

ISSN 1677-8464

Curvatura da Superfície de Proteínas no Java Protein Dossier

Paula Kuser Falcão¹
Christian Baudet²
Roberto Hiroshi Higa³
Goran Neshich⁴

Um aspecto elementar da biologia envolve o reconhecimento entre moléculas. Grande parte das informações estruturais farmacologicamente importantes residem na superfície das proteínas.

As interações entre proteínas e entre proteínas e peptídios, ácidos nucleicos ou ligantes têm um papel vital em todos os processos biológicos. Além da necessidade de serem energeticamente favoráveis, estas interações dependem fortemente das características da propriedade das superfícies que interagem. Uma ligação perfeita implica na complementaridade das superfícies. As funções das proteínas tais como catálise ou reconhecimento molecular ocorrem predominantemente na superfície da proteína ou próximo dela (Fersht, 1999; Lesk, 2001).

A informação sobre a superfície também pode fornecer detalhes sobre a estabilidade e solubilidade das proteínas, já que estas propriedades dependem de como os elementos da superfície da macromolécula interagem com o solvente e pequenos solutos presentes na solução.

Técnicas estruturais modernas como cristalografia de raios-X e ressonância nuclear magnética produzem estruturas moleculares de alta resolução. Para estudar o funcionamento das proteínas, além de utilizar as informações de estrutura, faz-se necessário ter também cálculos analíticos de parâmetros moleculares. No entanto, um grande volume de trabalho é necessário para realizar a análise de várias estruturas, compará-las e extrair a informação desejada. A explosão de técnicas recentes de modelagem de estruturas de proteínas causou um enorme interesse em métodos para modelar a interação de ligantes com os receptores. O primeiro obstáculo que deve ser eliminado, no entanto, é a localização do sítio ativo, o local onde o ligante se posiciona. Este é o tipo de problema que se pretende ajudar a resolver mapeando parâmetros nas estruturas de proteínas.

O trabalho desenvolvido no Núcleo de Bioinformática Estrutural (NBI), da Embrapa Informática Agropecuária, visa desenvolver ferramentas para a análise estrutural de moléculas biológicas. Uma das ferramentas em

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@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@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@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@cnptia.embrapa.br)

desenvolvimento é o módulo *Java Protein Dossier* (JPD) do *Sting Millennium Suíte* (SMS). O JPD oferece um tipo de visualização através de cores que leva a identificação de possíveis sítios de ligação na superfície dos alvos biológicos. Uma das propriedades que oferece indícios para identificar um possível sítio ativo é a curvatura local da superfície, que deve ser visualizada juntamente com outras propriedades como hidrofobicidade e potencial eletrostático, também mapeadas no JPD.

O objetivo deste trabalho é gerar valores de curvatura para os resíduos de todas as proteínas existentes no Protein Data Bank (PDB), depositório mundial de proteínas (Berman et al., 2000), para que estes valores possam ser incorporados no módulo JPD. Existem atualmente aproximadamente vinte mil estruturas no PDB. Gera-se todos esses valores em larga escala (highthroughput) através da utilização do programa SurfRace (Tsodikov et al., 2002) e com a construção de um *script* que possibilite acessar todo o banco de dados de proteínas e criar um novo banco de dados com os valores de curvatura para os átomos.

Definição de Superfície

No estudo das interações entre moléculas é necessário examinar suas superfícies. A característica de curvatura que se quer quantificar é a curvatura da superfície da molécula. Para entender como é definida esta superfície, três tipos de superfícies serão brevemente explicadas aqui para melhor entendimento do texto: a superfície de van der Waals, a superfície acessível ao solvente e a superfície molecular.

A superfície de van der Waals é construída a partir das esferas de van der Waals dos átomos. Estudos mostram que os átomos ficam a uma determinada distância (distância intermolecular) uns dos outros nas moléculas, sugerindo que eles devem ocupar um volume definido. As distâncias intermoleculares é interpretada visualizando os átomos como esferas. O raio dessas esferas é chamado de raio de van der Waals. A superfície de van der Waals corresponde ao envelope molecular contendo as esferas atômicas de raio igual ao raio de van der Waals. Alguns exemplos de valores de raios de van der Waals são: Carbono = 1.70Å, Cl = 1.75Å, Cu = 1.40Å, H = 1.20Å, He = 1Å, N = 1.55Å, O = 1.52Å, Pb = 2.02Å.

A superfície de van der Waals de uma molécula não é muito relevante quimicamente falando porque contém muitos vales e depressões entre átomos, que não podem ser acessados por outros átomos. Um conceito mais prático de superfície é a superfície acessível ao solvente, introduzida por Lee & Richards (1971). Esta superfície, traçada em torno de uma molécula, é descrita pelo centro de uma esfera teste (probe sphere) rolando em torno da proteína. Aumentando o raio de van der Waals de cada átomo com o valor do raio da

esfera teste, obtém-se a superfície acessível ao solvente. O valor do raio da esfera teste utilizado geralmente é o raio de uma molécula de água - 1.4 Å (Leach, 2001).

A superfície molecular (Richards, 1977; Conolly, 2003) é composta pelos dois tipos de superfície anteriormente descritos: a parte da superfície de van der Waals de cada átomo que é acessível à esfera teste, correspondendo à superfície de contato, e a face interna da esfera teste quando está, simultaneamente, em contato com mais do que um átomo (a superfície reentrante).

Estes conceitos ficam mais claros quando ilustrados na Fig. 1. A Fig. 2 representa a superfície molecular de um aminoácido para mostrar as regiões de curvatura côncavas ou convexas.

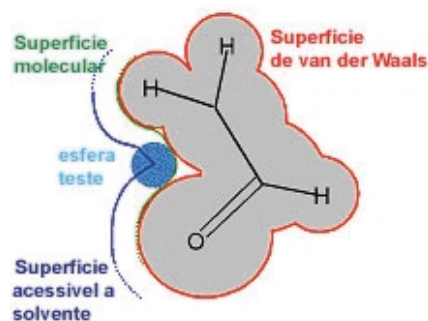


Fig. 1. Superfície de Van der Waals (vermelha), é a superfície relacionada com o raio de van der Waals dos átomos; a superfície acessível ao solvente (azul), que requer a rolagem de uma esfera imaginária de raio igual a 1.4 Å; e a superfície molecular (verde), que é composta por partes das duas outras superfícies.

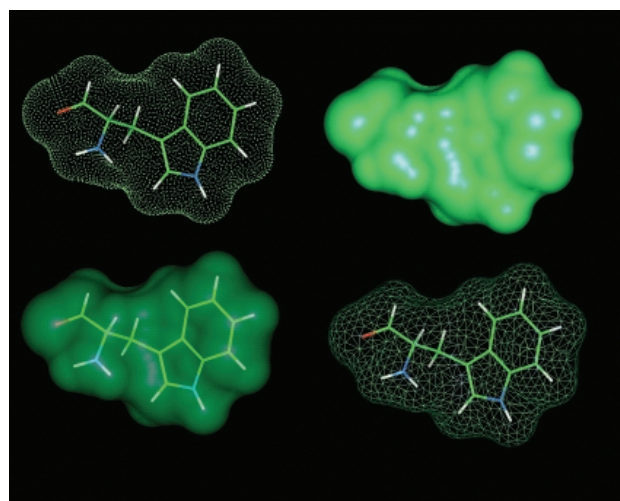


Fig. 2. Representação gráfica da superfície molecular do aminoácido triptofano, mostrando a curvatura da superfície, com regiões côncavas e convexas (Extraído de http://www.booksites.net/leach2/molecular/surfaces_1.7.gif).

Cálculo dos Valores de Curvatura com SurfRace

O programa SurfRace, desenvolvido por Tsodikov et al. (2002), é baseado no teorema de Gauss-Bonnet que relaciona a geometria de uma superfície com a sua topologia. SurfRace utiliza a idéia de correspondência entre a superfície acessível ao solvente e a superfície molecular.

O programa utiliza os arquivos do banco de dados de proteínas (PDB) como arquivo de entrada. Este arquivo de entrada contém os nomes de todos os átomos que compõem a proteína, com suas respectivas coordenadas (x , y , z). O arquivo de saída do programa SurfRace contém, além das coordenadas de todos os átomos no formato original, quatro colunas extras que se referem a: raio de van der Waals de cada átomo, área da superfície acessível, área da superfície molecular e curvatura da superfície molecular de cada átomo (Tsodikov et al., 2002).

O programa tem algumas opções de entrada:

- conjunto dos raios de Van der Waals que será utilizado no cálculo;
- nome do arquivo com código pdb;
- raio da esfera de teste em Angstroms;
- tipo de cálculo que se quer executar: área da superfície acessível a solvente, área das superfícies acessível e molecular, área das superfícies acessível e molecular e curvatura média da superfície molecular.

Utilizou-se o programa SurfRace para obter os valores de curvatura para todos os átomos do arquivo de coordenadas das proteínas. Para fazer o cálculo do valor

de curvatura para todos os arquivos de estruturas disponíveis no PDB (20.000 estruturas) o programa SurfRace rodou através de um *script* em Perl (Fig. 3). Este *script* lê os arquivos pdb um a um, cria um arquivo texto com todas as opções de entrada que o programa precisa para cada arquivo pdb e chama o programa. O programa é executado e gera o arquivo de saída com as informações de curvatura para cada arquivo pdb. Estes arquivos são colocados em um banco de dados que será consultado pelo módulo JPD para geração do sumário com as informações das proteínas que este produz. Um exemplo de um arquivo de saída é apresentado na Tabela 1.

```
#!/usr/sbin/perl
#
# surfRace.pl
#
# Chama e executa o surfRace: ./surfRace 1acb.pdb
# Para cada entrada dos arquivos *.pdb
# O executavel deve estar no mesmo diretorio
# do pdb
#
# Paula Kuser ago/2002
#

@files = `ls -c1 *.pdb`;
for ($i=0;$i<=$#files;$i++) {
    $pdb=$files[$i];
    open (INPUT,">input.txt")|| die ("Nao foi possivel abrir o arquivo");
    print INPUT "1\n";
    print INPUT "$pdb\n";
    print INPUT "1.45\n";
    print INPUT "3\n";
    print INPUT "\n";
    print INPUT "\n";
    system ("echo | ./surfRace < input.txt");
}
}
```

Fig. 3. *Script* para executar o programa SurfRace para as estruturas de proteínas do PDB.

Tabela 1. Exemplo de arquivo de saída do programa SurfRace onde a última coluna corresponde ao valor de curvatura do átomo. Valores menores que zero correspondem a regiões convexas, iguais a zero correspondem a regiões que não estão na superfície e valores maiores que zero correspondem a regiões côncavas.

	aa	x	y	z	curv
ATOM 442	OG1 THR E 62	-16.140	6.026	8.091	1.6 35.73 17.80 -0.28
ATOM 443	CG2 THR E 62	-17.377	7.925	7.035	2 21.27 19.24 -0.05
ATOM 449	OG SER E 63	-12.504	2.978	7.145	1.6 42.90 18.70 -0.34
ATOM 456	OD1 ASP E 64	-12.956	4.944	1.253	1.5 1.247 6.476 0.410
ATOM 457	OD2 ASP E 64	-10.901	4.346	1.641	1.5 6.637 11.10 0.137
ATOM 463	CG1 VAL E 65	-7.488	9.011	4.711	2 10.58 15.83 0.099
ATOM 464	CG2 VAL E 65	-6.346	9.935	2.652	2 2.823 10.92 0.235
ATOM 470	CG1 VAL E 66	-10.151	15.306	0.055	2 0 0 0
ATOM 471	CG2 VAL E 66	-11.607	13.394	0.833	2 0 0 0
ATOM 477	CG1 VAL E 67	-3.245	14.389	-0.850	2 0 0 0
ATOM 478	CG2 VAL E 67	-4.456	12.320	-0.045	2 0 0 0
ATOM 483	CB ALA E 68	-6.818	19.217	-0.767	2 0 0 0

As informações geradas com o cálculo de curvatura estão sendo incorporadas à versão 3.0 do aplicativo *Java Protein Dossier* do SMS, que está disponível na servidora beta do NBI (<http://beta.cbi.cnpia.embrapa.br/SMS>). A Fig. 4 mostra os resultados obtidos para uma proteína teste. O parâmetro é mapeado no JPD em tonalidades que vão da cor cinza até o verde, dependendo do valor da curvatura. Os aminoácidos em regiões côncavas estão pintados com a cor cinza e os aminoácidos em regiões convexas estão pintados de verde. Aqueles com valor de curvatura zero, ou seja, que estão em regiões planas da superfície da molécula, aparecem em branco.

Um excelente uso da propriedade de curvatura é para o cálculo do *docking* de duas moléculas. *Docking* é o processo de encaixar duas moléculas graficamente em três dimensões. Quando se faz o *docking* de duas estruturas de macromoléculas utilizando o parâmetro curvatura, é necessário ter superfícies geometricamente complementares que são iguais em magnitude, mas de sinais opostos. Este critério de curvatura pode ser usado como condição necessária (mas não suficiente) de complementariedade, permitindo eliminar uma série de arranjos não complementares.

Um exemplo de utilização da informação de complementariedade de superfícies pode ser visto na Fig. 5, que mostra o ligante benzamidina dentro do sítio ativo da enzima beta-tripsina (Morris et al., 1996). A posição do ligante foi determinada após vários ciclos de *docking*, até encontrar o encaixe perfeito na cavidade da superfície do sítio ativo da enzima.

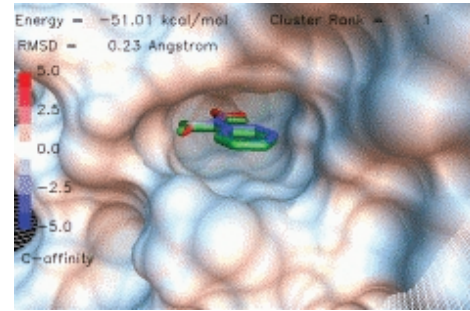


Fig. 5. Molécula benzamidina posicionada no sítio ativo da proteína beta-tripsina através da técnica de *docking*.

Conclusões

A análise obtida com o uso do JPD dá informações sobre as regiões funcionalmente importantes da proteína muito rapidamente. Com as informações iniciais de curvatura, juntamente com outras propriedades mapeadas no JPD, como por exemplo acessibilidade e entropia, já é possível ter idéias sobre os possíveis ligantes, sobre experimentos de mutação que podem ser feitos em laboratório e sobre interações com outras estruturas. O valor de curvatura das regiões também pode ajudar a identificar regiões de interesse para o *docking*, que são as cavidades e protrusões.

O programa SMS, desenvolvido no Núcleo de Bioinformática Estrutural (NBI), da Embrapa Informática Agropecuária, pode ser utilizado por qualquer pessoa

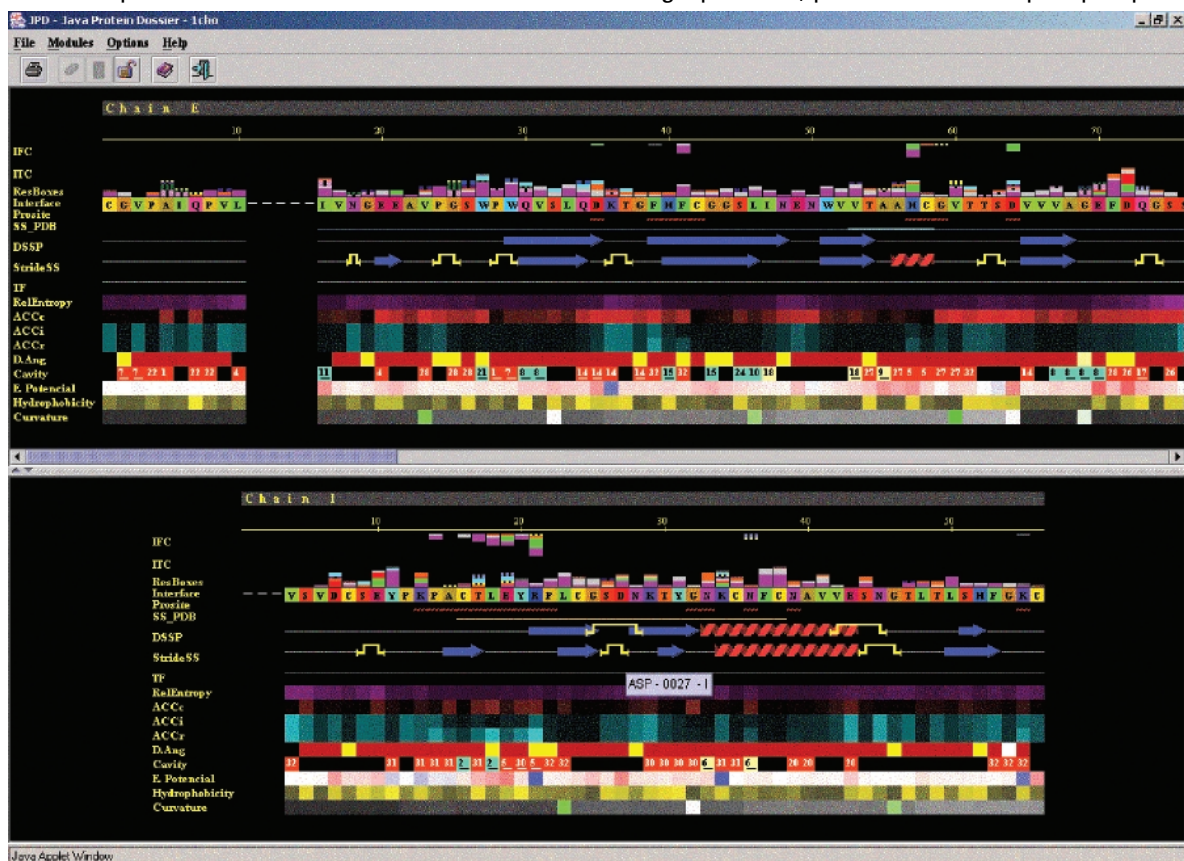


Fig. 4. Java Protein Dossier, com informações de curvatura incorporada. Informação contida na última linha (*Curvature*).

através do site <http://www.cbi.cnptia.embrapa.br>. A versão atual disponível no site é a versão 2.2; a nova versão que inclui o resultado exposto neste trabalho está disponível no site beta do NBI e será instalada para acesso externo no primeiro trimestre de 2003.

Referências Bibliográficas

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.

CONNOLLY, M. L. Molecular surface: a review. Disponível em: <<http://www.netsci.org/Science/Compchem/feature14.html>>. Acesso em: fev. 2003.

FERSHT, A. Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding. 3rd ed. New York: W.H. Freeman, 1999. 631 p.

LEACH, A. R. Molecular modelling: principles and applications. Upper Saddle River: Prentice Hall, 2001. 744 p.

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

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

MORRIS, G. M.; GOODSSELL, D. S.; HUEY, R.; OLSON, A. J. Distributed automated docking of flexible ligands to proteins: parallel applications of AutoDock 2.4. *J. Computer-Aided Molecular Design*, v. 10, p. 293-304, 1996.

RICHARDS, F. M. Areas, volumes, packing and protein structure. *Annu. Rev. Biophysics. Bioeng.*, v. 6, p. 151-176, 1977.

TSODIKOV, O. V.; RECORD JUNIOR, M. T.; SERGEEV, Y. V. A novel computer program for fast exact calculation of accessible and molecular surface areas and average surface curvature. *J. Comput. Chem.*, v. 23, p. 600-609, 2002.

Comunicado Técnico, 38

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*