplo, domínios múltiplos ou repetições.

1.5.2.3 Alinhamento múltiplo

Quando se dispõe de um banco de dados de sequências primárias o alinhamento múltiplo sempre será a melhor opção, já que um grande grupo de proteínas podem ser alinhadas e as regiões semelhantes ficam ainda mais destacadas. O alinhamento de três ou mais sequências primárias pode ser dividido em quatro categorias:

- Extensão de um alinhamento de pares utilizando programação dinâmica;
- Extensões hierárquicas de métodos;



- ♦ Métodos de seguimentação;

- ♦ Métodos consensuais.

O objetivo é alinhar resíduos estruturalmente equivalentes levando em conta características estruturais comuns, tais como, elementos de estrutura secundária e o sítio de ligação (BARTON, 1996).

1.5.3 Construção das coordenadas do modelo

A localização no espaço tridimensional do maior número possível de átomos da proteína alvo depende do alinhamento (PEITSCH, 1997). Uma variedade de métodos podem ser usados para construir um modelo tridimensional para a proteína alvo. Partindo do princípio de que a conformação da cadeia principal da estrutura molde pode ser transferida para ela. Existem muitos métodos de modelagem dessas regiões conservadas. Os mais importantes são:

- ♦ Modelagem pela união de corpos rígidos ("modeling by rigid-body assembly" - BLUNDELL *et al.*, 1987; GREER, 1990);
- ♦ Modelagem pela combinação de segmentos ("modeling by segment matching")(CLAESSENS *et al.*, 1989; LEVITT, 1992);
- ♦ Modelagem pela satisfação de restrições espaciais ("modeling by satisfaction of spatial restraints") (SALI e BLUNDELL, 1993; ASZODI e TAYLOR, 1996; SRINIVASAN e MARCH 1993).

O último método é o que está implementado no MODELLER para realizar a construção dos modelos.



1.5.4 Modelagem das regiões de alças

As diferenças funcionais entre membros de uma mesma família protéica são, em geral, consequência de diferenças estruturais na superfície das proteínas. Essas diferenças provêm de substituições, eliminações e inserções de resíduos nas cadeias de proteínas homólogas, principalmente nas alças, regiões mais expostas da proteína. Por serem regiões onde ocorrem variações estruturais, as alças geralmente determinam a especificidade das proteínas, contribuindo para a formação conformacional dos sítios de ligação (no caso de enzimas: sítios ativos) (FISER *et al.*, 2000). As conformações das alças devem ter baixa energia interna e não interagir desfavoravelmente com o restante da proteína. Devido a estas complicações, a modelagem de alças é, sem dúvida, a etapa mais difícil da predição de modelos de proteínas. A modelagem das alças pode ser vista como um problema de enovelamento de proteínas. As alças são geralmente pequenas e isto provoca distorções de enovelamento difíceis de avaliar. Além disso, a conformação de um dado segmento (de alça) é influenciada por toda estrutura da proteína (FISER *et al.*, 2000). Conseqüentemente, são nestas regiões que existem maiores probabilidades da ocorrência de erros na modelagem.

Os métodos de predição de alças são classificados em: métodos *ab initio* (FINE *et al.*, 1986; MOULT e JAMES, 1986; BRUCCOLERI e KARPLUS, 1987), que são baseados em uma análise conformacional que depende do ambiente estrutural e é guiada por uma função de energia e métodos comparativos, baseados na semelhança com fragmentos de estruturas conhecidas. Este método utiliza bancos de dados de estruturas de alças de proteínas armazenadas no PDB, onde é possível procurar, para cada alça da proteína alvo, estruturas que tenham as conformações energética e geometricamente mais favoráveis. Uma vez obtidos os modelos de alça, o sistema é otimizado por mecânica molecular. Esta metodologia é mais eficiente quando o número de aminoácidos das alças é pequeno, geralmente menor do que oito (GREER, 1980; JONES e THIRUP, 1986; CHOTHIA *et al.*, 1987).



É possível reduzir, em alguns casos, as limitações inerentes aos métodos comparativos pela aplicação de métodos híbridos, que combinam a abordagem por bancos de dados e os métodos *ab initio* (CHOTHIA *et al.*, 1986; MARTIN *et al.*, 1989; MAS *et al.*, 1992; VAN VLIJMEN e KARPLUS, 1997). As alças modeladas são selecionadas de acordo com critérios como a exposição de grupos hidrofóbicos na superfície e a necessidade de baixas energias conformacionais relativas (PEITSCH e CHARIFSON, 1997).

1.5.5 Modelagem das cadeias laterais

Como no caso das alças, as conformações das cadeias laterais também são modeladas a partir de dados estéricos e energéticos de estruturas semelhantes obtidas experimentalmente (VASQUEZ, 1996; SANCHEZ e SALI, 1997). A modelagem é feita com o auxílio de bancos de dados de rotâmeros permitidos de cadeias laterais (χ_n), derivados de estruturas 3D depositadas no PDB em alta resolução (PEITSCH e CHARIFSON, 1997). A figura 8 define esquematicamente os ângulos torcionais de proteínas.

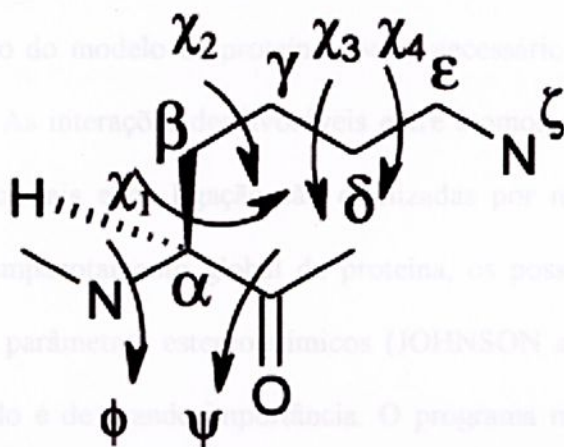


Figura 8 - Definição dos ângulos de torção da cadeia principal (Φ e Ψ) e das cadeias laterais (χ_n).

A utilização de "bibliotecas" de cadeias laterais (ou "coleções", termo menos utilizado na área) como, por exemplo, a desenvolvida por Dunbrack Jr. e colaboradores (BOWER *et al.*, 1997), é um dos métodos mais eficientes de predição da estrutura das cadeias laterais de proteínas. Elas permitem estimar a probabilidade de cada cadeia lateral assumir uma

determinada conformação (χ_n) em função dos ângulos conformacionais da cadeia principal (Φ e Ψ) e do tipo de resíduo. A precisão aproximada do método é de 82% para o ângulo diedro χ_1 e de 72% para os diedros χ_1 e χ_2 , tomados simultaneamente. Esta precisão é alcançada para pares de estruturas que compartilham de 30 a 90% de identidade.

Quando as cadeias laterais estão localizadas na superfície da proteína, deve-se incluir termos de solvatação no processo de modelagem (CREGUT *et al.*, 1994; WILSON *et al.*, 1993). Além disso, a inclusão de termos que descrevem as ligações de hidrogênio melhora significativamente a predição da estrutura das cadeias laterais (DUNBRACK e KARPLUS, 1993).

As pontes de dissulfeto é um problema especial e as cadeias laterais deste tipo são modeladas com o auxílio de informações estruturais derivadas de pontes de dissulfeto equivalentes em estruturas semelhantes (JUNG *et al.*, 1994; SALI e OVERINGTON, 1994).

1.5.6 Validação e otimização do modelo

Após a construção do modelo da proteína alvo é necessário realizar a otimização e a validação deste modelo. As interações desfavoráveis entre átomos não-ligados, bem como as energias de ângulos torcionais e de ligação são otimizadas por mecânica molecular. Deve avaliar a qualidade do empacotamento global da proteína, os possíveis erros estruturais em regiões localizadas e os parâmetros estereoquímicos (JOHNSON *et al.*, 1994). A qualidade estereoquímica do modelo é de grande importância. O programa mais utilizado para análise estereoquímica é o PROCHECK (www.biochem.ucl.ac.uk/~roman/procheck/procheck.html) (LASKOWSKI *et al.*, 1993). É possível utilizar também o programa VERIFY 3D (http://www.doe-mbi.ucla.edu/Services/Verify_3D/), que mede a compatibilidade da proteína alvo com sua sequência de aminoácidos, usando seu perfil tridimensional (BOWIE *et al.*, 1991).



Em muitos casos, a probabilidade de ocorrência de erros em um modelo de proteína construído pelo método da homologia é inversamente proporcional ao grau de identidade entre as seqüências de resíduos utilizadas. Em geral, os erros não se distribuem igualmente por toda a estrutura (D'ALFONSO *et al.*, 2001). Como um primeiro critério, os RMSD (*Root mean Squared Deviations*) das distâncias entre os átomos da cadeia principal de estruturas conservadas de pares de proteínas dependem do grau de identidade das seqüências utilizadas (CHOTHIA *et al.*, 1986). Pares de proteínas com grau de identidade superior a 50%, em geral, apresentam, valores de RMSD inferior a 1,0 Å para a proteína como um todo. Cada modelagem é um problema particular e é necessário lembrar que a construção de um modelo de proteína é uma tarefa que deve ser fundamentada em princípios de evolução molecular para cada grupo específico de proteínas homólogas.

A modelagem por homologia é uma área da bioinformática importante no campo de biologia estrutural, bioquímica e biofísica, particularmente em estudos relacionados com o genoma, permitindo acelerar o processo de elucidação de estruturas protéicas e avaliação das interações das proteínas com ligantes, em curto espaço de tempo a custos reduzidos.



OBJETIVOS

A proposta deste projeto foi modelar e analisar as estruturas dos seguintes complexos:

CDK1:Flavopiridol, CDK1:H717, CDK1:Purvalanol, CDK1:Estaurosporina, CDK1:Indirubina-5-Sulfonato, CDK1:Oxindol 16 (composto3), CDK1:Roscovitina, CDK1:Isopenteniladenina e CDK1:Olomoucina, utilizando os complexos de CDK2 com os mesmos inibidores como molde. Obter informações da interação dos inibidores propostos com a proteína em estudo e estabelecer bases estruturais para a inibição da CDK1, que permitirá o desenvolvimento de inibidores mais específicos.

3.1 Modelagem dos complexos

Na interface do programa PARMODEL foram fornecidos como arquivos de entrada o arquivo contendo a estrutura primária da CDK1 e o arquivo no formato pdb do complexo CDK2:Indol. Para cada sequência primária da CDK1 e cada conjunto de coordenadas atômicas (arquivo .pdb), foi determinado que deveriam ser construído 1000 modelos. A construção de vários modelos aumenta a possibilidade de se obterem estruturas com conformações espaciais diferentes, pois a restrição espacial imposta pelo molde CDK2:Indol é relaxada para cada um dos modelos. Assim, a partir de uma única sequência primária da CDK1, é possível obter uma série de estruturas com conformações espaciais diferentes, o que permite a análise da interação dos inibidores com a proteína em estudo.



MÉTODOS

Para realizar a modelagem por homologia dos complexos da CDK1 com diferentes inibidores foi usado o programa PARMODEL (Apêndice B) (UCHÔA *et al.*, 2004). O PARMODEL distribui seus trabalhos em *cluster Beowulf* de 16 nós em ambiente Linux, objetivando maior rapidez nas respostas. Ele utiliza uma biblioteca em linguagem C, MPI (*Message Passing Interface*), que controla a distribuição dos trabalhos para o *cluster*, permitindo a paralelização da execução do MODELLER (Apêndice A) e diminuindo o tempo de processamento do processo de modelagem. O PARMODEL foi projetado para integrar os principais programas usados no processo de modelagem por homologia, é subdividido em quatro módulos: modelagem, avaliação, otimização e visualização. O PARMODEL foi utilizado para a realização das modelagens e análises de resultados.

3.1 Modelagem dos complexos

Na interface do programa PARMODEL, foram fornecidos como arquivos de entrada, o arquivo contendo a estrutura primária da CDK1 e o arquivo no formato pdb do complexo CDK2:Inibidor. Para cada sequência primária da CDK1 e cada conjunto de coordenadas atômicas (arquivo .pdb), foi determinado que deveria ser construído 1000 modelos. A construção de vários modelos aumenta a possibilidade de obter boas estruturas porque as mudanças nas restrições espaciais iniciais podem gerar diferentes conformações na estrutura final. Foram enviados por endereço eletrônico um arquivo anexado compactado com a estrutura dos dez melhores modelos, o resultado dos alinhamentos, uma lista com a função objetiva do



MODELLER para todos modelos construídos e os resultados das análises para os melhores modelos. Esses modelos foram classificados com base na função objetiva do MODELLER que tinha o valor mais baixo. O valor da função objetiva do MODELLER não é uma medida absoluta, podendo ser usado somente para classificar modelos construídos para o mesmo alinhamento (SALI e BLUNDELL, 1993). Para realizar o alinhamento o PARMODEL usa o alinhamento perfil-perfil do comando SALIGN do MODELLER-6v2 (OHLSON *et al.*, 2004). O método de alinhamento perfil-perfil é ao menos 30% melhor do que o método padrão de alinhamento da sequência-perfil, ambos em sua habilidade de reconhecer super famílias de proteínas relacionadas e na qualidade dos alinhamentos obtidos (HOOFT *et al.*, 1996). Atualmente, o PARMODEL retorna os resultados das modelagens, contendo um arquivo anexado compactado com a estrutura do melhor modelo, o resultado do alinhamento, uma lista com a função objetiva do MODELLER para todos modelos construídos e os resultados das análises para o melhor modelo.

3.2 Análises dos modelos

As análises das qualidades estereoquímicas dos modelos finais dos complexos CDK1:Inibidores e das estruturas usadas como moldes foram realizadas com a execução do programa PROCHECK (LASKOWSKI *et al.*, 1993). O PROCHECK realiza uma verificação detalhada na estereoquímica da estrutura da proteína. Os arquivos de entrada do programa foram os arquivos de coordenadas 3D (formato .pdb) de cada complexo. Seus arquivos de saída compreendem vários arquivos no formato postScript (.ps) e uma lista detalhada de resíduo-por-resíduo. O PROCHECK faz uma avaliação da qualidade total da estrutura, utilizando parâmetros estereoquímicos derivados de estruturas resolvidas em alta resolução, destacam também as regiões que podem necessitar de uma investigação adicional. Por exemplo a geometria covalente, planaridade (cheça a planaridade de anéis aromáticos e grupos terminais através do



valor do RMS), ângulos diédricos, que é a distribuição através dos ângulos ϕ e ψ formados entre o carbono α e o nitrogênio (C_{α} -N) e entre o carbono e o carbono α (C - C_{α}), respectivamente (figura 9). Como cada padrão de estrutura secundária possuem faixas de valores preferenciais ou estereoquimicamente permitidos para esses ângulos, ocorre a formação de regiões específicas no diagrama onde os resíduos tendem a aparecer. Os resíduos podem ser localizados em uma das seguintes regiões: permitida (*Most Favorable*), adicionalmente permitida (*Additional Allowed*), generosamente permitida (*Generously Allowed*) e não permitida (*Disallowed*). Este programa é fundamental na análise dos modelos gerados e para uma melhor visualização da posição dos resíduos com relação ao diagrama de Ramachandran, determinando o percentual dos resíduos que estão em regiões permitidas (RAMAKRISHNAN e RAMACHANDRAN, 1965).



Figura 9 - Ângulos de torção de cadeia polipeptídica, de rotação repetida no plano dos ossos. Os ângulos ϕ e ψ são os ângulos de torção em torno da ligação C_{α} -N (ϕ) e da ligação C - C_{α} (ψ). Por convenção, ϕ e ψ são 180° na conformação estendida e alternam, como indicado, na direção horária quando visto a partir do C_{α} .

Com a execução do programa PROCHECK também foi possível obter o G-fator, que fornece a medida da desvio de ângulos torcionais, distâncias de ligações covalentes e ângulos da cadeia principal. O G-fator é essencialmente uma contagem das probabilidades baseadas nas distribuições observadas de ângulos torcionais estereocorrelacionados. Quando aplicado a um dado resíduo, um valor de G-fator baixo indica que a propriedade corresponde a uma baixa probabilidade conformacional. Os resíduos que caem nas regiões não permitidas do diagrama

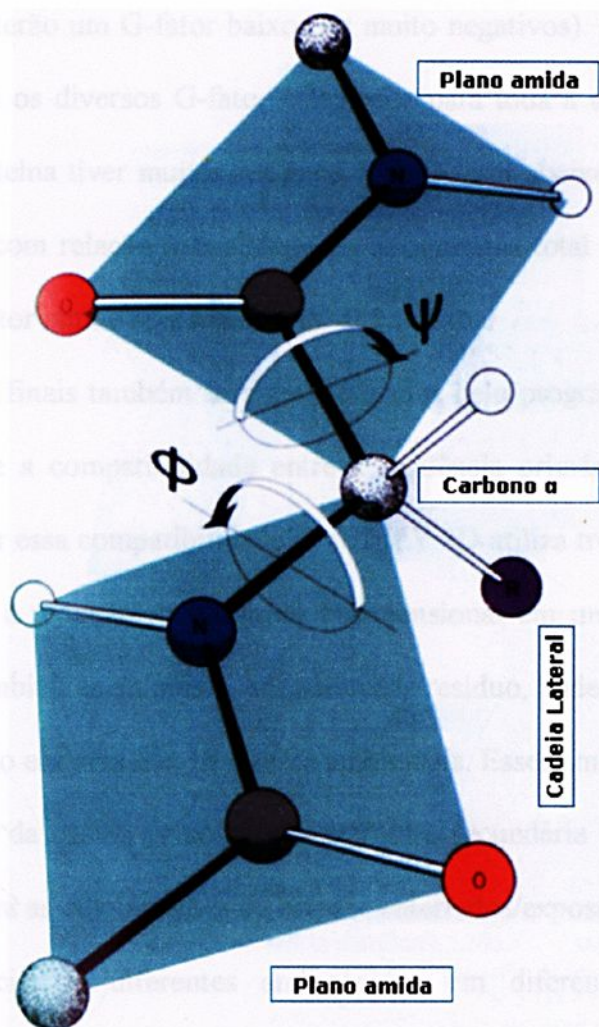


Figura 9 - Ângulos de torção da cadeia polipeptídica, dois grupos peptídicos planares são mostrados. Os únicos movimentos razoavelmente livres são a rotação em torno da ligação C_{α} -N (ϕ) e da ligação C- C_{α} (ψ). Por convenção, ϕ e ψ têm 180° na conformação apresentada e aumenta, como indicado, na direção horária quando visto a partir do C_{α} .

Com a execução do programa PROCHECK também foi possível obter o G-fator, que fornece a medida do desvio de ângulos torcionais, distâncias de ligações covalentes e ângulos da cadeia principal. O G-fator é essencialmente uma contagem das probabilidades baseadas nas distribuições observadas destes parâmetros estereoquímicos. Quando aplicado a um dado resíduo, um valor de G-fator baixo indica que a propriedade corresponde a uma baixa probabilidade conformacional. Os resíduos que caem nas regiões não permitidas do diagrama

de Ramachandran terão um G-fator baixo (ou muito negativos). É possível calcular o G-fator global que envolve os diversos G-fator calculados para toda a estrutura, num único número. Assim, se uma proteína tiver muitos resíduos com G-fator abaixo de $-1,00$, este valor indica que existem erros com relação a qualidade estereoquímica total do modelo obtido. O ideal é que o valor do G-fator obtido seja maior que $-0,5$.

Os modelos finais também foram examinados pelo programa VERIFY 3D (BOWIE *et al.*, 1991) que mede a compatibilidade entre a sequência primária e a estrutura 3D de uma proteína. Para medir essa compatibilidade o VERIFY 3D utiliza três operações:

- É realizada a redução da estrutura tridimensional em uma sequência unidimensional dentro de ambientes químicos, um para cada resíduo, onde cada resíduo na estrutura 3D é classificado em uma das 18 classes ambientais. Esses ambientes são categorizados por três estados da cadeia principal, da estrutura secundária e por seis estados da cadeia principal para as combinações de estados enterrados/exposto e polar/apolar.
- A preferência de diferentes aminoácidos em diferentes classes ambientais são empiricamente determinadas e representadas por uma matriz de comparação (matriz 3D-1D - BOWIE *et al.*, 1991) que descreve a probabilidade de se encontrar cada um dos 20 aminoácidos em cada uma das 18 classes de ambientes químicos. O VERIFY 3D determina a probabilidade, por resíduo, de acordo com o tipo de aminoácidos e a natureza do ambiente químico calculado no passo anterior. A matriz 3D-1D é previamente conhecida e derivada de uma análise de estruturas conhecidas de boa qualidade. O resultado é uma medida da compatibilidade entre a sequência primária e sua estrutura 3D descrita pelo seu perfil tridimensional (LUTHY *et al.*, 1992). É determinado empiricamente o índice global esperado de $S_{ideal} = \exp(-0,83 + 1,008 \times \ln(L))$ de compatibilidade entre a sequência e a estrutura 3D. Valores próximos de $0,45S_{ideal}$ indica que a qualidade do modelo é questionável. Se os valores forem menores



que 0,45 S_{ideal} , esses valores indicam que a estrutura é incorreta. Para os modelos corretos os valores considerados ideais são os valores acima de 0,45 S_{ideal} .

- Utilizando a Matriz 3D-1D a estrutura da proteína pode ser convertida para o perfil tridimensional. É realizado um alinhamento da sequência primária com o perfil tridimensional, onde a qualidade do alinhamento é a medida da compatibilidade da sequência com sua estrutura 3D descrita por seu perfil tridimensional (LUTHY *et al.*, 1992).

O escore total é a soma de todos 3D-1D escores que são as preferências estatísticas de cada resíduo presente na proteína. Esse cálculo pode ser representado pela seguinte expressão:

$$\text{Escore total} = \sum_{ij} N_{ij} \ln\left(\frac{p(i:j)}{p_i}\right)$$

onde N_{ij} é o número de resíduos i no ambiente j , $p(i:j)$ é a probabilidade de encontrar um resíduo i em um ambiente j e p_i é a probabilidade total de encontrar um resíduo i em um ambiente (BROWIE *et al.*, 1991). O escore ideal representa a soma das probabilidades individuais dos resíduos, que é calculado utilizando a expressão abaixo.

$$S_{ideal} = e^{(-0,83+1,008 \times \ln(L))}$$

onde L é o número de aminoácidos, e finalmente o S_{ideal} escore que descreve a compatibilidade da sequência com sua estrutura 3D, pode ser obtido pela razão do escore total com escore Ideal, logo:

$$S_{ideal} \text{Score} = \frac{\text{Total Score}}{\text{Ideal Score}}$$

Os valores Considerados ideais são aqueles que apresentam resultados acima de 0,45

S_{ideal} .



O X-PLOR (<http://xplor.csb.yale.edu/xplor/download/distributions/>) é um programa para refinamento de estruturas cristalográficas ou resolvidas por RMN. Ele possui várias rotinas de análise estrutural (Seu foco principal é a determinação da estrutura tridimensional de uma macromolécula usando dados de difração de raios X ou RMN). O programa é baseado em uma aproximação da função de energia: combinações arbitrárias empíricas, geométricas e termos de energia efetiva descritas experimentalmente podem ser usados. A função energia combinada pode ser minimizada por uma variedade de gradiente descendente, *Simulate Annealing* e procedimento de busca conformacional. O X-PLOR foi o primeiro programa a combinar dados de difração de Raios X e refinamento por dinâmica molecular. Desde então, o programa foi submetido a um desenvolvimento intenso, e o foco deslocou do refinamento para a determinação de estrutura. A rotina utilizada deste programa foi a que determina o RMSD dos modelos em relação à geometria ideal (ENGH e HUBER, 1991). Tal avaliação fornece um parâmetro indicativo da proximidade dos valores de distância de ligações, o ângulo de ligação e de torção em relação aos valores esperados (BRUNGER, 1992).

3.3 Análises das interações do inibidor com a proteína

A especificidade e afinidade entre uma enzima e seu inibidor dependem de interações iônicas, ligações de hidrogênio, interações eletrostáticas, interações hidrofóbicas, enfim, complementaridade das suas superfícies de contato (CANDURI *et al.*, 2001; DE AZEVEDO JR. *et al.*, 2001; KIM *et al.*, 1996). Para verificar a existência dessas interações e para analisar as interações existentes foram executados várias rotinas e programas. As rotinas e os programas utilizados foram:

- Para analisar as interações atômicas quanto à complementaridade de forma e complementaridade eletrostática das cargas na superfície da proteína, foi executado o programa GRASP (NICHOLLS *et al.*, 1991).



- O programa XtalView foi utilizado para verificar a existência e a distância das ligações de hidrogênio dos inibidores com a proteína alvo e a proteína molde.
- Para verificar os valores da área de contato entre o inibidor e a proteína em estudos, foram executadas as rotinas areaimol, diffarea e resarea, do pacote CCP4 (CCP4, 1994).
- Os programas mais utilizados nos processos de alinhamentos são o BLAST (ALTSCHUL *et al.*, 1997 - <http://www.ncbi.nlm.nih.gov/blast/bl2seq/bl2.html>), o FastA (PEARSON, 1990 - http://fasta.bioch.virginia.edu/fasta/fasta_list.html), o CLUSTAL (JEANMOUGIN *et al.*, 1998 - <http://dot.imgen.bcm.tmc.edu/multi-align/Help/clustalw.html>), e o MULTALIN (CORPET, 1988 - <http://prodes.toulouse.inra.fr/multalin/multalin.html>). O alinhamento das seqüências primárias da CDK1 e CDK2 foi realizado no MULTALIN.
- O programa MOLSCRIPT (KRAULIS, 1991) foi utilizado para gerar uma representação dos elementos de estrutura secundária dos complexos em estudo.

RESULTADOS E DISCUSSÕES

A análise da qualidade estereoquímica dos modelos finais dos complexos CDK1:Inibidores obtidos e da estrutura utilizada como molde foram executadas com base nos diagramas de Ramachandran, que é um dos resultados fornecidos pelo PROCHECK. No diagrama de Ramachandran pode ser visto a distribuição dos ângulos phi e psi dos resíduos pertencentes a uma determinada estrutura. A glicina é o único resíduo sem C_{β} , é menos limitado do que os outros resíduos de aminoácidos. Enquanto que, a cadeia lateral cíclica da prolina limita sua variação dos valores de phi a ângulos de -60° , tornando-a o resíduo de aminoácido mais restringido conformacionalmente. Com base nessas restrições no diagrama de Ramachandran esses resíduos de aminoácidos são representados por triângulos.

Nas figuras 10a-f, 11a-f e 12a-f são apresentados os diagramas de Ramachandran dos modelos obtidos dos complexos CDK1:Inibidor e dos complexos CDK2:Inibidor. Também é apresentado na tabela 2 o resultado das análises de qualidade estereoquímica para cada complexo. É possível fazer uma avaliação da qualidade média das estruturas em questão. Todos os complexos em estudos exceto o CDK1:purvalanol (que apresenta o mesmo número) apresentam maior número de resíduos do que seu molde na região permitida. Foi possível verificar que os modelos finais dos complexos CDK1:Inibidor em análise são apropriados. Foi observado também a existência de alguns resíduos nas regiões não permitidas. Para esses resíduos, foi verificado através do programa XtalView, que nenhum deles apresentam choques estereoquímicos, por ser resíduos de superfície ou estar muito enterrado na proteína.



Tabela 2 - Resultados das análises de qualidade estereoquímica.

Complexo	Regiões do Diagrama de Ramachandran (%)			
	Mais Favorável	Adicionalmente Permitida	Generosamente Permitida	Não Permitida
CDK1:Flavopiridol	86,3(79,7)	10,3(14,2)	2,7(4,2)	0,8(1,9)
CDK1:H717	89,7(86,6)	8,0(11,8)	1,1(1,3)	1,1(0,4)
CDK1:Purvalanol	88,8(88,8)	10,4(10,4)	0,4(0,8)	0,4(0,0)
CDK1:Estaurosporina	89,7(88,3)	9,3(11,3)	0,4(0,4)	0,4(0,0)
CDK1:Indirubina-5-Sulfonato	86,3(82,0)	7,6(11,8)	1,5(1,2)	0,8(0,8)
CDK1:Oxindol-16	93,9(88,1)	5,3(11,1)	0,4(0,0)	0,4(0,8)
CDK1:Roscovitina	90,8(87,1)	8,0(12,9)	0,8(0,0)	0,4(0,0)
CDK1:Isopenteniladenina	93,1(89,3)	4,6(8,4)	2,3(1,9)	0,0(0,4)
CDK:Olomoucina	89,7(87,9)	9,54(10,9)	0,4(0,8)	0,4(0,4)

❖ Os valores apresentados dentro dos parênteses são dos moldes.

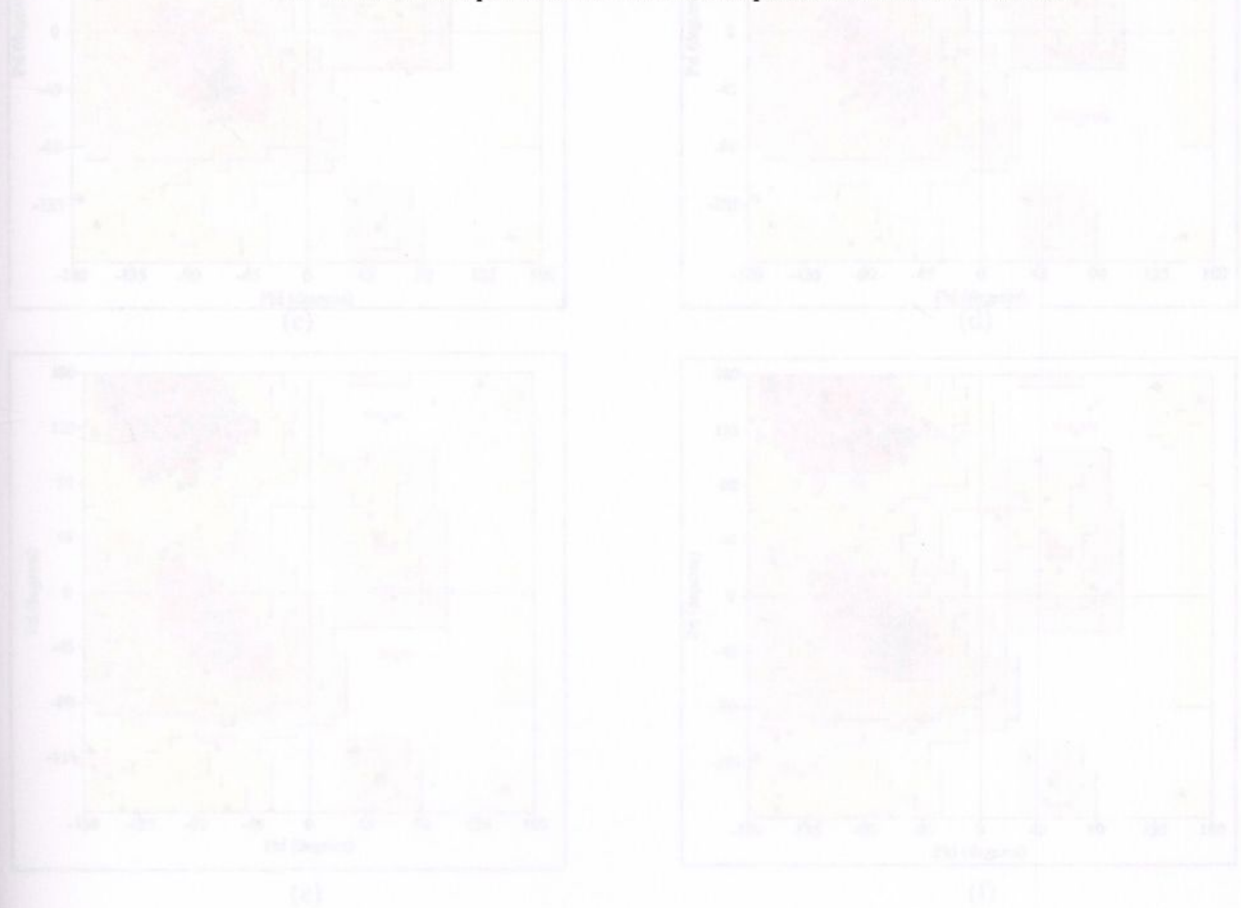
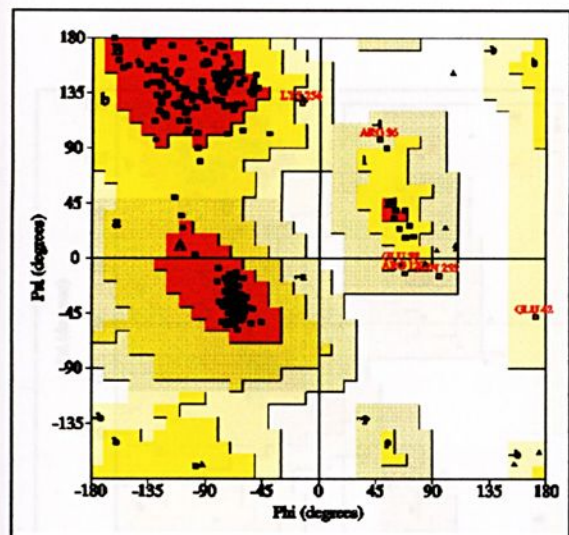
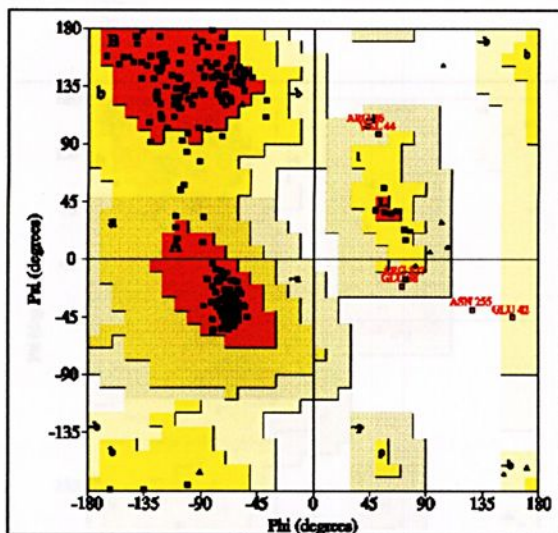


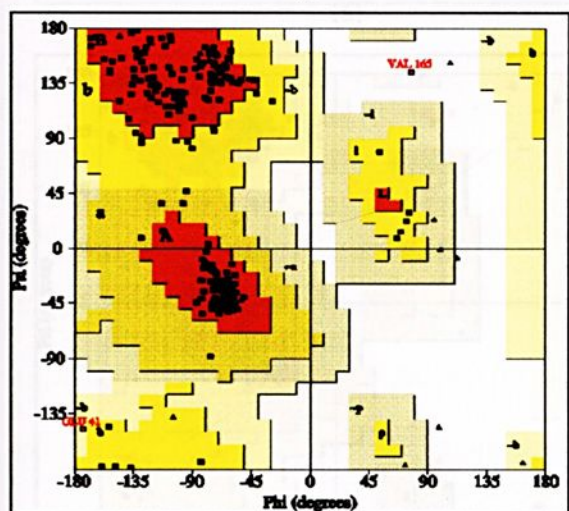
Figura 10 - Nos 21 graus de liberdade do diagrama de Ramachandran as regiões verde, amarela, cor-de-rosa e branca representam respectivamente as regiões permitidas, adicionalmente permitidas, generosamente permitidas e não permitidas. (a) complexo CDK1:Flavopiridol, (b) complexo CDK1:H717, (c) complexo CDK1:Purvalanol, (d) complexo CDK1:Estaurosporina, (e) complexo CDK1:Indirubina-5-sulfonato, (f) complexo CDK1:Roscovitina, (g) complexo CDK1:Isopenteniladenina, (h) complexo CDK1:Olomoucina.



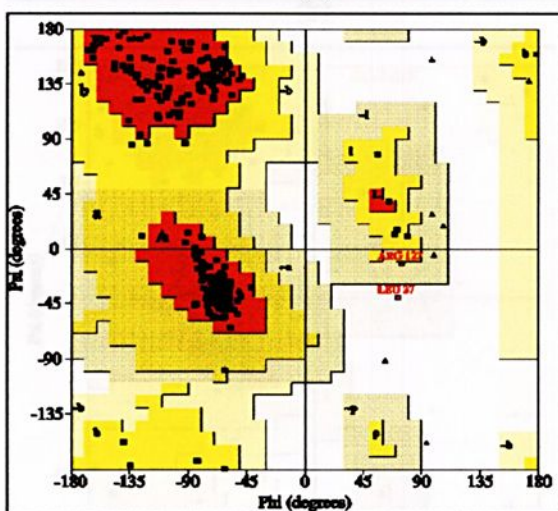
(a)



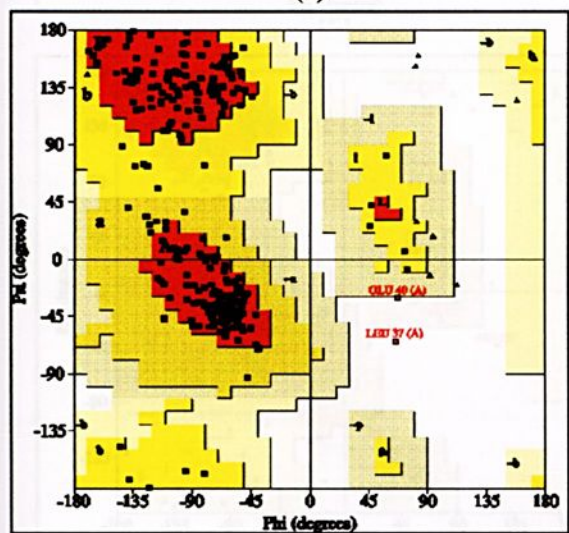
(b)



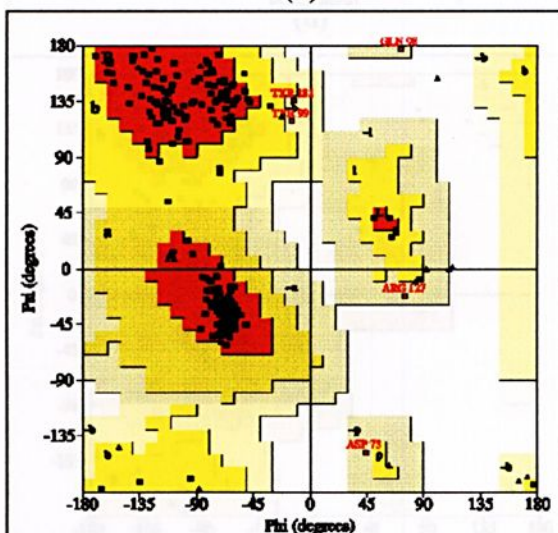
(c)



(d)

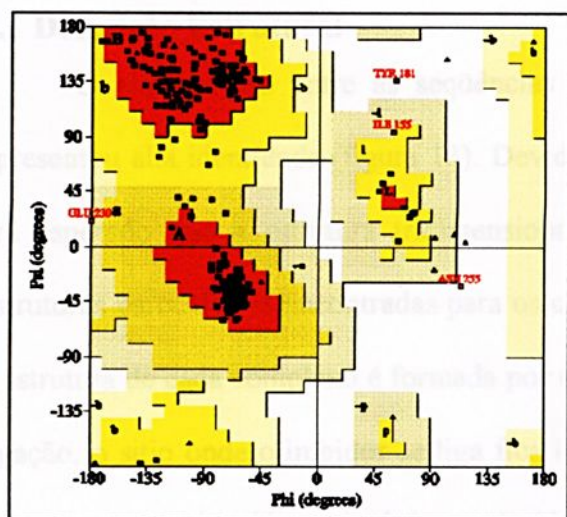


(e)

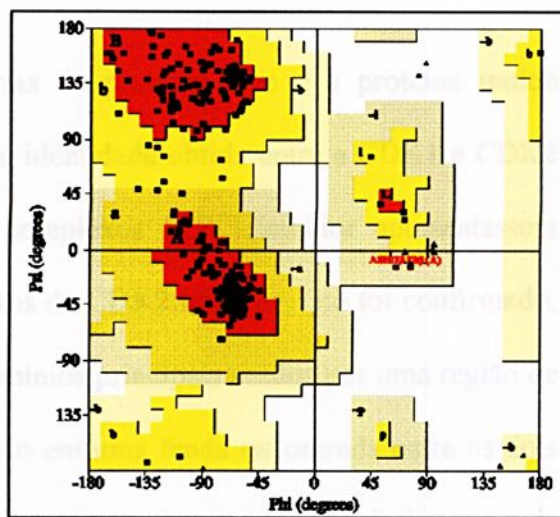


(f)

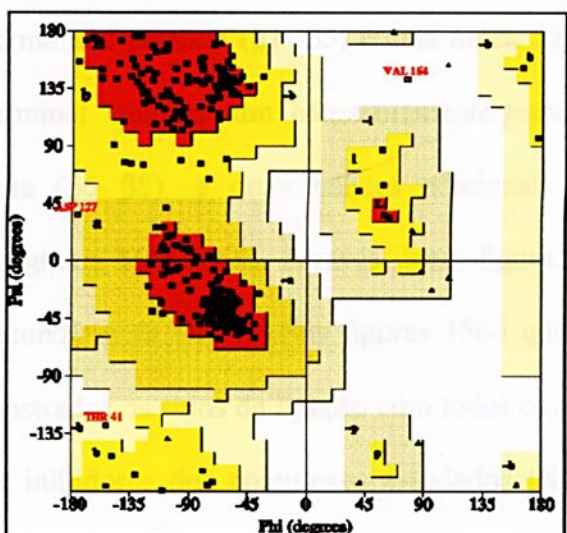
Figura 11 - Nos diagramas de Ramachandran as regiões vermelhas, amarelas, amarelas claro e brancas, representam respectivamente as regiões permitida, adicionalmente permitida, generosamente permitida e não permitida. (a) complexo CDK1:Isopenteniladenina (b) complexo CDK2:Isopenteniladenina (c) complexo CDK1:Olomoucina (d) complexo CDK2:Olomoucina (e) complexo CDK1:Oxindol-16 e (f) complexo CDK2:Oxindol-16.



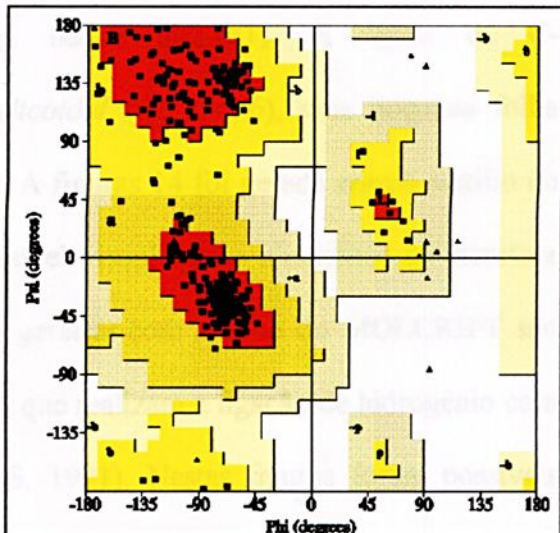
(a)



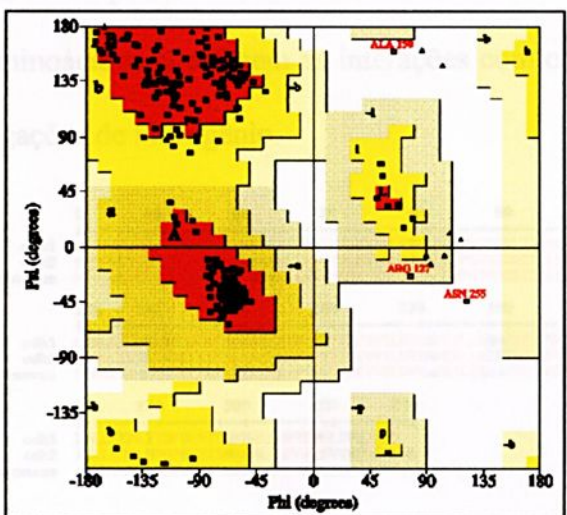
(b)



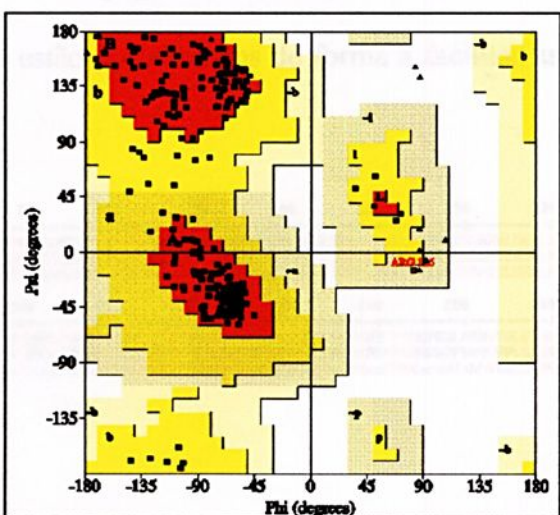
(c)



(d)



(e)



(f)

Figura 12 - Nos diagramas de Ramachandran as regiões vermelhas, amarelas, amarelas claro e brancas, representam respectivamente as regiões permitida, adicionalmente permitida, generosamente permitida e não permitida. (a) complexo CDK1:Purvalanol (b) complexo CDK2:Purvalanol (c) complexo CDK1:Roscovitina (d) complexo CDK2:Roscovitina (e) complexo CDK1:Staurosporina e (f) complexo CDK2:Staurosporina.

4.1 Descrição Estrutural

O alinhamento entre as seqüências primárias da proteína alvo e a proteína molde apresentou alta identidade (figura 13). Devido a alta identidade obtida entre a CDK1 e CDK2 era esperado que a estrutura tridimensional dos complexos CDK1:Inibidor apresentassem estruturas parecidas às encontradas para os complexos de CDK2:Inibidor. Isto foi confirmado, a estrutura de cada complexo é formada por dois domínios principais unidos por uma região de ligação, o sitio onde o inibidor se liga fica localizado em uma fenda encontrada entre os dois domínios. A região N-terminal é constituída de um conjunto de cinco folhas β dispostas de forma anti-paralela (β 1- β 5) e uma única e longa hélice alfa (α 1). A região de C-terminal contém um empacotamento *pseudo-4-helicoidal* (α 2, 3, 4, 6), uma pequena folha beta (β 6- β 8) e duas hélices adicionais (α 5, 7). A figuras 14 foi gerada com o auxilio do programa MOLMOL. Através dessa figura foi possível visualizar os elementos de estrutura secundária da CDK1. Nas figuras 15a-i que foram geradas com auxilio do MOLCRIPT são mostrados os sítios de ligação com todos os resíduos que realizam a ligação de hidrogênio com os inibidores dos complexos estudados (KRAULIS, 1991). Nestas figuras foram possíveis verificar que os inibidores estão localizados nos sítios de ligação do ATP e que os resíduos de aminoácidos que fazem as interações com o ligante estão posicionados de forma a facilitar as ligações de hidrogênio.

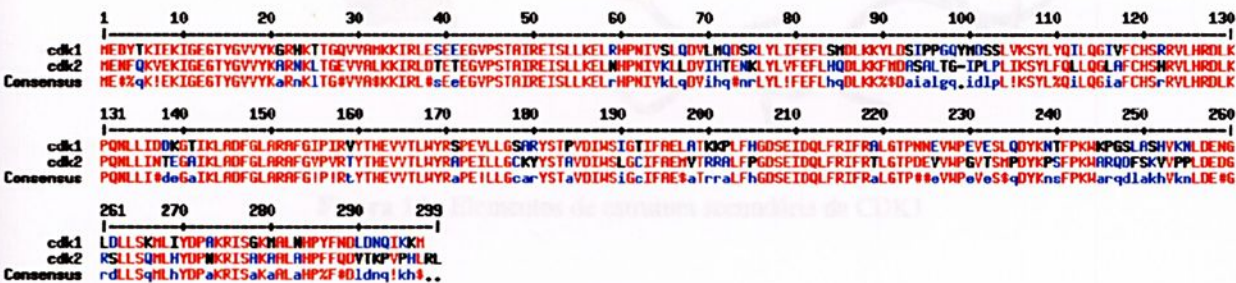


Figura 13 - Alinhamento das seqüências da CDK1 e CDK2. Há 67,34% de identidade entre as seqüências. Os resíduos idênticos ou conservados são representados em vermelho, já os resíduos que não apresentam nenhuma identidade estão em azul e pretos

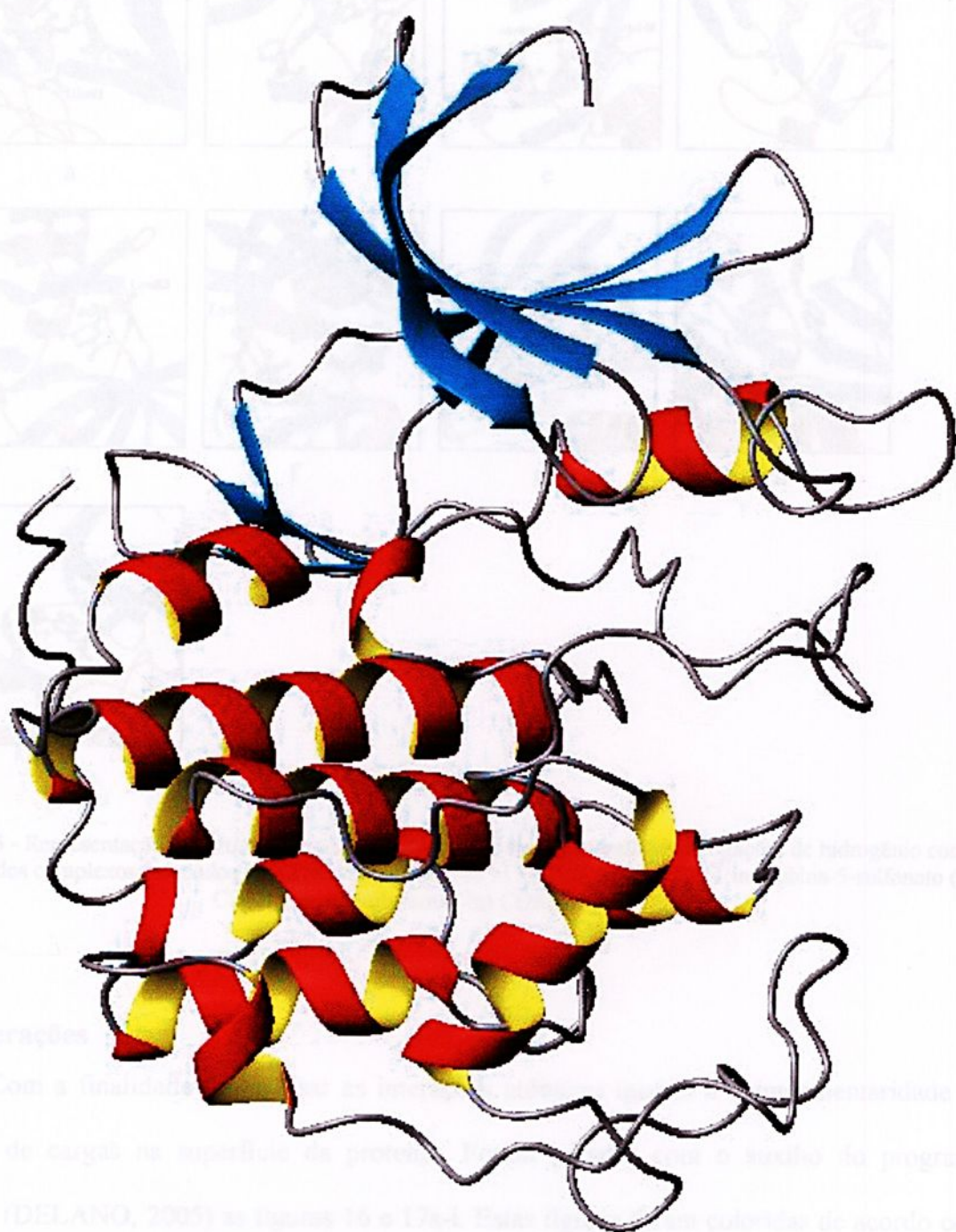


Figura 14 - Elementos de estrutura secundária da CDK1

CDK1. Os inibidores estão ligados ao bolsão de ligação que fica localizado na região da fenda entre os domínios N-terminal e C-terminal da CDK1 (figura 16). Não foram observadas existências de choques entre as cadeias laterais das moléculas.

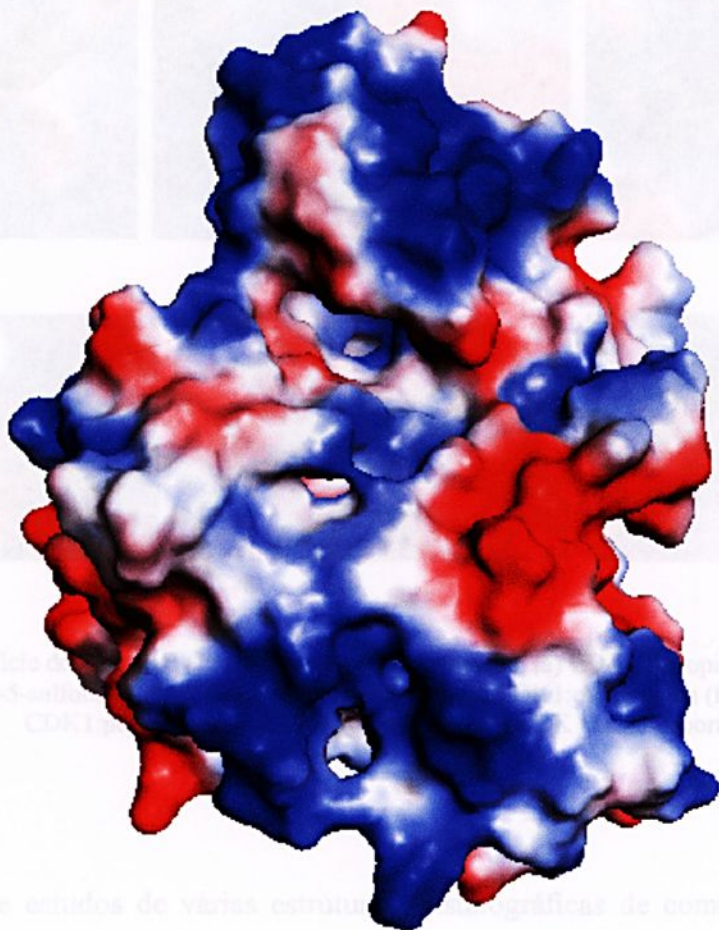


Figura 16 - Superfície do potencial eletrostático. A região em azul representa os resíduos carregados positivamente, em vermelho são os carregados negativamente e em branco os resíduos neutros.

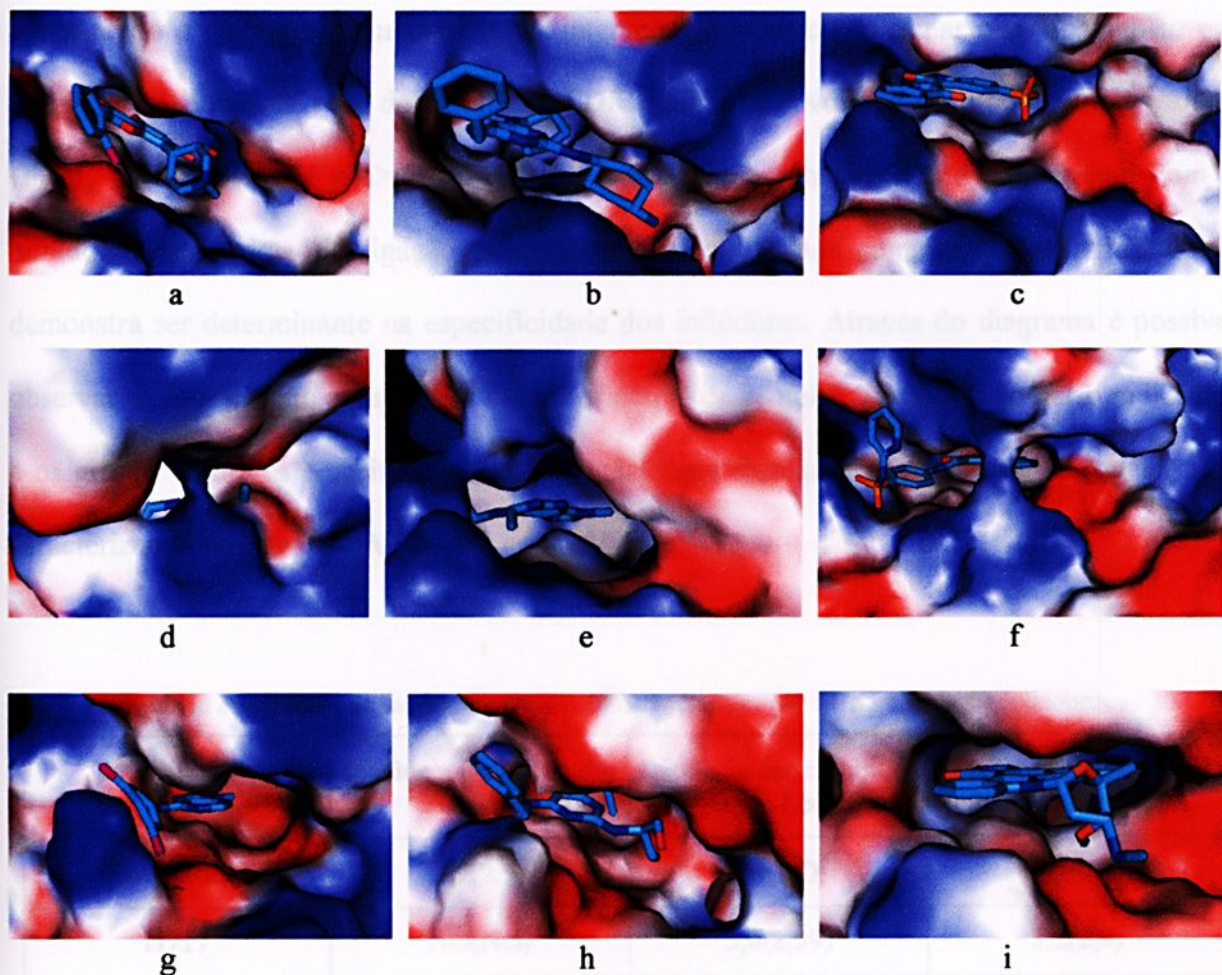


Figura 17 - Superfície do potencial eletrostático do sítio de ligação (a) CDK1:flavopiridol (b) CDK1:h717 (c) CDK1:indirubina-5-sulfonato (d) CDK1:isopenteniladenina (e) CDK1:olomoucina (f) CDK1:oxindol-16 (g) CDK1:purvalanol (h) CDK1:roscovitina e (i) CDK1:estaurosporina.

Através de estudos de várias estruturas cristalográficas de complexos formados pela proteína CDK2 e diferentes inibidores foi possível observar um padrão de interação conservado e apresenta ser determinante no reconhecimento de pequenas moléculas pelo sítio de ligação do ATP para essa classe de enzimas. Este padrão de ligação foi denominado Garfo Molecular (*Molecular fork* - DE AZEVEDO JR. *et al.*, 2002a). Na tabela 3 é descrito o padrão de ligação de hidrogênio que forma o Garfo Molecular, identificado no complexo CDK2:Inibidor (DE AZEVEDO JR. *et al.*, 2002b). Também são descritas as ligações de hidrogênio existentes entre os inibidores e os resíduos que caracterizam o garfo molecular da CDK1. Na figura 18 é

apresentado o diagrama esquemático das interações do Garfo Molecular que é composto por três ligações e que no caso do complexo CDK2:Inibidor, envolvem o grupo C=O do resíduo Glu81, os grupos N-H e C=O do resíduo Leu83 formando ligações de hidrogênio com o inibidor. Esse padrão de ligação também foi observado nos complexos CDK1:Inibidor e demonstra ser determinante na especificidade dos inibidores. Através do diagrama é possível observar os átomos que compõem a cadeia principal da proteína, na região do garfo molecular e os grupos doadores e aceptores de hidrogênio que formam as ligações de hidrogênio que caracterizam a interação (CANDURI. *et al.*, 2004).

Tabela 3 - ligações de hidrogênio entre o Garfo Molecular das CDK1 e CDK2 e Inibidores.

Inibidor	Distância entre inibidor e C=O em Glu81 (Å)	Distância entre inibidor e C=O em Leu83 (Å)	Distância entre inibidor e N-H em Leu83 (Å)
Flavopiridol	2,9(2,9)	2,8(NO)	NO(2,7)
H717	NO(NO)	3,0(2,99)	3,2(2,9)
Purvalanol	NO(NO)	3,2(2,7)	2,6(NO)
Estaurosporina	2,7(2,8)	NO(NO)	2,7(2,6)
Indirubina-5-Sulfonato	3,0(3,0)	3,1(2,8)	2,7(3,1)
Oxindol-16	3,0(2,98)	3,0(NO)	2,3(2,9)
Roscovitina	NO(NO)	3,4(2,9)	3,0(3,4)
Isopenteniladenina	2,8(NO)	NO(3,4)	NO(NO)
Olomoucina	NO(NO)	2,4(2,0)	NO(2,3)

❖ Os valores apresentados dentro dos parênteses são dos moldes.
❖ NO: Não Observado



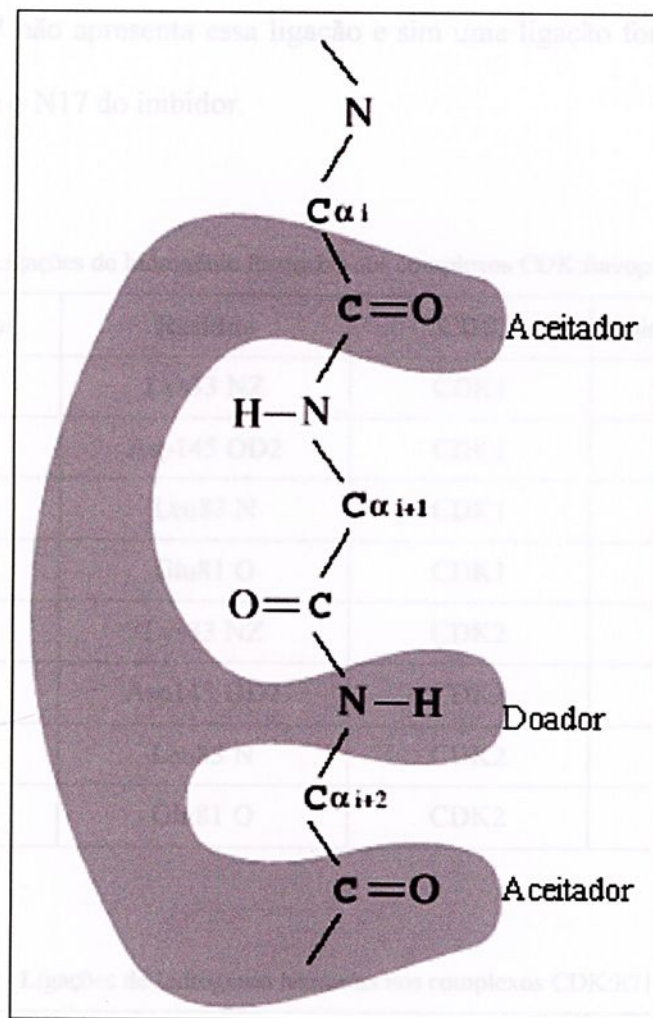


Figura 18 - Doadores e aceptores de hidrogênio da CDK.

A especificidade e a afinidade entre uma enzima e seu inibidor dependem de interações iônicas, ligação de hidrogênio, interações eletrostáticas, interações hidrofóbicas, e de complementaridade das suas superfícies de contato (CANDURI *et al.*, 2001; DE AZEVEDO JR. *et al.*, 2001; KIM *et al.*, 1996). Nas tabelas 4 – 12, estão relacionados todos os resíduos que realizam as ligações de hidrogênio e os valores das distâncias. É possível notar, que praticamente não houve mudanças dos complexos de CDK2:Inibidores para os complexos CDK1:Inibidores, apresentando uma discreta variação na distância entre os resíduos que fazem as ligações de hidrogênio do complexo modelo e os resíduos que fazem as ligações de hidrogênio do complexo molde. Foi verificado também que o complexo CDK1:H717 apresenta o resíduo Asn133 OD1 da CDK1 fazendo uma ligação de hidrogênio com o N17 do inibidor, já

o complexo CDK2:H717 não apresenta essa ligação e sim uma ligação formada entre resíduo Asp132 OD1 da CDK2 e o N17 do inibidor.

Tabela 4 - Ligações de hidrogênio formadas nos complexos CDK:flavopiridol.

Flavopiridol	Resíduo	CDK	Distância (Å)
O3	Lys33 NZ	CDK1	3,2
O3	Asp145 OD2	CDK1	3,2
O4	Leu83 N	CDK1	2,8
O5	Glu81 O	CDK1	2,9
O3	Lys33 NZ	CDK2	3,0
O3	Asp145 OD2	CDK2	3,2
O4	Leu83 N	CDK2	2,8
O5	Glu81 O	CDK2	2,9

Tabela 5 - Ligações de hidrogênio formadas nos complexos CDK:h717.

H717	Resíduo	CDK	Distância (Å)
N17	Asn133 OD1	CDK1	2,9
N17	Asp146 OD2	CDK1	2,7
N7	Leu83 N	CDK1	3,2
N18	Leu83 O	CDK1	3,0
N17	Asp132 OD1	CDK2	2,9
N17	Asp145 OD1	CDK2	2,7
N7	Leu83 N	CDK2	3,2
N18	Leu83 O	CDK2	3,0



Tabela 6 - Ligações de hidrogênio formadas nos complexos CDK:indirubina-5-sulfonato.

Indirubina-5-Sulfonato	Resíduo	CDK	Distância (Å)
O23	Asp146 N	CDK1	3,0
N14	Glu81 O	CDK1	3,0
O13	Leu83 N	CDK1	2,7
N7	Leu83 O	CDK1	3,1
O24	Lys33 NZ	CDK1	2,9
O22	Lys33 NZ	CDK1	3,2
O24	Asp145 N	CDK2	3,3
O23	Asp146 N	CDK2	3,0
N14	Glu81 O	CDK2	3,0
O13	Leu83 N	CDK2	2,7
O24	Lys33 NZ	CDK2	2,9
O22	Lys33 NZ	CDK2	3,2

Tabela 7 - Ligações de hidrogênio formadas nos complexos CDK:olomoucina.

Olomoucina	Resíduo	CDK	Distância (Å)
N6	Leu83 N	CDK1	2,4
N7	Leu83 N	CDK1	2,5
N6	Ser84 O	CDK1	3,5
N6	Leu83 O	CDK2	2,0
N7	Leu83 O	CDK2	3,2
N6	His 84 O	CDK2	3,3

Tabela 8 - Ligações de hidrogênio formadas nos complexos CDK:oxindol-16.

Oxindol-16	Resíduo	CDK	Distância (Å)
O24	Asp 86 OD2	CDK1	3,5
O24	Asp 86 N	CDK1	3,2
N3	Glu81 O	CDK1	3,0
O1	Leu83 N	CDK1	3,0
N8	Lys33 NZ	CDK1	3,2
O23	Lys89 NZ	CDK1	3,1
O24	Asp 86 N	CDK2	3,2
O24	Asp 86 OD2	CDK2	3,5
N3	Glu81 O	CDK2	3,0
O1	Leu83 N	CDK2	3,0
O23	Lys89 NZ	CDK2	3,1

Tabela 9 - Ligações de hidrogênio formadas nos complexos CDK:purvalanol.

Purvalanol	Resíduo	CDK	Distância (Å)
N6	Leu83 O	CDK1	2,6
N7	Leu83 N	CDK1	3,2
N6	Leu83 O	CDK2	2,8

Tabela 10 - Ligações de hidrogênio formadas nos complexos CDK:roscovitina.

Roscovitina	Resíduo, átomo	CDK	Distância (Å)
N6	Leu83 O	CDK1	2,9
N7	Leu83 N	CDK1	3,5
N6	Leu83 O	CDK2	2,9
N7	Leu83 N	CDK2	3,4



Tabela 11 - Ligações de hidrogênio formadas nos complexos CDK:estaurosporina.

Estaurosporina	Resíduo	CDK	Distância (Å)
N1	Glu81 O	CDK1	2,8
O5	Leu83 N	CDK1	2,7
N4	Asp86 OD2	CDK1	2,9
N4	Gln132 O	CDK1	2,9
O6	Gln132 O	CDK1	3,2
N1	Glu81 O	CDK2	2,8
O5	Leu83 N	CDK2	2,6
N4	Asp86 OD2	CDK2	2,8
N4	Gln131 O	CDK2	2,8
O6	Gln131 O	CDK2	3,3

Tabela 12 - Ligações de hidrogênio formadas nos complexos CDK:isopenteniladenina.

Isopenteniladenina	Resíduo	CDK	Distância (Å)
N7	Lys33 NZ	CDK1	3,1
N9	Glu81 O	CDK1	2,8
N7	Lys33 NZ	CDK2	3,1
N9	Glu81 O	CDK2	2,8

Em termos de interações em níveis atômicos, algumas características como ligações de hidrogênio, interações eletrostáticas são conservadas. A existência das complementaridades nas ligações ou interações entre a proteína e inibidores foram checados. Nas figuras 19 – 27 estão ilustradas a sobreposição do bolsão de ligação dos complexos CDK1:Inibidor e CDK2:Inibidor. Foi possível notar que os resíduos que realizam ligações de hidrogênio com os inibidores se sobrepõem no mesmo plano. Foi observada discreta mudança conformacional no complexo CDK1:estaurosporina. Essa diferença influenciou nos valores obtidos de RMSD da sobreposição dos carbonos alfa (α).

Figure 20 - Sobreposição dos bolsões de ligação dos complexos, em azul é apresentado o complexo CDK1:estaurosporina e em vermelho o complexo CDK2:estaurosporina

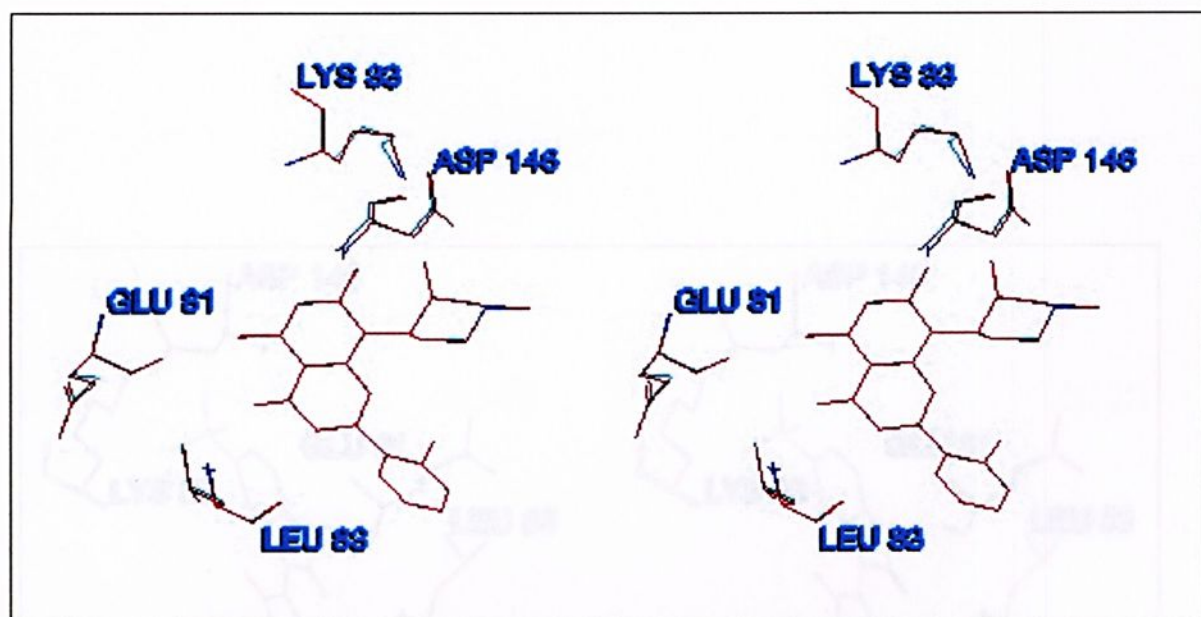


Figura 19 - Sobreposição dos bolsões de ligação dos complexos, em azul é apresentado o complexo CDK1:flavopiridol e em vermelho o complexo CDK2:flavopiridol.

Figure 21 - Sobreposição dos bolsões de ligação dos complexos, em azul é apresentado o complexo CDK1:estaurosporina e em vermelho o complexo CDK2:estaurosporina

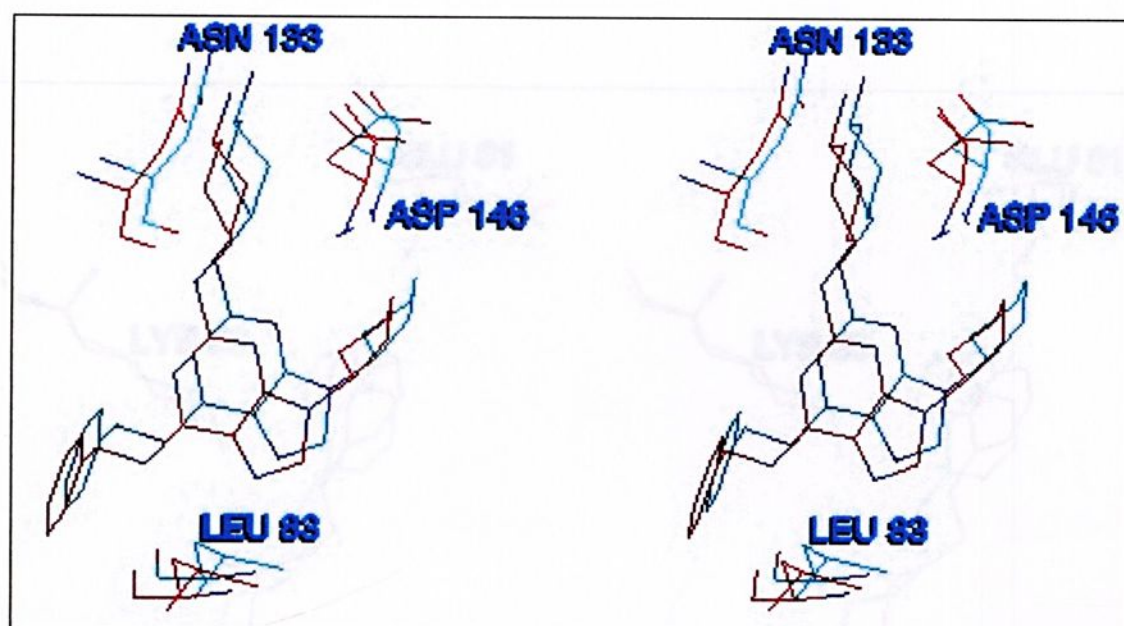


Figura 20 - Sobreposição dos bolsões de ligação dos complexos, em azul é apresentado o complexo CDK1:h717 e em vermelho o complexo CDK2:h717.

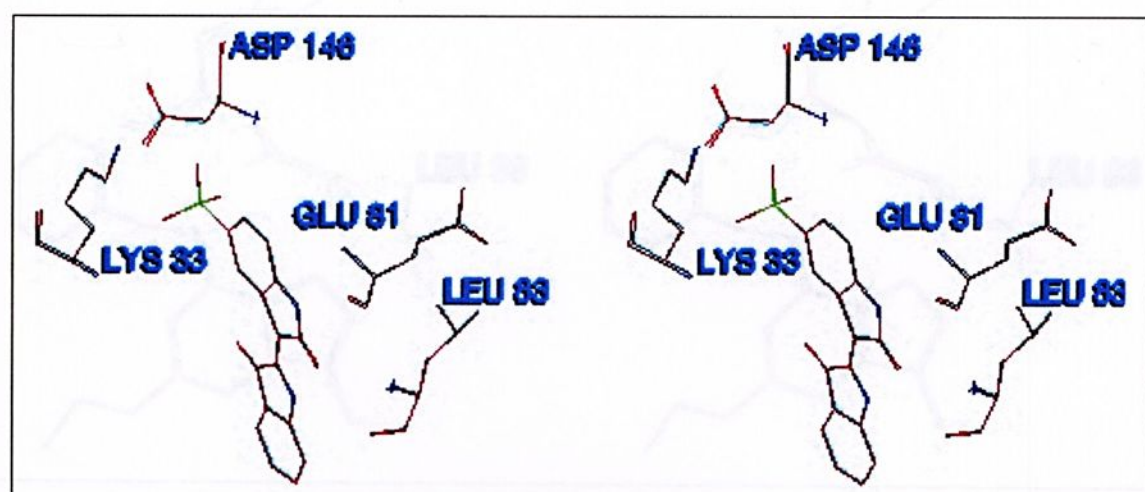


Figura 21 - Sobreposição dos bolsões de ligação dos complexos, em azul é apresentado o complexo CDK1:indirubina-5-sulfonato e em vermelho o complexo CDK2:indirubina-5-sulfonato.

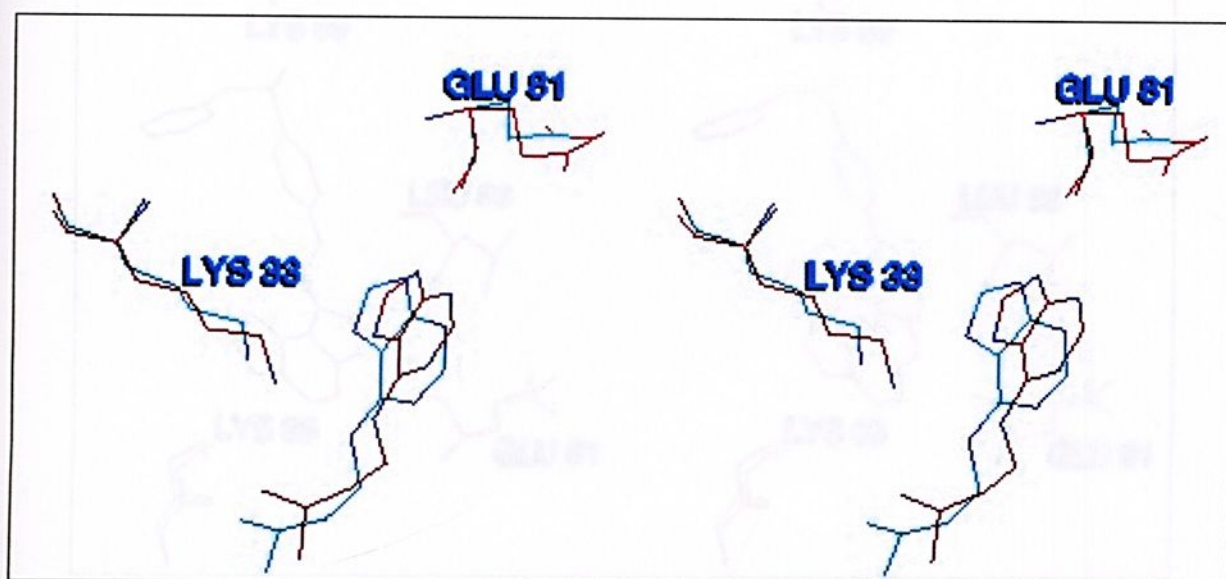


Figura 22 - Sobreposição dos bolsões de ligação dos complexos, em azul é apresentado o complexo CDK1:isopenteniladenina e em vermelho o complexo CDK2:isopenteniladenina.

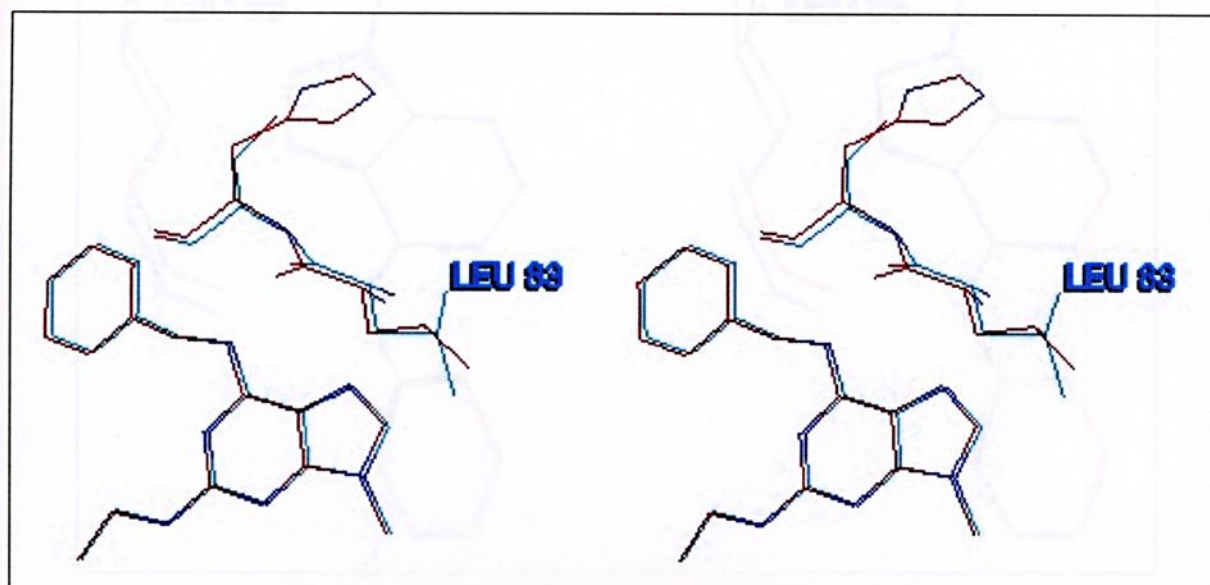


Figura 23 - Sobreposição dos bolsões de ligação dos complexos, em azul é apresentado o complexo CDK1:olomoucina e em vermelho o complexo CDK2:olomoucina.

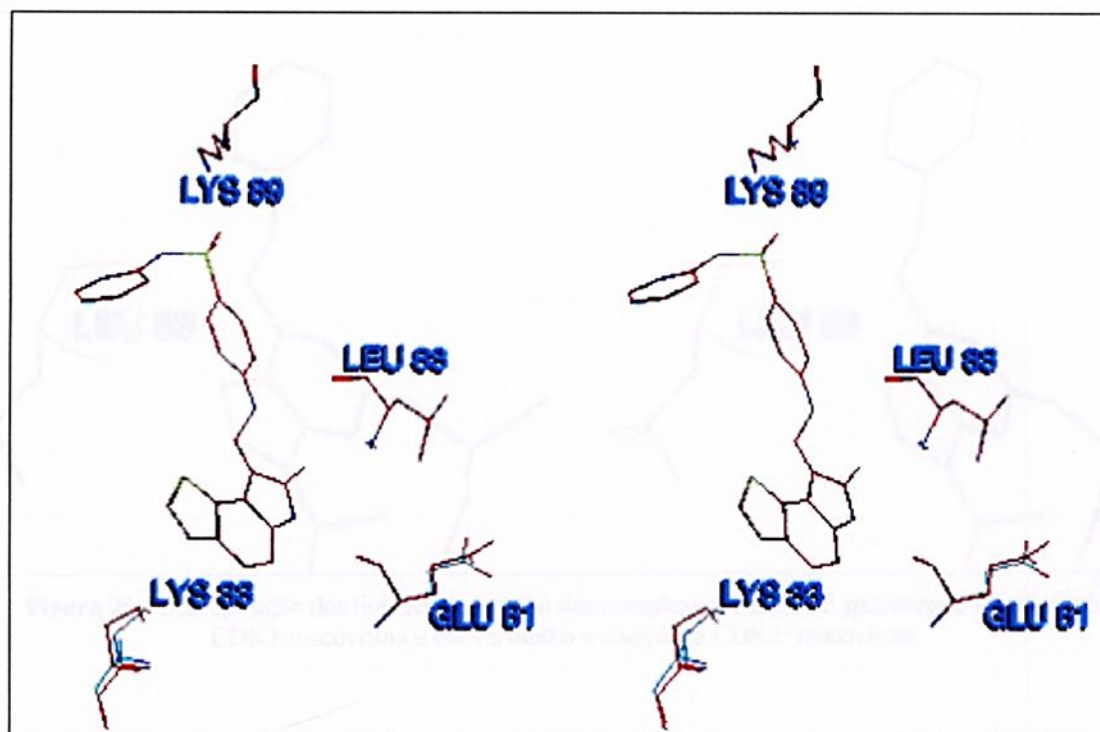


Figura 24 - Sobreposição dos bolsões de ligação dos complexos, em azul é apresentado o complexo CDK1:Oxindol-16 e em vermelho o complexo CDK2: Oxindol-16.

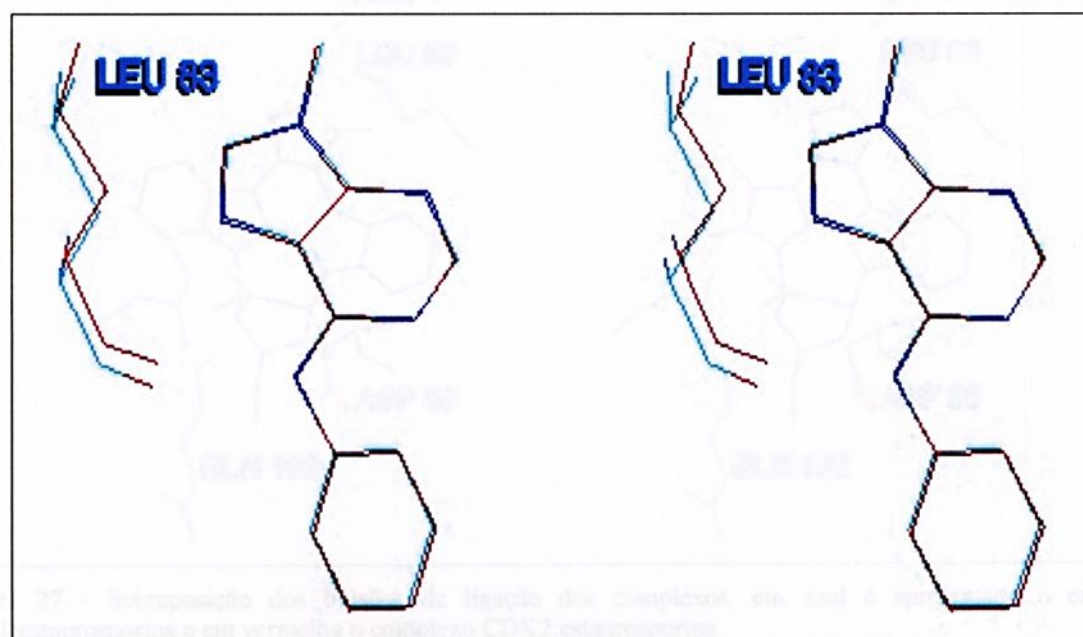


Figura 25 - Sobreposição dos bolsões de ligação dos complexos, em azul é apresentado o complexo CDK1:purvalanol e em vermelho o complexo CDK2:purvalanol.

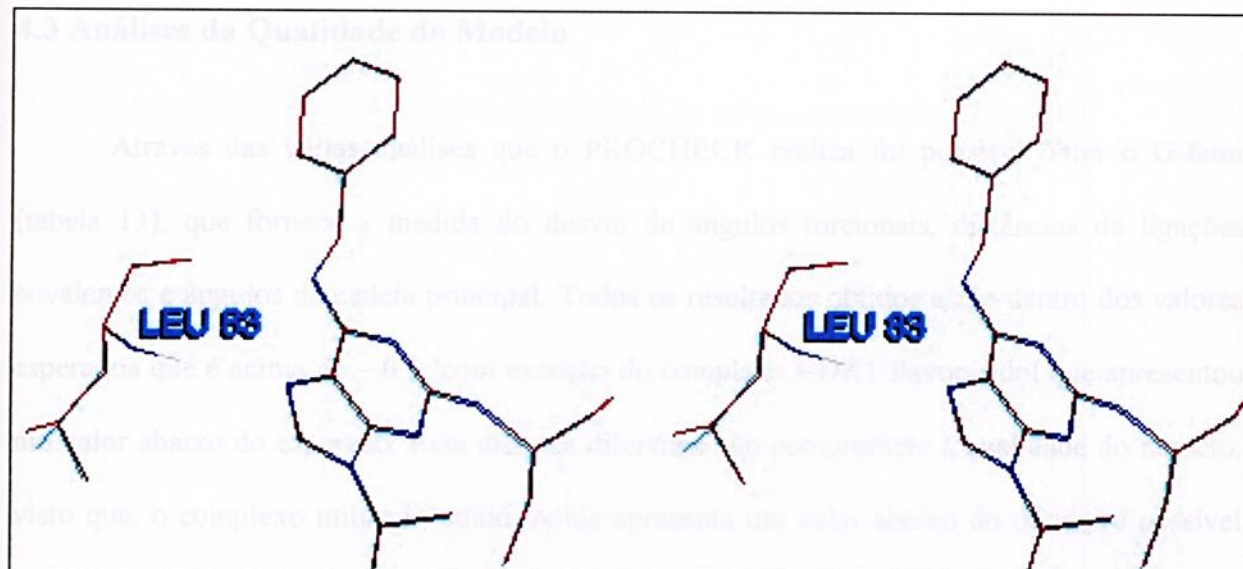


Figura 26 - Sobreposição dos bolsões de ligação dos complexos, em azul é apresentado o complexo CDK1:roscovitina e em vermelho o complexo CDK2: roscovitina

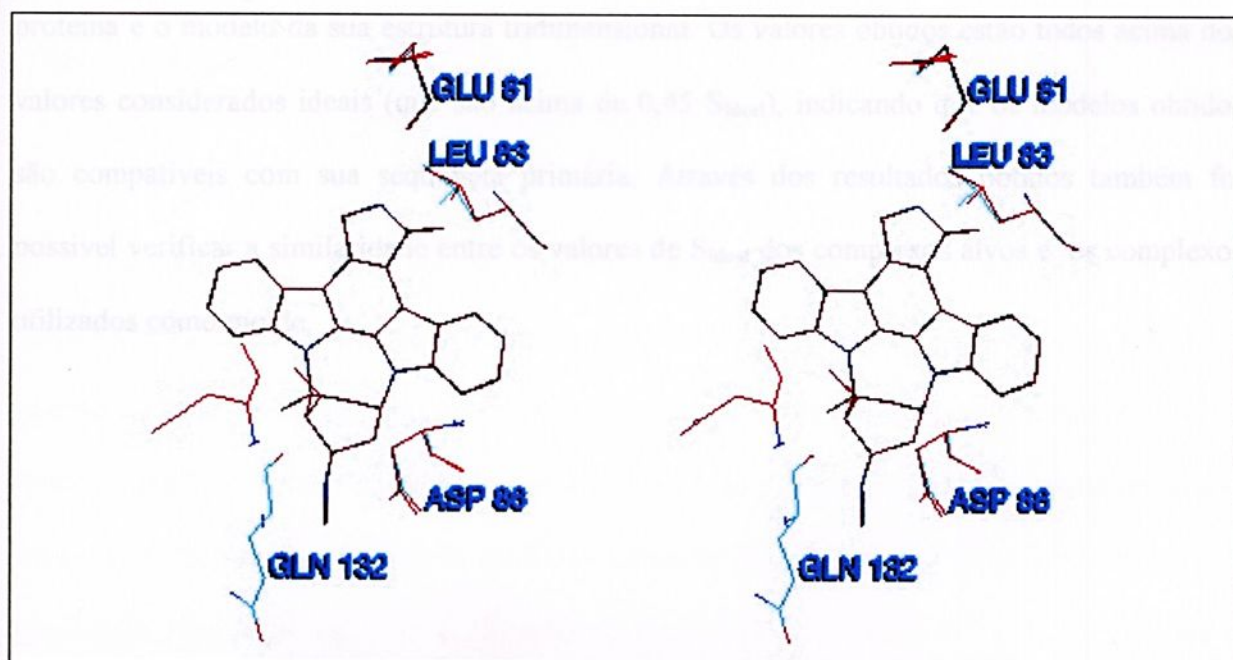


Figura 27 - Sobreposição dos bolsões de ligação dos complexos, em azul é apresentado o complexo CDK1:estaurosporina e em vermelho o complexo CDK2:estaurosporina.

4.3 Análises da Qualidade do Modelo

Através das várias análises que o PROCHECK realiza foi possível obter o G-fator (tabela 13), que fornece a medida do desvio de ângulos torcionais, distâncias de ligações covalentes e ângulos da cadeia principal. Todos os resultados obtidos estão dentro dos valores esperados que é acima de $-0,5$, com exceção do complexo CDK1:flavopiridol que apresentou um valor abaixo do esperado. Essa discreta diferença não compromete a qualidade do modelo, visto que, o complexo utilizado como molde apresenta um valor abaixo do obtido. é possível concluir que os modelos obtidos apresentam boa qualidade, não necessitando de novas verificações quanto aos ângulos de torções e geometria covalente. Na tabela 13 são relacionados também os resultados obtidos através da execução do programa VERIFY-3D (LUTHY et al., 1992) que mede a compatibilidade entre a sequência de aminoácidos de uma proteína e o modelo da sua estrutura tridimensional. Os valores obtidos estão todos acima dos valores considerados ideais (que são acima de $0,45 S_{Ideal}$), indicando que os modelos obtidos são compatíveis com sua sequência primária. Através dos resultados obtidos também foi possível verificar a similaridade entre os valores de S_{Ideal} dos complexos alvos e os complexos utilizados como molde.



Tabela 13 - Resultado de G-fator e 3D Profile do complexo CDK1:Inibidor e moldes.

Complexo	G-fator		3D Profile		
	Ângulos de torção	Geometria Covalente	Escore Total	Escore Ideal	S _{ideal} Escore
CDK1:flavopiridol	-0,28(-0,71)	-0,51(-0,73)	102,69(109,38)	135,54(136,00)	0,758(0,804)
CDK1:h717	-0,10(0,06)	-0,31(0,50)	119,60(124,63)	135,54(125,42)	0,882(0,994)
CDK1:purvalanol	-0,02(0,14)	-0,10(0,54)	127,75(130,76)	135,54(127,26)	0,942(1,027)
CDK1:estaurosporina	-0,08(0,00)	-0,23(0,23)	124,79(118,76)	135,54(126,34)	0,921(0,940)
CDK1:indirubina-5-sulfonato	-0,11(-0,05)	-0,37(0,43)	128,85(125,73)	135,54(134,62)	0,950(0,934)
CDK1:oxindol-16	-0,05(0,11)	-0,24(0,52)	130,63(134,57)	135,54(136,00)	0,964(0,989)
CDK1:roscovitina	-0,11(-0,48)	-0,29(0,45)	129,74(128,81)	135,54(130,48)	0,957(0,987)
CDK1:isopenteniladenina	-0,07(-0,05)	-0,26(-0,23)	133,33(131,21)	135,54(135,54)	0,984(0,968)
CDK1:olomoucina	-0,06(-0,20)	-0,37(-1,14)	131,38(129,94)	135,54(134,62)	0,970(0,965)

❖ Os valores apresentados dentro dos parênteses são dos complexos moldes.

Na tabela 14 estão descritos os valores obtidos de RMSD da geometria ideal (KEARSLEY, 1989) que calcula o quanto a qualidade estrutural do modelo, com relação a posição de cada átomo em uma estrutura, varia para um conjunto de confôrmeros. Para cada átomo o RMSD é a raiz quadrada da soma do quadrado das distâncias entre o referido átomo em cada confôrmero e sua posição média no conjunto de confôrmeros. O desvio máximo dos ângulos de ligação e distância de ligação aceito é respectivamente 10° e 0,1 Å (ENGH e HUBER, 1991). Todos os valores obtidos estão dentro dos valores esperados. Na tabela 15 são descritos os valores do RMSD obtidos pela sobreposição dos complexos. Estes resultados foram usados para comparar estruturas moleculares das proteínas alvo e molde, também foram calculados os valores do RMSD do sítio de ligação. Os valores considerados ideais para os



complexos em estudo seria abaixo de 1, visto que, as proteínas que apresentam um grau de identidade acima de 50% é esperado valores de RMSD para a proteína como um todo maior ou igual a 1. Para todos os complexos CDK1:Inibidores em análise obtidos valores abaixo de 1, exceto CDK1:isopenteniladenina que apresentou um valor acima do esperado. Essa alteração é devido a diferenças conformacionais entre a proteína alvo e a proteína molde. Já o valor do RMSD do sítio de ligação apresentou resultado acima do esperado, o valor apresentado pelo complexo (CDK1:estaurosporina) foi devido a não sobreposição dos resíduos Gln132 e Asp132 por estarem em conformações diferentes. Os outros complexos apresentam valores dentro do esperado mostrando que os modelos obtidos apresentam boa qualidade.

Tabela 14 - Valores de RMSD para a Geometria Ideal.

Complexo	RMSD Geometria Ideal	
	Ângulos de ligação (°)	Distância de ligação (Å)
CDK1:flavopiridol	4,8	0,02
CDK1:h717	3,4	0,02
CDK1:purvalanol	3,4	0,02
CDK1:estaurosporina	3,1	0,02
CDK1:indirubina-5-sulfonato	3,8	0,02
CDK1:oxindol-16	3,1	0,02
CDK1:roscovitina	3,5	0,02
CDK1:isopenteniladenina	3,1	0,02
CDK1:olomoucina	3,2	0,02

❖ Os valores apresentados entre parênteses são para CDK2.



Tabela 15 - Valores de RMSD obtidos através da sobreposição dos complexos e para os sítios de ligação.

Complexo	RMSD (Proteína)	RMSD (Sítio)
CDK1:flavopiridol	0,131	0,021
CDK1:h717	0,194	0,021
CDK1:purvalanol	0,113	0,009
CDK1:estaurosporina	0,082	1,314
CDK1:indirubina-5-sulfonato	0,147	0,035
CDK1:oxindol-16	0,074	0,118
CDK1:roscovitina	0,149	0,098
CDK1:isopenteniladenina	1,187	0,009
CDK1:olomoucina	0,228	0,061

Na tabela 16, a segunda coluna descreve os códigos dos arquivos de coordenadas atômicas dos moldes depositados no PDB, na terceira coluna estão descritos as resoluções em que as estruturas foram resolvidas, que varia entre 2,00 e 2,80 Å. Na quarta coluna são apresentados os valores da área de contato entre a proteína e o inibidor em estudo. Estes resultados foram obtidos com a execução das rotinas do pacote CCP4; areaimol, diffarea e resarea (CCP4, 1994). Quanto maior for o valor de área de contato, maior será o contato do inibidor com a proteína, o que resulta em numero maior de ligações de hidrogênio. Foi observado que todos os complexos em estudo com exceção de três: CDK1:Purvalanol, CDK1:Isopenteniladenina e CDK1:Olomoucina, mostram valores de contato de área maior que os valores apresentados pelos complexos da CDK2, indicando que nos complexos em estudo o inibidor faz maior contato com a proteína. São apresentados também os valores de IC50 dos complexos da CDK1 e dos complexos da CDK2. IC50 é a concentração de inibidor necessária



complexos da CDK1 e dos complexos da CDK2. IC50 é a concentração de inibidor necessária para causar a inibição de 50% da atividade da proteína. Os valores de IC50 são determinados experimentalmente (KNOCKAERT *et al.*, 2002). Foi possível verificar através dos valores relacionados na tabela 16 que são valores iguais ou similares os moldes, indicando que os inibidores também são específicos para inibição da CDK1.

Tabela 16 - Código e resolução das estruturas cristalográficas de CDK2 usadas como moldes, valores de contato de superfícies e de IC₅₀ dos complexos CDK1:Inibidores e CDK2:Inibidores.

Complexo	Código de Acesso PDB	Área de Contato (Å²)	IC ₅₀ (µM)	Referência
CDK1:Flavopiridol	-	325 (320)	0,4(0,1)	DE AZEVEDO JR. <i>et al.</i> , 1996.
CDK1:H717	1G5S	351(346)	0,2 (0,05)	DREYER <i>et al.</i> , 2001.
CDK1:Purvalanol	1CKP	226(239)	0,006 (0,006)	GRAY, N. <i>et al.</i> , 1998.
CDK1:Estaurosporina	1AQ1	347 (340)	0,006(0,007)	LAWRIE <i>et al.</i> , 1997.
CDK1:Indirubina-5-Sulfonato	1E9H	270 (260)	0,06(0,04)	HOESSEL <i>et al.</i> , 1999.
CDK1:Oxindol-16	1FVV	338 (337)	0,8(0,06)	BRAMSON <i>et al.</i> , 2001.
CDK1:Roscovitina	-	336 (320)	0,5(0,7)	DE AZEVEDO JR. <i>et al.</i> , 1997
CDK1:Isopenteniladenina	-	189 (200)	45(50)	SCHULZE-GAHMEN <i>et al.</i> , 1995.
CDK1:Olomoucina	-	274 (288)	7(7)	SCHULZE-GAHMEN <i>et al.</i> , 1995.

❖ Os valores apresentados entre parênteses são para CDK2.

CONCLUSÃO

As estruturas obtidas dos complexos indicam que os inibidores estão fortemente ligados ao bolsão de ligação do ATP, com áreas de contato de similares. Existem algumas variações nos valores de área de contato obtidos para os complexos de CDK1 em relação aos de CDK2. Esta variação que é discreta não tem nenhuma correlação direta com a atividade destes inibidores. Todos os inibidores estudados apresentam afinidade similar tanto para CDK1 como para a CDK2.

O Garfo Molecular identificado por estudos anteriores na estrutura CDK2 parece ser conservado na estrutura CDK1, e também participa na formação do conjunto de ligações de hidrogênio com o inibidor. Através dos resultados de área de contato obtidos foi possível verificar algumas variações que possivelmente são ocasionadas devido ao padrão de ligação de hidrogênio que são feitas entre o inibidor e o garfo molecular. Sugerindo que esse padrão ligação exerce um papel fundamental para a determinação na especificidade de inibidores para a CDK9, CDK5 e CDK2 e CDK1.

Os inibidores em estudo apresentam valores de IC50 similares aos valores que foram determinados experimentalmente (estudos anteriores) para a inibição da CDK2 que foram usados como molde. Existem discretas diferenças que provavelmente se devem a algumas mudanças conformacionais de cadeias laterais que influenciam na formação de ligação de hidrogênio e na área de contato. A similaridade nos valores de IC50 indicam que os inibidores em estudo podem inibir a CDK1 com eficiência similar a CDK2.



REFERÊNCIAS BIBLIOGRÁFICAS

- 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, p.107-132, 1987.
- ALEEM, E., BERTHET, C., KALDIS, Cdk2 as a master of S phase entry: fact or fake? *Cell Cycle*, v. 3, p. 35-37, 2004.
- ALTSCHUL, S. F. *et al.* Basic local alignment search tool. *J. Mol. Biol.*, v. 215, p. 403-410, 1990.
- ALTSCHUL, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, v.25, p.3389-3402, 1997.
- ASZODI, A.; TAYLOR, W. R. Homology modelling by distance geometry. *Fold. Des.*, v. 1, n. 5, p. 325 – 334, 1996.
- BACON, D. J.; ANDERSON, W. F. A fast algorithm for rendering space-filling molecule pictures, *J. Mol. Graphics.*, v. 6, p. 219-220, 1988.
- BARKER, W. C. *et al.* The Protein Information Resource (PIR). *Nucleic Acids Res.*, v. 28, p. 41-44, 2000.
- BARTON, G. J. Protein structure Prediction – a practical approach, Oxford University Press, 1996.
- BENSON, D. A. *et al.* GenBank. *Nucleic Acids Res.*, v. 28, p.15 – 18, 2000.
- BERENDSEN, H. J. C.; VAN DER SPOEL, D.; VAN DRUNEN, R. GROMACS: a message-passing parallel molecular dynamics implementation. *Comp. Phys. Commun.*, v. 91, p 43-56, 1995.
- BERMAN, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.*, v. 28, p. 235-242, 2000.
- BERNSTEIN, F. C. *et al.* The Protein Data Bank: a computer-based archival file for macromolecular structures. *J Mol Biol.*, v. 112, n. 3, p. 535-542, 1977.
- BERTHET, C. *et al.* Cdk2 knockout mice are viable. *Curr. Biol.*, v.13, p. 1775-1785, 2003.
- BLUNDELL, T. L. *et al.* Knowledge-based prediction of protein structures and the design of novel molecules. *Nature*, v.326, n.6111, p. 347 – 352, 1987.



BOWER, M. J.; COHEN, F. E.; DUNBRACK JR, R.L Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: a new homology modeling tool. *J. Mol. Biol.*, v. 267 n. 5, p. 1268-1282, 1997.

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

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

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.

BRUCCOLERI, R. E.; KARPLUS, M. Prediction of the folding of short polypeptide segments by uniform conformational sampling. *Biopolymers*, v. 26, n.1, p.137 – 168, 1987.

BRUCE A. *et al.* Molecular biology of the cell. 3rd ed. New York: London Garland, 1994.

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

CAI, D. *et al.* Stable expression of wild-type p53 gene in human lung cancer cells after retrovirusmediated gene transfer. *Human Gene Therapy*, v. 4, p. 617-624, 1993.

CANDURI, F., UCHOA, H. B., DE AZEVEDO, W. F. JR. Molecular models of cyclin-dependent kinase 1 complexed with inhibitors. *Biochem. Biophys. Res. Commun.*, v. 324, p. 661-666, 2004.

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

CANDURI, F. *et al.* Crystal structure of human uropepsin at 2.45Å resolution. *Acta Crystallogr. Sect. D- Biol. Crystallogr.*, v. 57, p. 1560-70, 2001.

CLAESSENS, M. *et al.* Modelling the polypeptide backbone with 'spare parts' from known protein structures. *Protein Eng.*, v. 2, p. 335 – 345, 1989.

COLLABORATIVE COMPUTACIONAL PROJECT N° 4. The CCP4 Suite: Programs for proteínas crystallography. *Acta Crystallogr.*, v. D50, p. 760-763, 1994.

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

CREGUT, D.; LIAUTARD, J. P.; CHICHE, L. Homology modelling of annexin I: implicit solvation improves side-chain prediction and combination of evaluation criteria allows recognition of different types of conformational error. *Protein Eng.*, 1 v. 7, 1p. 333-1344, 1994.

CHAO, K. M.; PEARSON, W. R.; MILLER, W. Aligning two sequences within a speci.ed diagonal band, *Comput. Appl. Biosci.*, v.8, p. 481–487, 1992.



CHOTHIA, C. *et al.* The predicted structure of immunoglobulin D1.3 and its comparison with the crystal structure. *Science*, v.233, n. 4765, p. 755 – 758, 1986.

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

D'ALFONSO, G.; TRAMONTANO, A.; LAHM, A. Structural conservation in single-domain proteins: implications for homology modeling. *J. Struct. Biol.*, v. 134, n. 2-3, p.246 – 256, 2001.

DAYHOFF, M. O.; ECK, R.V. (eds) Atlas of Protein Sequence and Structure (Natl. Biomed. Res. Found., Silver Spring, MD), Vol. 3, p.33 1968.

DE AZEVEDO JR, W. F. *et al.* Inhibition of cyclin-dependent kinases by purine analogues: crystal structure of human cdk2 complexed with roscovitine. *Eur. J. Biochem.*, v. 243, n. 1, p. 518-526, 1997.

DE AZEVEDO JR, W. F. *et al.* Molecular model for the binary complex of uropepsin and pepstatin. *Biochem. Biophys. Res. Commun.*, v. 287 n. 1, p. 277-281, 2001.

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

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

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

DELANO, W. L. The case for open-source software in drug discovery. *Drug Discov Today.*, v. 10, n. 3, p. 213-217, 2005.

DORER R, PRYCIAK P. M, HARTWELL L. H. *Saccharomyces cerevisiae* cells execute a default pathway to select a mate in the absence of pheromone gradients. *J. Cell Biol.*, v. 131, p. 845-861, 1995.

DREYER, M. K. *et al.*, Crystal structure of human cyclin-dependent kinase 2 in complex with the adenine-derived inhibitor H717. *J. Med. Chem.*, v. 44, p. 524–530, 2001.

DUNBRACK, R. L.; KARPLUS, M. Backbone-dependent rotamer library for proteins. Application to side-chain prediction. *J. Mol. Biol.*, v. 230, n. 2, p. 543-574, 1993.

ENGH, R. A.; HUBER, R. Accurate bond and angle parameters for X-ray protein structure refinement. *Acta crystallogr.*, v. A47, p. 392-400, 1991.

EU 3-D Validation Network, Who checks the checkers? Four validation tools applied to eight atomic resolution structures, *J. Mol. Biol.*, v. 276, p. 417–436, 1998.



- FINE, R. M. *et al.* Predicting antibody hypervariable loop conformations. II: Minimization and molecular dynamics studies of MCPC603 from many randomly generated loop conformations. *Proteins*, v.1, n. 4, p. 342 – 362, 1986.
- FISER, A.; DO, R. K. G.; SALI, A. Modeling of loops in protein structures [In Process Citation]. *Protein Sci.*, v. 9 p. 1753 – 1753, 2000.
- FITCH, W. M. An improved method of testing for evolutionary homology. *J. Mol. Bio.*, v.16, n. 1, p. 9 – 16, 1966.
- FLOHIL, J. A.; VRIEND, G.; BERENDSEN, H. J. Completion and refinement of 3-D homology models with restricted molecular dynamics: application to targets 47, 58, and 111 in the CASP modeling competition and posterior analysis. *Proteins.*, v. 48, p. 593–604, 2002.
- GASTEIGER, E.; JUNG, E.; BAIROCH, A. SWISS-PROT: connecting biomolecular knowledge via a protein database. *Curr. Issues Mol. Biol.*, v. 3, n. 3, p. 47-55, 2001.
- GIBBS, A. J.; MCLNTYRE, G. A. The diagram, a method for comparing sequences. Its use with amino acid and nucleotide sequences. *Eur. J. Biochem.*, v.16, p. 1 – 11, 1970.
- GRAY, N. *et al.*, Exploiting chemical libraries, structure, and genomics in the search for new kinase inhibitors. *Science*, v. 281, p. 533–538, 1998.
- GREER, J. Comparative modeling methods: application to the family of the mammalian serine roteases. *Proteins*, v. 7, n. 4, p. 317 – 334, 1990.
- GREER, J. Model for Haptoglobin Heavy Chain Based upon Structural Homology. *PNAS.*, v. 77, p. 3393 – 3397, 1980.
- HARTWELL, L.; WEINERT, T. Checkpoints: Controls that ensure the order of cell cycle events. *Science*, v. 246, p. 629-634, 1989.
- HEINEMANN, U. Structural genomics in Europe: slow start, strong finish?. *Nat. Struct. Biol.*, v. 7, p. 940-942, 2000.
- HOESSEL R. *et al.*, Indirubin, the active constituent of a Chinese antileukaemia medicine, inhibits cyclin-dependent kinases. *Nat. Cell Biol.*, v.1, p. 60–67, 1999.
- HOOFT, R.W. *et al.* Errors in protein structures. *Nature*, v. 381, p. 272, 1996.
- HUNT, T. Cell biology. Cell cycle gets more cyclins. *Nature*, v. 350, n. 6318, p. 462-43, 1991.
- JEANMOUGIN, F. *et al.* Multiple sequence alignment with Clustal X. *Trends Biochem., Sci.* v. 23, n.10, p. 403-405, 1998.
- JEFFREY, P. *et al.* Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex. *Nature*, v. 376, n. 6538, p. 313-320, 1995.
- JOHNSON, M. S. *et al.* Knowledge-based protein modeling. *Crit. Rev. Biochem. Mol. Biol.*, v. 29, p. 1 – 68, 1994.



- JONES, T. A.; THIRUP, S. Using known substructures in protein model building and crystallography. *EMBO J.*, v. 5, p. 819 – 822, 1986..
- JUNG, S. H.; PASTAN, I.; LEE, B. Design of interchain disulfide bonds in the framework region of the Fv fragment of the monoclonal antibody B3. *Proteins*, v. 19, n. 1, p. 35 – 47, 1994.
- KEARSLEY, S. K. On the orthogonal transformation used for structural comparisons. *Acta Crystallogr.*, v. A45, p. 208-210, 1989.
- 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.
- KNOCKAERT, M.; GREENGARD, P AND MEIJER, L. Pharmacological inhibitors of cyclin-dependent kinases. *Trends Pharmacol Sci.*, v. 23, v. 9, p. 417-25, 2002.
- KOEHL, P.; LEVITT, M. A brighter future for protein structure prediction. *Nat. Struct. Biol.*, v. 6, p. 108–111, 1999.
- KRAULIS, P. J. Molscript: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, v. 24, p. 137-145, 1991.
- LANDER, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature*, v. 409, p. 860-921, 2001.
- LASKOWSKI, R. A.; MACARTHUR, M. W.; MOSS, D. S.; THORNTON, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, v. 26, p. 283-291, 1993.
- LAWRIE, A. M.; ZOLLE, O.; SIMPSON A. W. Protein kinase inhibition by staurosporine revealed in details of the molecular interaction with CDK2. *Nat. Struct. Biol.*, v. 4, p. 796–800, 1997.
- LEE, M.G.; NURSE, P. Complementation used to clone a human homologue of the fission yeast cell cycle control gene *cdc2*. *Nature*, v. 327, n. 6117, p. 31-35, 1987.
- LEVITT, M. Accurate modeling of protein conformation by automatic segment matching. *J. Mol. Biol.*, v. 226, p. 507 – 533, 1992.
- LINDAHL, E.; HESS, B.; VAN DER SPOEL, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Mod.*, v. 7, p. 306–317, 2001.
- LUTHY, R.; BOWIE J.; EISENBERG, D. Assessment of protein models with three-dimensional profiles. *Nature.*, v. 356 p. 83-85, 1992.
- MAGGIO, E. T.; RAMNARAYAN, K. Recent developments in computational *proteomics*. *Trends Biotechnol*, v. 19, n. 7, p. 266-272, 2001.
- MALUMBRES, M. *et al.* Mammalian cells cycle without the D-type cyclin-dependent kinases Cdk4 and Cdk6. *Cell*, v. 118, p. 493-504, 2004.



MARTIN, A. C. R.; CHEETHAM, J. C.; REES, A. R. Modeling antibody hypervariable loops: a combined algorithm. *Proc. Natl. Acad. Sci., U.S.A.* v. 86, n.23, p. 9268 – 9272, 1989.

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

MAS, M. T. *et al.* Modeling the anti-CEA antibody combining site by homology and conformational search. *Proteins*, v. 14, n. 4, p. 483 – 498, 1992.

MERRITT, E. A. RASTER3D VERSION 2.0. A program for photorealistic molecular graphics, *Acta Cryst.*, v. 50 p. 869–873, 1994.

MOULT, J. The current state of the art in protein structure prediction. *Curr. Opin. Biotechnol.*, v. 7, p 422–427, 1996.

MOULT, J.; JAMES, M. N. An algorithm for determining the conformation of polypeptide segments in proteins by systematic search. *Proteins*, v. 1, n. 2, p. 146 – 163, 1986.

NEEDLEMAN, S. B.; WUNSCH, C. D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.*, v. 48, n. 3, p. 443 – 453, 1970.

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.

NORVELL, J. C.; MACHALEK, A. Z. Structural genomics programs at the US National Institute of General Medical Sciences. *Nat. Struct. Biol.*, v. 7, p. 931, 2000.

OHLSON, T.; WALLNER, B.; ELOFSSON, A. Profile–profile methods provide improved fold-recognition: a study of different profile–profile alignment methods. *Proteins.*, v. 57, p. 188–197, 2004

ORTIZ, A. R.; KOLINSKI, A.; SKOLNICK, J. Fold assembly of small proteins using monte carlo simulations driven by restraints derived from multiple sequence alignments. *J. Mol. Biol.*, v. 277, n. 2, p. 419-448, 1998.

OSGUTHORPE, D. J. Analysis of the predicted structures of domain 1 of protein g3 (T0030) and NK-lysin (T0042). *Proteins Suppl*, v. 1, p. 172-178, 1997.

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

PEITSCH, M. C. CHARIFSON, P. S. Em Practical Application of Computer-Aided Drug Design, ed.; Marcel Dekker: New York, 1997.

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



ROBERT, S. Structural genomics, round 2 "As NIH plans to extend its high-speed structural biology program for another 5 years, researchers remain divided on how to best allocate its shrinking budget". *Science*, v. 307, p. 1554-1557, 2005.

ROST, B.; SANDER, C. Bridging the protein sequence-structure gap by structure predictions. *Annu. Rev. Biophys. Biomol. Struct.*, v. 25, p.113-136, 1996.

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

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

SANCHEZ, R.; SALI, A. Evaluation of comparative protein structure modeling by MODELLER-3. *Proteins*, v. 1, p. 50 – 58, 1997.

SANTOS FILHO, O. A. DE ALENCASTRO, R. B.; FIGUEROA-VILLAR, J. D. Homology modeling of wild type and pyrimethamine/cycloguanil-cross resistant mutant type *Plasmodium falciparum* dihydrofolate reductase. A model for antimalarial chemotherapy resistance. *Biophys. Chem.*, v. 91, p. 305-317, 2001.

SCHULZE-GAHMEN, U. *et al.* 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, n. 4, p. 378-391, 1995.

SCHULZE-GAHMEN, U. *et al.* Multiple modes of ligand recognition: crystal structures of cyclin-dependent protein kinase 2 in complex with ATP and two inhibitors, olomoucine and isopentenyladenine. *Proteins Struct. Funct. Genet.*, v. 22, n. 4, p. 378-391, 1995.

SIMONS, K. T. *et al.* Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *J. Mol. Biol.*, v. 268, p. 209-225, 1997.

SIMONS, K. T. *et al.* Improved recognition of native-like protein structures using a combination of sequence-dependent and sequence-independent features of proteins. *Proteins*, v. 34, n. 1, p. 82-95, 1999.

SRINIVASAN, S.; MARCH, C. J.; SUDARSANAM, S. An automated method for modeling proteins on known templates using distance geometry. *Protein Sci.*, v. 2, p. 227 – 289, 1993.

TERWILLIGER, T. C. Structural genomics in North America. *Nat. Struct. Bio.*, v. 7, p. 935-939, 2000.

UCHOA, H. B. *et al.* PARMODEL: a web server for automated comparative modeling of proteins. *Biochem Biophys Res Commun*, v. 325, n. 4, p.1481-1486, 2004.

VALENTINI, C. *et al.* Activity of cyclin-dependent protein kinase (CDK)-inhibitor CYC202 (R-roscovitine) in mantle cell lymphoma. *Proc. Am. Assoc. Cancer Res.*, v. 45: p. 827, 2004.

VAN GUNSTEREN, W. F. *et al.* Biomolecular Simulations: The GROMOS96 Manual and User Guide. *VdF Hochschulverlag ETHZ*, Zurich, 1996.



VAN VLIJMEN, H. W.; KARPLUS, M. PDB-based protein loop prediction: parameters for selection and methods for optimization. *J. Mol. Biol.*, v. 267, n. 4, p. 975 – 1001, 1997.

VASQUEZ, M. Modeling side-chain conformation. *Curr. Opin. Struct. Biol.*, v. 6, n. 2, p. 217 – 221, 1996.

VENTER, J. C. *et al.* The Sequence of the Human Genome. *Science*, v. 291, p. 1304-1351, 2001.

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

WARD, L. S. Understanding the Molecular Process of Tumorigenesis. *Arq Bras Endocrinol Metab.*, v. 46, n. 4, p. 351-360, 2002.

WILSON, C.; GREGORET, L. M.; AGARD, D. A. Modeling side-chain conformation for homologous proteins using an energy-based rotamer search. *J. Mol. Biol.*, v. 229, n. 4, p.996-1006, 1993.

YOKOYAMA, S. *et al.* Structural genomics projects in Japan. *Nat. Struct. Biol.*, v. 7, p. 943-945, 2000.



Neste apêndice é apresentado um resumo do principal artigo sobre o MODELLER (SALI e BLUNDELL, 1993).

MODELLER é um software que modela a estrutura 3D de proteínas utilizando a técnica de satisfação de restrições na forma de distância átomo-átomo e de ângulos diedrais que são extraídos das estruturas já conhecidas. As entradas para o programa de forma geral são restrições na estrutura espacial da sequência de aminoácidos e ligantes a ser modelada. A saída é uma estrutura tridimensional que melhor satisfaz essas restrições. 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 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. O programa pode também executar múltiplas comparações de sequências de proteínas ou de 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.

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 o alinhamento 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 carbonos α equivalentes ou entre ângulos diedrais equivalentes em duas proteínas relacionadas. Esses relacionamentos foram descritos como funções de densidade



de probabilidade (pdf) 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 seqüências 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 e GO, 1985) que aplica o método de gradiente conjugado para as posições de todos os átomos que não sejam hidrogênio.

Derivação Restrições Espaciais

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 a integral sobre todos os valores possíveis de x resulte 1. A probabilidade de um evento $X_1 \leq X < X_2$ é obtida pela integração de p :

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

A expressão de uma restrição em termos de uma pdf fornece mais informações sobre os possíveis valores da característica a ser restrita do 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 RMN multidimensional.



Derivação de pdfs básicas.

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

$$P(x / a, b, \dots, c) \quad (2)$$

Isto é uma pdf condicional que nos dá a densidade de probabilidade para x , quando $a, b, \dots, c$ 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 na associação $(x/a, b, \dots, c)$ exceto x devem ser conhecidas no momento da predição. Contudo, x deve ser uma característica espacial da sequência a ser modelada. A característica conhecida mais útil é aquela do mesmo tipo de x , mas associada com as posições equivalentes nas estruturas conhecidas relacionadas. Por exemplo, quando restringi uma certa distância equivalente nas estruturas relacionadas.

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

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

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



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

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

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

Banco de dados local

Um banco de dados local, contendo dados estruturais de membros de 17 famílias de proteínas retirados do PDB (ABOLA *et al.*, 1987; BERNSTEIN *et al.*, 1977), foi criado dentro do próprio MODELLER para a implementação do processo de modelagem comparativa ou por homologia. As famílias de proteínas escolhidas para compor o banco de dados local foram as seguintes: Proteinases Aspárticas, Serino Proteinases, Azurinas, Imunoglobulinas, Gama-Cristalinas, Cisteíno Proteinases, Lisozimas, Globinas, Fosfolipases, Citocromos 1 e 2, Centros de Reação Fotossintética, Ferredoxinas, Flavodoxinas, Dehidrofolato Redutases, Dehidrogenases, e Oxidoreduções. Na tabela 17, pode ser visto maiores detalhes sobre o tamanho e composição deste banco de dados.



Tabela 17 - Tamanho do banco de dados local.

Alinhamento	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

Composição do Banco de dados local

O banco de dados local contém membros representativos de 4 classes estruturais de proteínas: α , β , $\alpha+\beta$ e α/β (tabela 18), sendo a classe α formada predominantemente por 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 . A frequência de resíduos na conformação de folhas- β é similar à frequência de resíduos na conformação de hélice. A percentagem de identidade para as seqüências das estruturas que compõem o banco de dados varia de 6% a 98%, isto também se reflete na diferença de vizinhança local de resíduos, para os pares de resíduos equivalentes. A maior parte das estruturas de proteínas no banco local foi resolvida a média e a alta resolução, ou seja, acima de 2,5Å. Embora, existem também algumas estruturas resolvidas a baixa resolução.



Tabela 18 - Classes Estruturais de Proteínas.

A	Globinas, Fosfolipases, Citocromos 1 e 2 e Centros de Reação Fotosintética.
B	Proteinases Aspárticas, Serino Proteinases, Azurinas, Imunoglobulinas, Gama-Cristalinas.
$\alpha+\beta$	Cisteino Proteinases e Lisozimas.
α/β	Ferrodosinas, Flavodosinas, Dehidrofolato Redutases, Dehidrogenases e Oxidoreditases.

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 estabelecer os relacionamentos gerais entre as mais variadas características das estruturas de proteínas.

Derivação das pdfs de características a partir de pdfs básicas

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

O exemplo a seguir demonstra a aplicação destas definições. O objetivo é construir uma pdf característica para uma distância particular $C\alpha-C\alpha$ em uma determinada seqüência. Suponha duas estruturas conhecidas relacionadas com distâncias equivalentes. Então, temos 2 pdfs básicas correspondentes para a distância $C\alpha-C\alpha$ na seqüência alvo. Além disso, também sabemos que cada uma das duas 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 modo a combinar todas essas informações, temos que combinar as três pdfs básicas em uma única pdf característica. Para fazer isso podemos utilizar todos os possíveis alinhamentos destas três proteínas no banco de dados local. Um exemplo de dependência da distância $C\alpha-C\alpha$ para duas distâncias equivalentes em duas estruturas $p(d/d', d'')$, é mostrado na figura 28a.

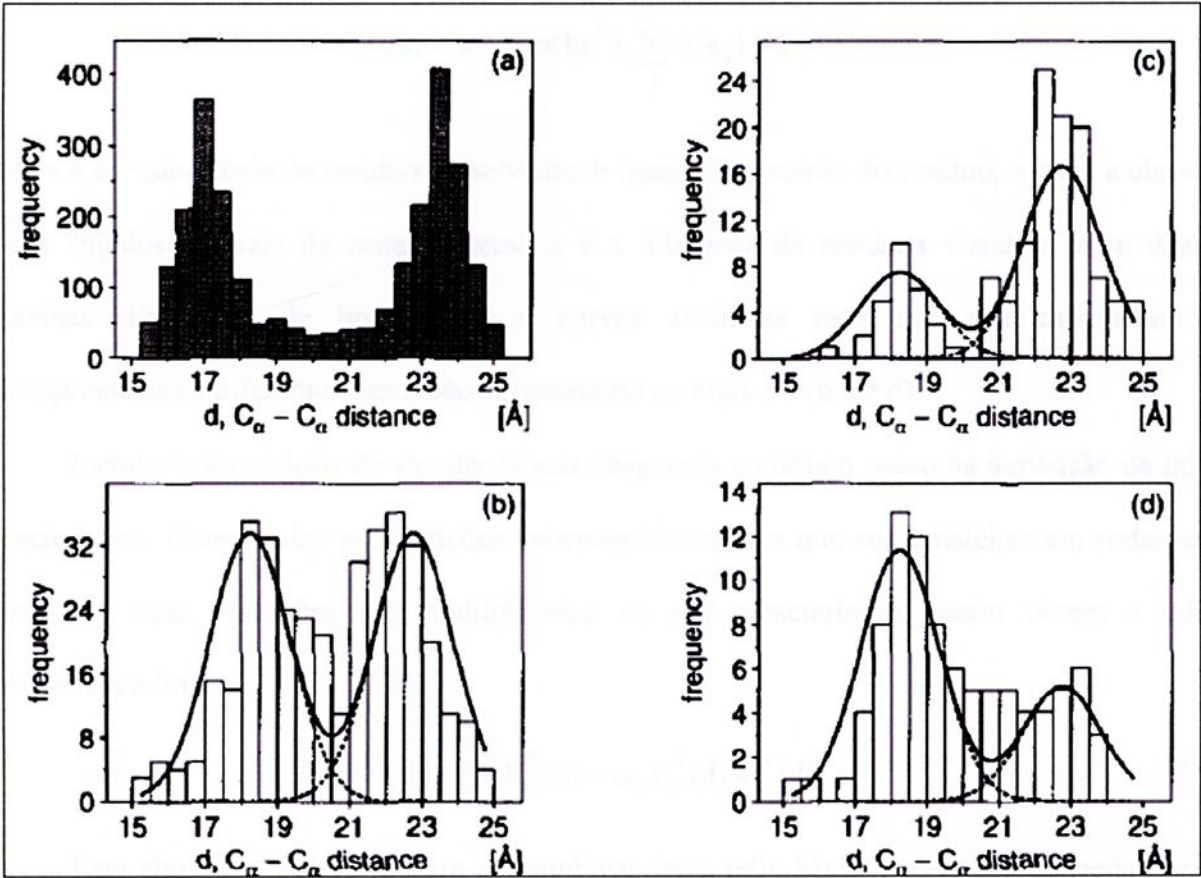


Figura 28 - Derivação de uma pdf característica a partir de pdfs básicas. Em todos os gráficos, $18,0 < d' < 18,5$ e $22,5 < d'' < 23,0$. a) $W'(d/d', d'')$. b) $W'(d/d', d'', s', s'')$, onde $0,2 < s' < 0,4$ e $0,2 < s'' < 0,4$. c) $W'(d/d', d'', s', s'')$, onde $0,2 < s' < 0,4$ e $0,4 < s'' < 0,6$. d) $W'(d/d', d'', s', s'')$, onde $0,4 < s' < 0,6$ e $0,2 < s'' < 0,4$.

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

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

esta relação gera claramente um melhor ajuste dos dados que um simples produto das pdfs individuais o que acabaria fazendo com que a distância d se tornasse um média de d' e d'' . O



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

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

onde

(6)

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

onde a é acessibilidade do resíduo ao solvente, b igual a B_{iso} médio do resíduo, c igual a classe χ dos ângulos diedrais da cadeia lateral, s é a diferença de resíduos vizinhos entre duas proteínas. Exemplos de histogramas e curvas analíticas para esta pdf característica correspondentes a diferentes pesos são apresentadas na Figura 1 (b até d).

Incluindo a restrição de van der Waals chegamos ao último passo na derivação da pdf característica. Como todas as restrições estereoquímicas tem que ser satisfeitas em todas as estruturas, estas restrições são multiplicadas na pdf característica, assim obtém a pdf característica final.

$$P^D(d) = [\omega_1 P_1^d(d) + \omega_2 P_2^d(d)] P^v(d) \quad (7)$$

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

Derivação da pdf molecular a partir das pdfs características

O último estágio na derivação de uma pdf é a combinação de todas as pdfs 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 uma probabilidade de ocorrência de qualquer combinação destas 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 características são independentes uma das outras, a pdf molecular é simplesmente um produto das pdfs características.

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

Assim, maximizando a função P podemos encontrar o modelo mais provável para a estrutura 3D desconhecida, dado seu alinhamento com as estruturas conhecidas.

Otimização da pdf molecular

A maximização da função P que representa a pdf molecular obtida, fornecerá o modelo de estrutura tridimensional. A função que é atualmente otimizada é uma transformação da pdf molecular P :

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

onde todas as características são expressas em termos de coordenadas atômicas no espaço cartesiano. A função F é chamada de função objetiva. O mesmo conjunto de coordenadas Cartesianas que maximiza P também minimiza F , todavia F é computacionalmente mais indicada para ser otimizada que P , visto que, é realizada a substituição da multiplicação de termos em P pela adição de termos em F , além disso o problema de *overflow* em ponto flutuante é menor em F do que em P .

A segunda transformação da pdf molecular original, que é útil para a otimização, é a mudança de escala dos desvios padrão das restrições básicas. Essa opção permite a mudança de escala independente para cada tipo de característica a se restringir (tamanhos de ligação,



ângulos de ligação, distâncias entre Ca-Ca , etc.). Aumentando o desvio padrão, as restrições se tornam menos poderosas, maiores violações tendem a ocorrer, similar para o efeito que um aumento na constante de força faz na minimização de energia.

Inicialmente o método da função alvo variável foi utilizado para o cálculo de estruturas tridimensionais de proteínas consistentes com restrições obtidas de RMN (BRAUN e GO, 1985). A diferença principal entre o método original e a implementação atual é que o procedimento original otimiza os ângulos diedrais e o processo de otimização atual é feito no espaço cartesiano. De acordo com o método da função 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 que está sendo modelada. A função alvo $P(\Delta r)$ é obtida da mesma forma que a pdf molecular, exceto que somente aquelas restrições cujos átomos que as originam pertençam a resíduos não mais que Δr resíduos separados na sequência, são incluídas. O cálculo como um todo consiste de um grande número de otimizações de gradientes conjugado da função alvo $P(\Delta r)$ com valores crescentes de Δr . A conformação inicial $P(1)$ a ser otimizada também é uma estrutura estendida ou uma conformação derivada de uma cadeia estendida através de rotação dos ângulos diedrais das cadeias principal e lateral. Nos passos subsequentes do método de função variável a conformação inicial a ser otimizada é o modelo final do passo anterior. Um conjunto de diferentes modelos pode ser obtido variando-se as conformações iniciais.



O PARMODEL é uma ferramenta automatizada para modelagem comparativa de estrutura 3D de proteínas. Este servidor *Web* foi projetado para integrar os principais os principais programas usados no processo de modelagem por homologia, é subdividido em quatro módulos: modelagem, avaliação, otimização e visualização.

Modelagem

Para construir os modelos, o PARMODEL conta com o MODELLER (Apêndice A). Através da interface (figura 29) que o PARMODEL fornece, o usuário deverá entrar com o arquivo contendo a sequência de aminoácidos da proteína alvo e até oito arquivos no formato pdb dos moldes. Além disso, o usuário deverá entrar com o seu endereço eletrônico (para que ele possa receber os resultados posteriormente). Embora as buscas por modelos seja implementada no MODELLER, isto não é usado, devido a existência de várias ferramentas disponíveis na Internet, como BLAST (ALTSCHUL *et al.*, 1990), PSI-BLAST (ALTSCHUL *et al.*, 1997), e FASTA (CHAO *et al.*, 1992) que fornecem alinhamentos flexíveis e rápidos envolvendo grande número de bases de dados de sequência. O usuário tem liberdade de escolher seus próprios moldes (UCHÔA *et al.*, 2004). O usuário também deve especificar o número de modelos a serem construídos para uma mesma proteína, esse número pode variar de 1 a 1000. A construção de vários modelos aumenta a possibilidade de obter boas estruturas porque as mudanças nas restrições espaciais iniciais podem gerar diferentes conformações na estrutura final. Contudo, somente o melhor deles será retornado. O valor da função objetiva do MODELLER não é uma medida absoluta, podendo ser usado somente para classificar modelos construídos para o mesmo alinhamento (SALI e BLUNDELL, 1993). Para realizar o alinhamento, o PARMODEL usa o alinhamento perfil-perfil do comando SALIGN do MODELLER-6v2 (OHLSON *et al.*, 2004). O



método perfil-perfil é ao menos 30% melhor do que o método padrão do sequência-perfil, ambos em sua habilidade de reconhecer super famílias de proteínas relacionadas, e na qualidade dos alinhamentos obtidos (HOOFT *et al.*, 1996). Após ter realizado a modelagem, o resultado do alinhamento são retornados ao usuário. O usuário deve indicar se a modelagem será realizada na presença ou ausência de ligantes. No caso da construção do modelo com ligantes, o número de ligantes presente nos moldes devem ser especificados. O usuário pode escolher três programas para avaliar os modelos: Procheck (HOOFT *et al.*, 1996), Whatcheck (LASKOWSKI *et al.*, 1993), e 3DAnalysis. A avaliação será feita apenas para o modelo melhor qualificado, de acordo com a função objetiva do MODELLER. Objetivando maior rapidez nas respostas, o PARMODEL distribui seus trabalhos a um *cluster Beowulf* de 16 nós em ambiente Linux. O PARMODEL utiliza uma biblioteca em linguagem C, MPI (*Message Passing Interface*), que controla a distribuição dos trabalhos para o *cluster*. Isto permite a paralelização da execução do MODELLER e diminui o tempo no processo de modelagem. O PARMODEL retorna os resultados das modelagens via *e-mail*, contendo um arquivo anexado compactado com a estrutura do melhor modelo, o resultado do alinhamento, uma lista com a função objetiva do MODELLER para todos modelos construídos e os resultados das análises para o melhor modelo.



Parmodel Modeling

Parmodel Modeling is a web tool that performs comparative modeling for protein structures. To build the models this tool relies on MODELLER, that is the main software used in comparative modeling. To make the server respond more rapidly, Parmodel Modeling submits jobs to a Beowulf cluster of 16 PC's running Linux. The user can build several models for a same protein (up to 1000) and the final models are sorted according to a MODELLER objective function. Parmodel Modeling returns the results of the modeling by e-mail. This e-mail contains an attached compacted file with the structure of the best model, the result of the alignment, a list with the MODELLER objective function for all built models and the results of the analysis for the best model.

User's data:

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

User's Name

Institution

E-mail address*

Necessary parameters for modeling:

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

Number of Models* From to

Sequence* (In one letter code)

Templates* (In pdb format)

Experimental Method (Templates)

☒ X-ray Diffraction

☐ NMR

☐ Modeling

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

☒ 0 ligand

☐ 1 ligand

☐ 2 ligands

☐ 3 ligands

Assessment of the model:

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

☐ Procheck

☐ Verify3D

Parameter used to choose the best model:

☒ MODELLER Objective Function

☐ Stereochemical Quality

Figura 29 - Interface do ParModel para modelagem

Avaliação

Além de realizar a avaliação da qualidade dos modelos, o PARMODEL também oferece uma interface para o módulo de avaliação executada pelo PARMODEL que apresenta três programas para avaliar estruturas de proteínas, permitindo ao usuário obter resultados de forma fácil e rápida. Estes programas são WHATCHECK, PROCHECK, (figura 30) e 3Danalysis. Estas ferramentas também podem ser usadas para a avaliação de estruturas resolvidas experimentalmente.

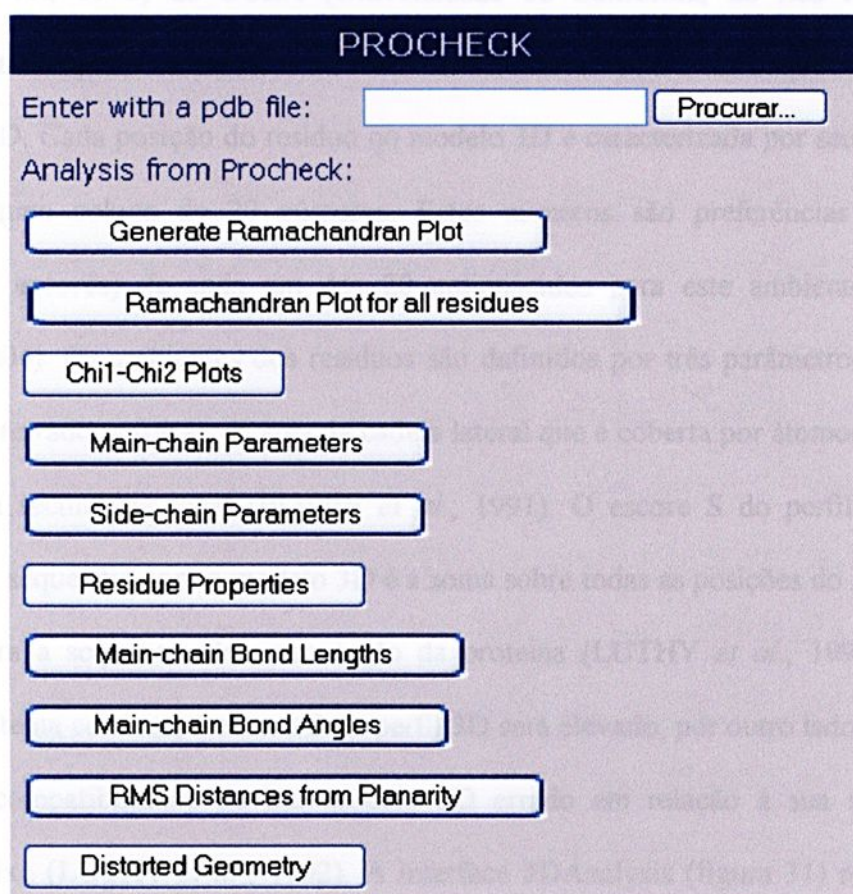


Figura 30 - Interface do ParModel para a avaliação realizada pelo PROCHECK.

O propósito do PROCHECK e WHATCHECK é (a) verificar a sintaxe do arquivo, (B) verificar a consistência de um modelo atômico com a biblioteca atual e identificar erros para

investigações futuras, (c) captar erros grosseiros nas estruturas (d) checar anormalidades estereoquímicas locais, e (e) produzir um critério global de qualidade estereoquímica (EU 3-D, 1998). Na interface do módulo de avaliação executada pelo PARMODEL, tanto PROCHECK como WHATCHECK exigem como entrada de dados o arquivo no formato pdb de uma proteína. Como resultado, o WHATCHECK gera um arquivo no formato pdf com um relatório descrevendo todas as análises realizadas, o PROCHECK gera 10 arquivos no formato pdf diferentes, com a análise individual da proteína.

3DAnalysis é um servidor *web* desenvolvido com base no Verify3D (BOWIE *et al.*, 1991; LUTHY *et al.*, 1992) de UCLA (Universidade de Califórnia, de Los Angeles). O 3DAnalysis mede a compatibilidade de um modelo tridimensional com sua sequência primária, usando um perfil 3D. Cada posição do resíduo no modelo 3D é caracterizada por seu ambiente e representada por uma coluna de 20 números. Estes números são preferências estatísticas (chamadas 3D-1D escores) de cada um dos 20 aminoácidos para este ambiente (MARTIRENOM *et al.*, 2004). Os ambientes dos resíduos são definidos por três parâmetros: a área do resíduo que está enterrado; a fração da área da cadeia lateral que é coberta por átomos polares (O e N) e a estrutura secundária local (BOWIE *et al.*, 1991). O escore S do perfil 3D para a compatibilidade da sequência com o modelo 3D é a soma sobre todas as posições do resíduo, dos 3D-1D escores para a sequência de aminoácido da proteína (LUTHY *et al.*, 1992). Para os modelos 3D de proteína corretos, o escore S do perfil 3D será elevado, por outro lado, o escore S do perfil para a compatibilidade de um modelo 3D errado em relação à sua sequência é frequentemente baixa (LUTHY *et al.*, 1992). A interface 3DAnalysis (figura 31) requer como entrada de dados um arquivo no formato pdb. Como resultado, gera um gráfico do perfil da proteína com o 3D-1D escores para cada aminoácido, o escore S do perfil 3D para sua sequência de aminoácido e o escore ideal S_{ideal} , que é calculada a partir do comprimento da proteína. Os



modelos que apresentam qualidade ruim possuem escores menores que $0,45 S_{ideal}$. Um escore igual ou acima de S_{ideal} indica que a estrutura é confiável (LUTHY *et al.*, 1992)

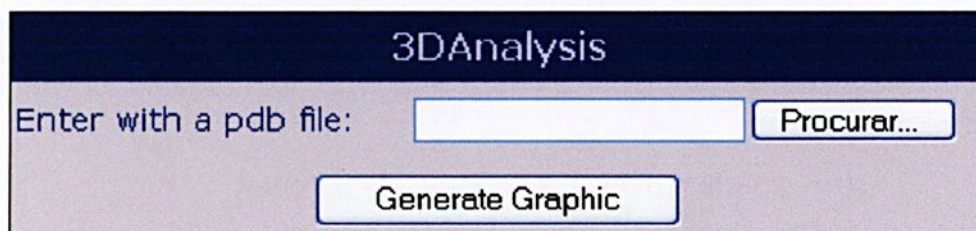


Figura 31 - Interface do ParModel para a avaliação realizada pelo 3DAnalysis.

Otimização

Os desvios na geometria da estrutura da proteína, que foram introduzidos pelo algoritmo de modelagem ao unir fragmentos rígidos, são regularizados na última etapa da modelagem pela minimização de energia usando o campo da força GROMOS96 (VAN GUNSTEREN *et al.*, 1996) do *software* Gromacs (BERENDSEN *et al.*, 1995; LINDAHL *et al.*, 2001). Campos de força empíricos são úteis para detectar partes do modelo com conformação errada. Diversos estudos (KOEHL e LEVITT, 1999; MOULT, 1996) mostram que a minimização de energia ou métodos de dinâmica molecular, em geral, não são capazes de melhorar com exatidão os modelos e são usados somente para regularizar a estrutura. Entretanto, a aplicação sucessiva de dinâmica molecular para melhorar a homologia dos modelos foram realizadas recentemente para teste de alguns casos (FLOHIL *et al.*, 2002). Encontra-se disponível uma ferramenta individual para a otimização (figura 32), a qual não está presente na ferramenta da modelagem executada pelo PARMODEL. Assim, através da interface da *web* o usuário pode submeter o arquivo no formato pdb da proteína e pode ainda alterar três parâmetros: o algoritmo usado para executar a minimização de energia, a força máxima (kJ/mol) para parar a minimização, e o número máximo de etapas da minimização.

Como resultado, a otimização executada pelo PARMODEL gera um arquivo no formato pdb com a energia minimizada.

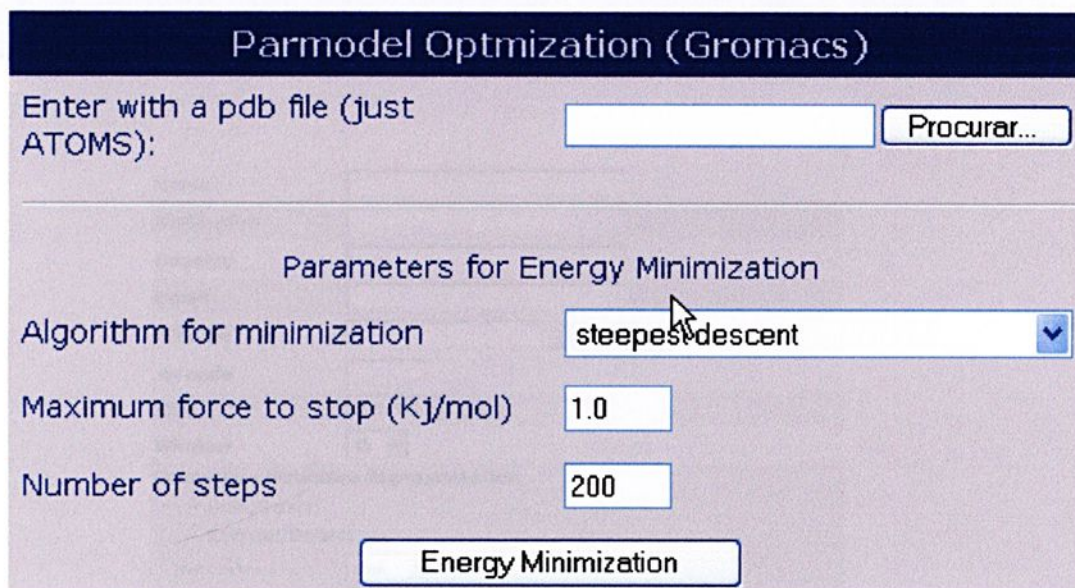


Figura 32 - Interface do ParModel para otimização do modelo.

Visualização

Para a Visualização do modelo, o PARMODEL usa o ProtGif, ferramenta que permite gerar imagens animadas para estruturas de proteínas. Para construir as imagens, o ProtGif conta com os *softwares* Raster3D (BACON E ANDERSON, 1988; MERRITT, 1994.) e Molscript (KRAULIS, 1991). Quando incluída em uma página *web*, estas ferramentas permitem que imagens animadas de moléculas sejam apresentadas por todos os padrões *browsers* sem depender de outro programa de visualização. O ProtGif gera uma representação dos elementos de estruturas secundária para as moléculas da proteína. Na interface (figura 33) as exigências mínimas de entrada são as informações sobre o usuário (*e-mail*, instituição, etc.) e o arquivo no formato pdb da proteína. O usuário pode alterar o tamanho da imagem, cores, número de *frames* que serão gerados, número de passos da rotação, e formato da imagem (ps, gif, jpeg, e vml). O usuário pode também especificar em que resíduo da proteína a imagem deve ser centrada.



HOME

SUBMIT

GALLERY

INSTRUCTIONS

CONTACTS

Name

Institution

Country

Email

PDB file

Procurar...

or code

Size

300x300

pixels

Window

45

Secondary Structure Representation

☒ Helix/Strand

☐ Cylinder/Strand

Helix Color

pink

Sheet Color

cyan

Coil Color

yellow

Still

Rotation Axis

y

☒ No

Frames

18

Time/Frames

0.15

s

☐ Yes

Rotation of the still image (0-360):

x

0

y

0

z

0

Format

☐ JPG

☒ GIF

☐ VRML

☐ PS

Center

☒ Default

☐ Number of residue:

Submit

Clear

Figura 33 - Interface do ParModel para a visualização realizada ProtGif.

88

