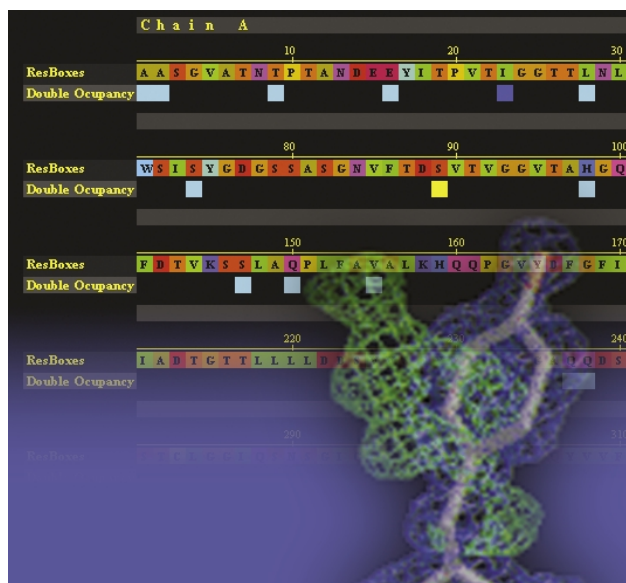


ISSN 1677-8464

Incorporação das Propriedades Rotâmeros e Ocupância em Métodos de Análise Estrutural de Proteínas

Paula Kuser Falcão¹
Christian Baudet²
Roberto Hiroshi Higa³
Goran Neshich⁴

Proteínas são moléculas que desempenham os mais diversos papéis no processo da vida. Existem proteínas estruturais, proteínas catalíticas, proteínas de transporte e armazenamento, proteínas do sistema imune, e proteínas regulatórias (Branden & Tooze, 1999; Lesk, 2001; Alberts et al., 2002; Creighton, 2003).

As proteínas são polímeros que tem uma “cadeia principal” de unidades repetitivas – os peptídeos –

com uma –“cadeia lateral” ligada a eles (Fig. 1). A cadeia principal é formada pelos átomos: nitrogênio (N), carbono alfa ($C\alpha$), carbono (C) e oxigênio (O). O carbono alfa faz a ligação com a cadeia lateral. As proteínas naturais são formadas de um repertório básico de 20 aminoácidos. Esses aminoácidos têm propriedades físico-químicas variadas em função de suas cadeias laterais. As estruturas das cadeias laterais dos 20 aminoácidos estão ilustradas na Fig. 2.

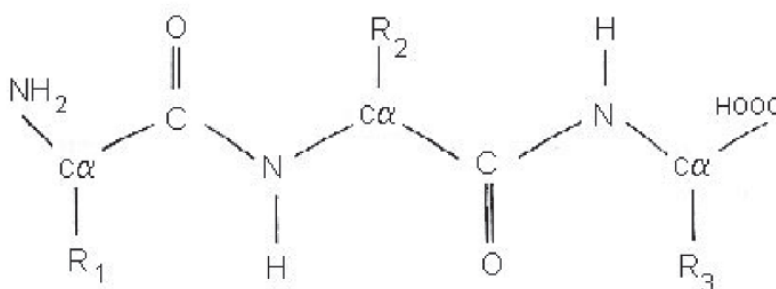


Fig. 1. Formação da cadeia polipeptídica. Em cada proteína há uma cadeia principal repetitiva e uma seqüência individual de cadeias laterais.

¹ Ph.D. em Física Aplicada, Cristalografia de Proteínas, Pesquisadora da Embrapa Informática Agropecuária, Caixa Postal 6041, Barão Geraldo - 13083-970- Campinas, SP. (e-mail: paula@cbi.cnptia.embrapa.br)

² Estudante de Engenharia da Computação, Estagiário da Embrapa Informática Agropecuária, Caixa Postal 6041, Barão Geraldo - 13083-970- Campinas, SP. (e-mail: christian@cbi.cnptia.embrapa.br)

³ M.Sc. em Engenharia Elétrica, Pesquisador da Embrapa Informática Agropecuária, Caixa Postal 6041, Barão Geraldo - 13083-970- Campinas, SP. (e-mail: roberto@cbi.cnptia.embrapa.br)

⁴ Ph.D. em Biofísica, Pesquisador da Embrapa Informática Agropecuária, Caixa Postal 6041, Barão Geraldo - 13083-970- Campinas, SP. (e-mail: neshich@cbi.cnptia.embrapa.br)

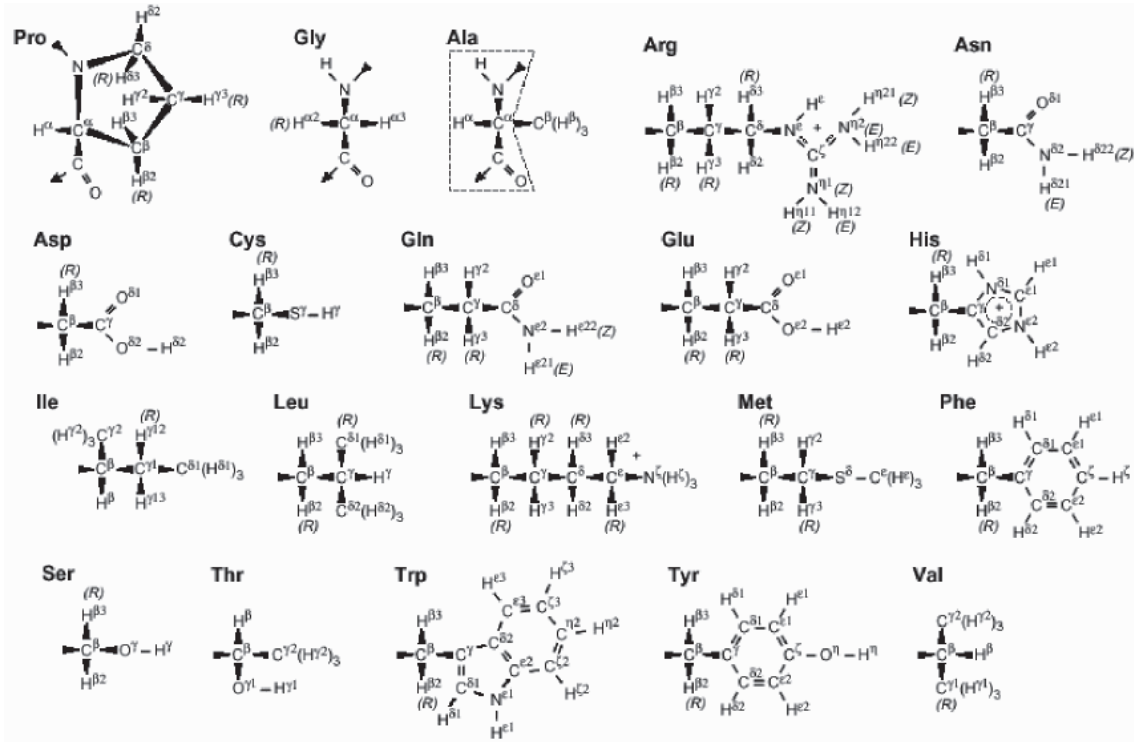


Fig. 2. Os vinte aminoácidos que ocorrem naturalmente em proteínas. Os vinte aminoácidos podem ser representados por símbolos de uma ou três letras: Ala(A)=alanina, Arg (R)=arginina, Asn(N)=asparagina, Cys(C)=cisteína, Glu(E)=ácido glutâmico, Gln(Q)=glutamina, Gly(G)=glicina, His(H)=histidina, Ile(I)=isoleucina, (Leu(L)=leucina, Lys(K)=lisina, Met(M)=metionina, Phe(F)=fenilalanina, Pro(P)=prolina, Ser(S)=serina, Thr(T)=treonina, Trp(W)=triptofano, Tyr(Y)=tirosina, Val(V)=valina. (Extraído de <http://www.le.ac.uk/biochem/mp84/teaching/lecture3.html>)

A união dos aminoácidos que gera a cadeia polipeptídica acontece quando os átomos C_{i-1} e N_i da cadeia principal se ligam através de uma ligação peptídica. A seqüência dos aminoácidos que formam a proteína é chamada de estrutura primária. A estrutura secundária está relacionada com o dobramento local da cadeia polipeptídica. A estrutura terciária é o arranjo dos elementos da estrutura secundária em três dimensões.

O dobramento da cadeia polipeptídica pode ser descrito de acordo com os ângulos de rotação interna em torno das ligações das cadeias principais (Fig. 3). As ligações entre o N e C_{α} , e entre o C_{α} e C, são ligações simples. As rotações internas em torno dessas ligações são restringidas por possíveis colisões dos átomos quando a proteína se dobra. A rotação em torno da ligação N - C_{α} é representada pelo ângulo ϕ , e a rotação em torno de C_{α} - C é representada por ψ . Algumas combinações de ϕ e ψ produzem conformações estereoquimicamente proibidas.

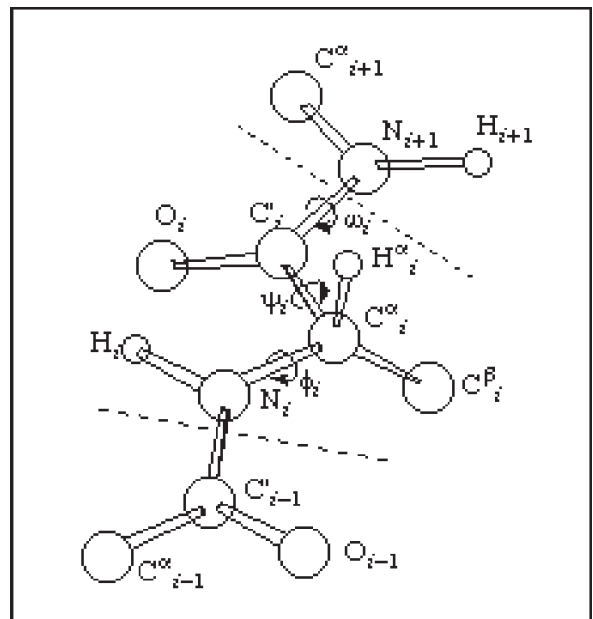


Fig. 3. Ângulos que descrevem o dobramento da cadeia polipeptídica.

A explosão da genômica, colocando à disposição dos pesquisadores seqüências genômicas inteiras, fez aumentar o número de seqüências de proteínas nos bancos de dados. No entanto, a resolução das estruturas destas proteínas não é tão rápida. Este aumento no número de seqüências está fazendo com que estruturas sejam modeladas, baseando-se nas informações de estruturas já resolvidas.

Métodos de análise de estruturas de proteínas estão se tornando fundamentais para entender o funcionamento destas macromoléculas, para a modelagem de novas moléculas e para a modelagem de pequenas moléculas complexadas às proteínas. Grande parte das reações que as proteínas realizam resultando no funcionamento das mesmas, resulta de interações entre os resíduos que as formam. Procura-se então conhecer todos os tipos de mudanças que a proteína pode sofrer e todas as características estruturais desta.

O processo de construção de um modelo de uma proteína por homologia normalmente segue alguns passos:

- é feito um alinhamento das seqüências da proteína que se quer modelar e a proteína homóloga. Quanto maior a percentagem de identidade neste alinhamento, mais fácil será a modelagem;
- baseado neste alinhamento é gerado um modelo da cadeia principal da proteína;
- após obtido o modelo da cadeia principal, as cadeias laterais devem ser colocadas no modelo. No momento de colocar as cadeias laterais dos aminoácidos nos modelos, o maior problema é que muitos aminoácidos têm várias conformações favoráveis (rotâmeros).

Métodos de análise de estruturas são utilizados no estudo de modelagem. O método de análise estrutural desenvolvido no grupo de bioinformática estrutural da Embrapa Informática Agropecuária (<http://www.cbi.cnptia.embrap.br/SMS>) procura mapear o maior número de informações relativas as estruturas das proteínas com estruturas conhecidas, depositadas no Protein Data Bank (Berman et al., 2000), formando um sumário completo da proteína em estudo. O objetivo deste trabalho é descrever duas novas propriedades que serão incorporadas neste método, rotâmeros e ocupância. A colocação correta das cadeias laterais no modelo é uma das etapas principais da construção de um modelo por homologia. Portanto, todas as informações referentes às várias conformações que as cadeias laterais podem assumir em uma estrutura é de grande importância no momento da construção de um modelo tridimensional de uma proteína.

Conformação das Cadeias Laterais (Rotâmeros)

Os átomos dos 20 aminoácidos geralmente encontrados nas proteínas, são nomeados para facilitar sua identificação. O carbono central da cadeia principal do aminoácido é chamado de alfa (α). Neste carbono alfa está ligada a cadeia lateral do aminoácido e os átomos das cadeias laterais são denominados β , γ , δ , ϵ , e ζ , em ordem de distância do carbono α . Veja como exemplo dois aminoácidos mostrados na Fig. 4. Está mostrado aí somente o carbono alfa da cadeia principal e a cadeia lateral do aminoácido. Os átomos N, C e O da cadeia principal não estão presentes. No caso do aminoácido lisina, por exemplo, o carbono alfa é o primeiro na parte inferior do desenho, sendo seguido pelo carbono beta (β), carbono gama (γ), carbono delta (δ) carbono epsilon (ϵ) e finalmente o nitrogênio zeta (ζ).

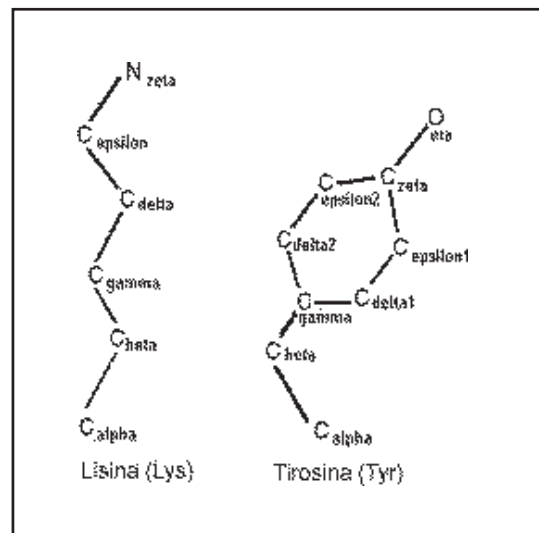


Fig. 4. Exemplos da nomenclatura dos átomos das cadeias laterais de dois aminoácidos: lisina e tirosina.

As conformações das cadeias laterais, assim como as conformações da cadeia principal como descrito acima, também são descritas por ângulos de rotação interna (ângulos de torção), denominados χ_1 , χ_2 , χ_3 , etc., como ilustrado na Fig. 5 para o caso do aminoácido lisina. Cadeias laterais diferentes, têm quantidades de graus de liberdade diferentes. A cadeia lateral de uma arginina, por exemplo, tem cinco ângulos de rotação interna. Os aminoácidos glicina e alanina não têm nenhum ângulo de rotação (Lesk, 2001).

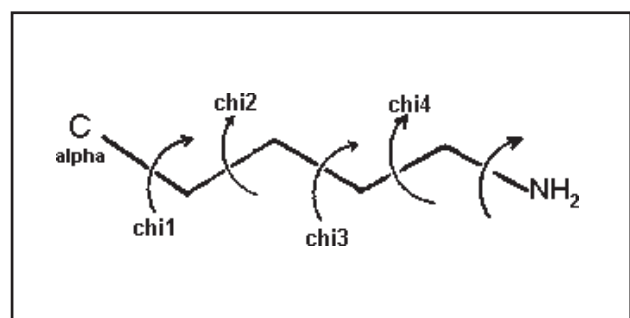


Fig. 5. Ângulos de torção da lisina.

As conformações de qualquer cadeia lateral correspondendo a diferentes combinações de valores dos ângulos chi são chamadas **rotâmeros**. Os rotâmeros são geralmente definidos como as conformações de menor energia.

Em geral, os valores de chi1 - a rotação em torno da ligação C α - C β - ficam em torno de 60° (gauche-), 180° (trans), 60° (gauche+). A análise estatística dos padrões dos ângulos conformacionais produziram as "bibliotecas de rotâmeros" (Lovell et al., 2000; Dunbrack Junior & Cohen, 1997), coleções das conformações das cadeias laterais preferenciais. A utilização das informações contidas nas bibliotecas de rotâmeros permite que os pesquisadores que estejam determinando ou modelando a estrutura de uma proteína possam usar as conformações mais prováveis, economizando tempo e produzindo uma estrutura que tem maiores chances de estar correta. Uma observação que também é útil para a modelagem de proteínas relacionadas é que as conformações das cadeias laterais tendem a ser conservadas. Ou seja, resíduos homólogos em proteínas relacionadas tem conformação similar das cadeias laterais. Isso acontece devido às interações que ocorrem com os resíduos vizinhos.

Dupla Ocupância

Devido à melhoria dos equipamentos utilizados para determinar estruturas de proteínas através de difração de raios X (detetores, luz síncrotron), está sendo possível "enxergar" mais detalhes das estruturas. Os cristais que são difratados nos raios X são formados por milhares de moléculas arranjadas simetricamente e cercadas por água (30-60% do volume total do cristal é água), o que permite a mobilidade das moléculas. Alguns resíduos, principalmente aqueles com cadeia lateral maior, podem ocupar posições diferentes nas

várias moléculas que compõem o cristal. Esta flexibilidade algumas vezes pode ser percebida nos mapas de difração que são analisados na determinação da estrutura (Fig. 6).

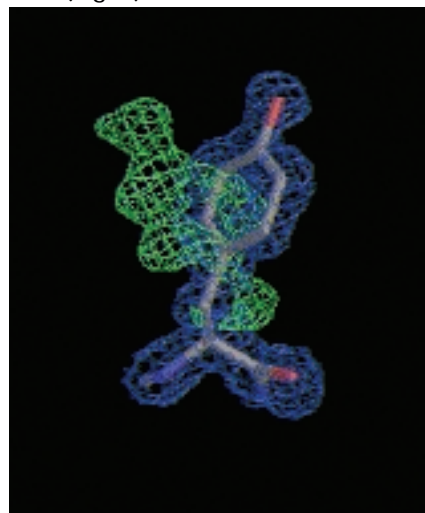


Fig. 6. Exemplo de densidade eletrônica obtida em um experimento de cristalografia por difração de raios X de uma proteína. O resíduo tirosina tem densidade eletrônica visível para ser modelado em duas posições.

Quando o cristalógrafo identifica que o resíduo tem essa flexibilidade, ele determina as coordenadas (x, y, z) para as duas possíveis posições dos átomos do resíduo, determinando inclusive o grau de ocupância (0 a 100%) nos campos 4 e 12 do arquivo de coordenadas da estrutura. Este arquivo é um arquivo padrão, chamado de arquivo pdb (Berman et al., 2000), que contém as informações da estrutura da proteína. Neste arquivo as coordenadas dos átomos devem ter o seguinte formato:

Campo	Colunas	Formato	Descrição
1	1 - 6	A6	ATOM ou HETATM
2	7 - 11	I5	número serial do átomo
-	12 - 12	1X	
3	13 - 16	A4	nome do átomo (" CA ", " ND1")
4	17 - 17	A	código alternativo (se houver)
5	18 - 20	A3	código de três letras do amino ácido
-	21 - 21	1X	
6	22 - 22	A1	código identificador da cadeia
7	23 - 26	I4	número sequencial do resíduo
8	27 - 27	A1	código de inserção (se houver)
-	28 - 30	3X	
9	31 - 38	F8.3	coordenada x do átomo
10	39 - 46	F8.3	coordenada y do átomo
11	47 - 54	F8.3	coordenada z do átomo
12	55 - 60	F6.2	valor da ocupância
13	61 - 66	F6.2	fator de temperatura
-	67 - 67	1X	
14	68 - 68	I3	número da nota de rodapé

Exemplo

```

1          2          3          4          5          6
12345678901234567890123456789012345678901234567890123456789012345678
-----
ATOM      1751  N   GLY  C  250      32.286   1.882  43.206   1.00  22.00
ATOM      1752  CA  GLY  C  250      32.365   1.086  41.969   1.00  21.39
ATOM      1753  C   GLY  C  250      31.538   1.735  40.864   1.00  20.79
ATOM      1754  O   GLY  C  250      30.621   2.527  41.152   1.00  21.58

```

Java Protein Dossier

A ferramenta *Java Protein Dossier* do pacote de programas *Sting Millennium Suite* (SMS-<http://www.cbi.cnptia.embrapa.br/SMS/>) tem como função fornecer um sumário completo de várias características estruturais das proteínas como por exemplo: estrutura secundária, fatores de temperatura, contatos entre resíduos, entropia, qualidade estereoquímica, entre vários outros. Este sumário representa várias propriedades através de códigos de cores, apresentando de uma maneira única uma informação bastante completa das propriedades de uma estrutura de macromolécula. O JPD está sempre expandindo sua lista de propriedades adicionando parâmetros considerados importantes na análise da estrutura de uma proteína.

Os dois parâmetros descritos, rotâmero e dupla ocupância são informações estruturais importantes para a análise estrutural de uma macromolécula. Estes parâmetros serão integrados ao relatório de análise gerado pelo JPD. A Fig. 7 mostra a informação de ocupância já marcados no JPD. Com isso o pesquisador visualiza imediatamente quantas e quais possibilidades de conformação aquela cadeia lateral pode assumir numa determinada molécula. Nesta Fig., a primeira linha se refere a seqüência de aminoácidos da proteína, e a segunda é a linha da dupla ocupância. As cores correspondem ao grau de ocupância que foi determinada no arquivo depositado no banco de dados (amarelo =21-40%, azul claro=41-60%, azul escuro=61-80%). Esta informação vai ser colocada na versão 3.0 do JPD, que ainda está em desenvolvimento.

Aqueles resíduos que podem assumir conformações diferentes da cadeia lateral serão marcados no JPD e as opções de rotâmero existentes serão listadas baseadas nas bibliotecas de rotâmeros (Lovell et al., 2000; Dunbrack Junior & Cohen, 1997). A implementação da marcação dos rotâmeros está em desenvolvimento.

Quando o arquivo com as coordenadas da proteína tiver algum resíduo com dupla ocupância, o JPD mostra uma linha onde aqueles resíduos ficam coloridos, permitindo uma visualização imediata dos resíduos com esta propriedade.

Mais detalhes sobre o funcionamento completo do módulo JPD poderão ser encontrados no artigo sobre esta ferramenta (Higa et al., 2003).

Considerações Finais

A demanda por modelos de estruturas para os quais não existem dados experimentais é muito grande para muitos projetos de desenho racional de drogas. No entanto, a determinação da estrutura a partir da seqüência é um problema que tem se mostrado muito difícil. A modelagem por homologia tem se mostrado muito útil para esta situação. Em modelagem por homologia, um modelo é gerado baseado na estrutura conhecida de uma proteína homóloga.

Informação sobre possíveis combinações e conformações dos aminoácidos, rotâmeros e possibilidades de localização das cadeias laterais, é de uma ajuda inestimável no momento de decidir como os aminoácidos serão dispostos ao longo da cadeia polipeptídica (Bower et al., 1997).

Nos últimos quatro anos, o número de estruturas de moléculas biológicas depositadas no banco de dados de proteínas (PDB) duplicou. Existem no momento quase 20.000 estruturas disponíveis. Este aumento de dados requer uma maneira fácil e rápida de acesso a estes dados e muita organização. Não apenas as informações presentes nos arquivos de estruturas são de interesse para a comunidade científica, mas também as informações derivadas dos dados estruturais que não se encontram nos arquivos. Além disso, há muita informação adicional que pode ser oferecida. Quanto maior o volume de informação disponível sobre as estruturas de proteínas, mais fácil será a modelagem de novas estruturas.

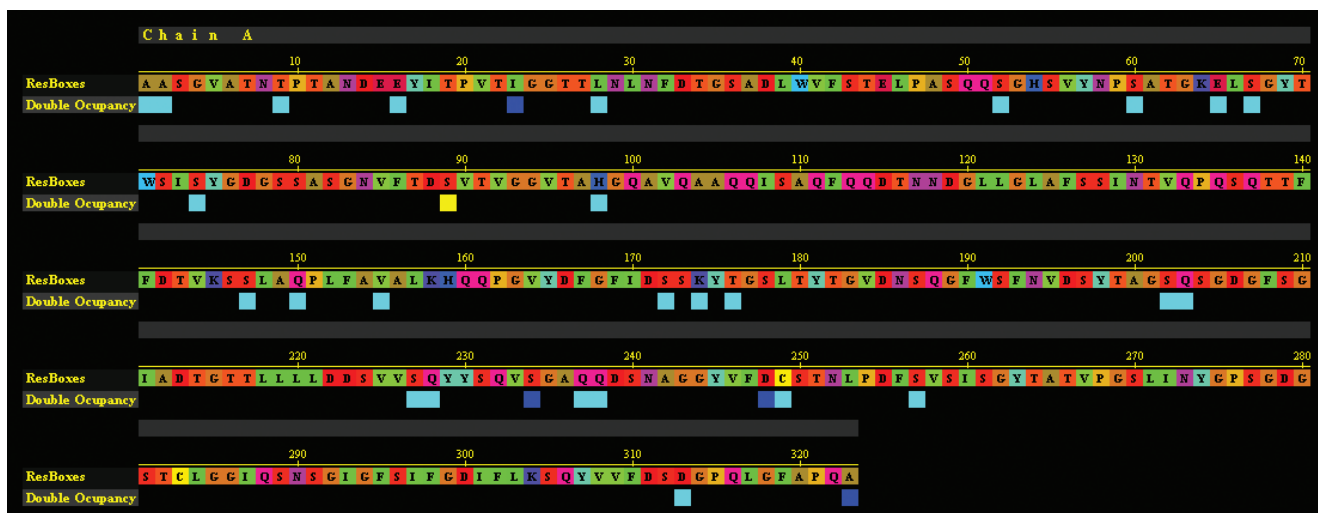


Fig.7. Sumário do JPD para uma parte da proteína de código 1bxo mostrando os resíduos com dupla ocupância marcados.

Referências Bibliográficas

ALBERTS, B.; JOHNSON, A.; LEWIS, J.; RAFF, M.; ROBERTS, K.; WALTER, P. **Molecular biology of the cell**. 4. ed. New Yor: Garland Pub., 2002. 1616 p.

BERMAN, H. M.; WESTBROOK, J.; FENG, Z.; GILLILAND, G.; BHAT, T. N.; WEISSIG, H.; SHINDYALOV, I. N.; BOURNE, P. E. The protein data bank **Nucleic Acids Research**, v. 28, p. 235-242, 2000.

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

BRANDEN, C.-U.; TOOZE, J. **Introduction to protein structure**. 2nd ed. New York: Garland Science, 1999. 410 p.

CREIGHTON, T. E. **Proteins: structures and molecular properties**. 2nd ed. New York: W. H. Freeman, 2003.

DUNBRACK JUNIOR, R. L.; COHEN, F. E. Bayesian statistical analysis of protein sidechain rotamer preferences. **Protein Science**, v. 6, p. 1661-1681, 1997.

HIGA, R. H.; BAUDET, C.; FALCÃO, P. K.; MANCINI, A.; NESHICH, G. **Apresentação gráfica de parâmetros protéicos utilizando o Java Protein Dossier**. Campinas: Embrapa Informática Agropecuária, 2002. (Embrapa Informática Agropecuária. Comunicado Técnico). No prelo.

LESK, A. M. **Introduction to protein architecture: the structural biology of proteins**. New York: Oxford University Press, 2001. 304 p.

LOVELL, S. C.; WORD, J. M.; RICHARDSON, J. S.; RICHARDSON, D. C. The penultimate rotamer library. **Proteins: Struc. Function and Genetics**, v. 40, p. 389-408, 2000.

Comunicado Técnico, 34

Embrapa Informática Agropecuária Área de Comunicação e Negócios (ACN)

Av. André Tosello, 209
Cidade Universitária - "Zeferino Vaz"
Barão Geraldo - Caixa Postal 6041
13083-970 - Campinas, SP
Telefone (19) 3789-5743 - Fax (19) 3289-9594
e-mail: sac@cnptia.embrapa.br

1ª edição
2002 - on-line
Todos os direitos reservados

Comitê de Publicações

Presidente: José Ruy Porto de Carvalho
Membros efetivos: Amarindo Fausto Soares, Ivanilde Dispatto, Luciana Alvim Santos Romani, Marcia Izabel Fugisawa Souza, Suzilei Almeida Carneiro
Suplentes: Adriana Delfino dos Santos, Fábio Cesar da Silva, João Francisco Gonçalves Antunes, Maria Angélica de Andrade Leite, Moacir Pedroso Júnior

Expediente

Supervisor editorial: Ivanilde Dispatto
Normalização bibliográfica: Marcia Izabel Fugisawa Souza
Capa: Intermídia Publicações Científicas
Editoração Eletrônica: Intermídia Publicações Científicas