

ISSN 1677-8464

Cálculo de Área Acessível por Solvente Utilizando SURFV – Definição de Interface Intramolecular pelo SMS

Roberto Hiroshi Higa¹
Adauto Luiz Mancini²
Paula Kuser Falcão³
Goran Neshich⁴



A superfície molecular da proteína é fonte importante para a compreensão de sua função, pois esta se dá basicamente através de interações intermoleculares (Branden & Tooze, 1999). Por isso, desde os anos 70, diversos métodos e ferramentas para cálculo e apresentação da superfície molecular e outras propriedades da molécula a ela mapeadas têm sido desenvolvidas (Connolly, 2002; O'Donnell, 2002).

O Sting Millennium Suite (SMS) (Structural Bioinformatics Group, 2002) compreende um conjunto de programas para análise da estrutura de proteínas acessível através da Web. Dentre suas funcionalidades, ele permite a visualização simultânea da estrutura e da seqüência da proteína e, através do módulo Protein Dossier, apresenta um conjunto de propriedades físico-químicas mapeadas para cada resíduo.

Dentre essas propriedades, um subconjunto refere-se especificamente a propriedades da superfície da molécula protéica (área acessível por solvente, curvatura e interface intermolecular), enquanto outro refere-se a propriedades que são de maior interesse para resíduos que pertencem à superfície molecular (potencial eletrostático e hidrofobicidade).

Entretanto, o Protein Dossier é uma apresentação bidimensional das propriedades da proteína (um *cartoon*) e não uma apresentação tridimensional, onde a representação da superfície é imediata. Por outro lado, tem a vantagem de permitir que diversas propriedades sejam apresentadas simultaneamente. Assim, no Protein Dossier a superfície molecular é indicada pelo parâmetro área acessível por solvente, ou seja, resíduos que participam da superfície molecular serão apresentados como possuindo área acessível por solvente diferente de zero. Além disso, para arquivos PDB (Berman et al., 2000) que representem um complexo (mais de uma molécula), o parâmetro área acessível por solvente também é utilizado para definição de interface intermolecular.

Este trabalho apresenta como o parâmetro área acessível por solvente é calculado e utilizado pelo SMS, especificamente no módulo JPD. Inicialmente é apresentada a definição de área acessível por solvente e o método utilizado para seu cálculo. Em seguida é apresentado como as interfaces intermoleculares são identificadas. Finalmente, é apresentada uma discussão dos resultados obtidos e perspectivas de trabalhos futuros.

¹ 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)

² Bacharel em Ciência da Computação, Pesquisador da Embrapa Informática Agropecuária, Caixa Postal 6041, Barão Geraldo - 13083-970 - Campinas, SP. (e-mail: adauto@cnptia.embrapa.br)

³ 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)

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)

Definição de Superfície Acessível por Solvente

O conceito de superfície acessível por solvente (*solvent accessible surface* – AS) foi proposto inicialmente por Lee & Richards (1971) como uma forma de avaliar quantitativamente a formação do núcleo hidrofóbico no processo de dobramento de proteínas. Ela descreve a área sobre a qual um contato entre a proteína e o solvente pode ocorrer. Lee & Richards (1971) consideram que cada átomo de uma molécula protéica pode ser representado por uma esfera de raio igual ao seu respectivo raio de Van der Waals, da mesma forma que uma molécula de solvente ou esfera *probe*, em geral uma molécula de água. A área acessível por solvente, foi definida como o lugar geométrico originado pelo centro da esfera *probe* à medida que esta rola sobre a superfície de Van der Waals da proteína. A Fig. 1 ilustra este conceito.

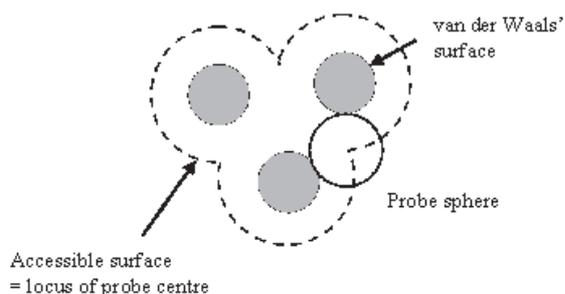


Fig. 1. Superfície acessível por solvente – AS.

Posteriormente, Richards (1977) definiu o conceito de superfície molecular (*molecular surface* – MS), como o lugar geométrico originado pela face interna da mesma esfera *probe* rolando sobre a superfície de Van der Waals da proteína (Fig. 2).

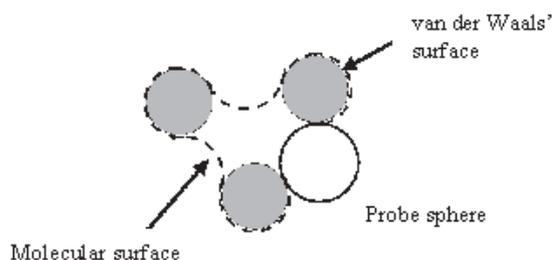


Fig. 2. Superfície molecular – MS.

Desde então, diversos algoritmos têm sido propostos para determinação da área tanto de MS quanto de AS. Para uma revisão mais completa sobre o assunto, pode-se consultar Connolly (2002). Aqui, descreve-se brevemente apenas o algoritmo desenvolvido por Shrake & Rupley (1973) para o cálculo de AS, que é o algoritmo implementado pelo programa SURFV (Sridharan et al., 1992), utilizado pelo SMS.

Cálculo de AS

1. Considere a superfície expandida (raio de Van der Waals do átomo + raio da esfera *probe*) para cada átomo da molécula protéica.
2. Para cada par de átomos, distribua um número fixo de pontos nas respectivas superfícies expandidas. Verifique quais pontos de cada um deles estão contidos na esfera definida pela superfície expandida do outro. O conjunto de pontos que não estão contidos na esfera expandida de qualquer outro átomo é considerado acessível por solvente.
3. A área acessível por solvente para cada átomo é determinada multiplicando-se o número de pontos acessíveis por solvente pela densidade de pontos dada a área de cada átomo.

Shrake & Ripley (1973) utilizaram um número de 92 pontos em seu trabalho e argumentaram que em testes por eles realizados, variando o número de pontos entre 400 e 92, a variação encontrada no valor de área da AS para 90% dos átomos da proteína lisozima foi inferior a 2%. Na implementação de SURFV, Sridharan et al. (1995) propõem uma distribuição de pontos mais uniformes sobre a superfície atômica expandida e a utilização da conectividade entre esses pontos para reduzir o tempo de computação. Eles também aumentaram o número de pontos distribuídos na superfície da esfera expandida. As opções de número de pontos disponíveis para utilização com SURFV são 122, 482, 1922 e 7682. Uma avaliação deste algoritmo e uma comparação com o algoritmo original de Lee & Richards (1971), utilizando implementações do pacote CCP4 (Collaborative Computational Project Number 4, 1994), podem ser encontradas em Briggs (2000).

Utilização de SURFV para Cálculo da Área da AS e Identificação de Interface

SURFV é um programa desenvolvido no Laboratório de Barry Honig na Columbia University, escrito em Fortran, e que implementa o algoritmo de Shrake & Ripley (1973). Ele é utilizado em forma de linha de comando, tendo a seguinte sintaxe e opções de entrada:

```
surfV [-s nlevel] (resolução) (raio da esfera probe)
(sizefile) (arquivo PDB) 1(0) 1(0) 1(0)
```

onde:

- resolução indica a granularidade da malha de pontos gerada sobre a superfície atômica expandida. Ele pode assumir valores inteiros de 1 a 4 (1 para 122 pontos; 2 para 482 pontos; 3 para 1922 pontos e 4 para 7682 pontos). No SMS, utiliza-se o valor 2 para esta opção;
- raio da esfera *probe* é o raio a ser considerado para a esfera *probe*. Geralmente, o valor utilizado é de 1.4 Å;
- *sizefile* indica o arquivo que contém o raio de Van der Waals para cada átomo no arquivo PDB;

- arquivo PDB indica o arquivo PDB a ser processado;
- o quinto parâmetro pode assumir 0 (zero) ou 1 (um). Se o seu valor for diferente de zero, será gerado um arquivo de nome "arquivo PDB".res com a área acessível por solvente para cada resíduo;
- o sexto parâmetro tem uma semântica similar. Se o seu valor for diferente de zero, será gerado um arquivo de nome "arquivo PDB".atm com a área acessível por solvente para cada átomo;
- o sétimo parâmetro também pode assumir 0 (zero) ou 1 (um). Por default, para cálculo das áreas acessíveis por solvente, os átomos de hidrogênio não são considerados, mesmo se presentes no arquivo PDB. Se o valor deste parâmetro for diferente de zero, os átomos de hidrogênio passam a ser considerados no cálculo.

Os arquivos res e atm, por default são multicolunas. A opção "-s nlevel" permite que se especifique o número de colunas nos arquivos de saída.

O SMS, especificamente para o Protein Dossier, utiliza o SURFV para:

- cálculo da área da **AS** para cada molécula isolada do arquivo PDB. Os resíduos da molécula que possuem área acessível por solvente são considerados como fazendo parte da superfície da molécula;
- caso o arquivo PDB possua mais de uma molécula, cálculo da área da **AS** para o complexo como um todo.

A interface intermolecular é determinada fazendo-se a diferença entre a área acessível para solvente no caso da molécula isolada e no caso em que ela está em complexo. Os resíduos da superfície de uma molécula que, quando em complexo, perdem área acessível para solvente são considerados como pertencentes à interface intermolecular.

A Fig. 3 ilustra como esses dados são apresentados pelo módulo Protein Dossier do SMS.

As duas últimas linhas apresentam de forma qualitativa os parâmetros "Acc. Complex" e "Acc. Isolation", onde valores em verde mais brilhante significa mais área acessível por solvente. Para os resíduos que apresentam diferença deste valor para os casos em complexo e em isolamento, uma linha vermelha abaixo da especificação do resíduo indica que ele participa da interface intermolecular.

Discussão e Trabalhos Futuros

A superfície molecular de proteínas pode fornecer importantes informações sobre seu funcionamento. Além disso, a verificação de alguns parâmetros mapeados por resíduo torna-se mais importante se estes resíduos estão na superfície da proteína. O SMS utiliza o conceito de área acessível para solvente para determinar se um resíduo está na superfície da proteína ou se faz parte de seu núcleo hidrofóbico e, no caso de um complexo, para determinar a interface intermolecular. Para isso, é utilizado o programa SURFV.

Embora, o SURFV tenha sido utilizado com sucesso no desenvolvimento do SMS, ele apresenta algumas limitações de uso:

- para um arquivo PDB com mais de uma molécula, ele não tem opção para utilização de apenas uma ou um subconjunto dessas moléculas no cálculo da área acessível por solvente. Isso fez com que, no desenvolvimento do programa para determinação de interface molecular, fosse necessário criar um *script* que criasse um arquivo PDB para cada molécula no arquivo PDB original para então utilizar o SURFV;
- ele não calcula outros parâmetros interessantes da superfície como curvatura e cavidades;
- ele é um programa *stand alone* escrito em Fortran e que não se integra com a biblioteca de programas C++ do SMS.

Por isso, no futuro, pretende-se reimplementar em C++ o algoritmo de Shrake & Rupley (1977) na biblioteca do SMS. Além disso, por se tratar de um algoritmo já

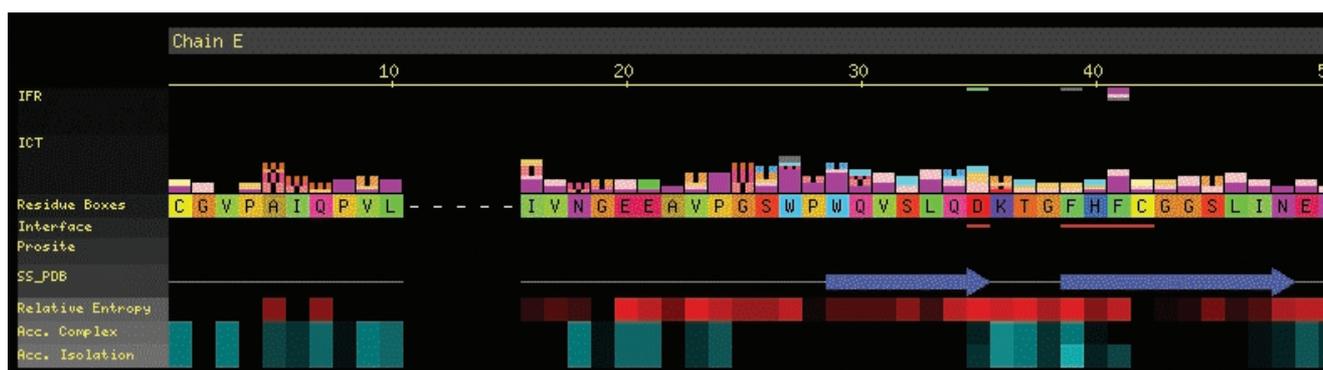


Fig. 3. Apresentação da área acessível por solvente e interface intermolecular no módulo Protein Dossier do SMS.

antigo, pretende-se estudar a possibilidade de implementar esta funcionalidade com um algoritmo mais moderno (Connolly, 2002).

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.

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

BRIGGS, P. Comparison of SURFACE and AREAIMOL for accessible surface area calculations. **CCP4 Newsletter on Protein Crystallography**, n. 38, Apr. 2000. Disponível em: <http://www.ccp4.ac.uk/newsletter38/03_surfarea.html>. Acesso em: 20 dez. 2002.

COLLABORATIVE COMPUTATIONAL PROJECT NUMBER 4. The CCP4 Suite: programs for protein crystallography. **Acta Cryst.**, v. D50, p. 760-763, 1994.

CONNOLLY, M. L. **Molecular surface**: a review. Disponível em: <<http://www.netsci.org/Science/Compchem/feature14.html>>. Acesso em: 20 dez. 2002.

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

O'DONNELL, T. J. **The scientific and artistic uses of molecular surfaces**. Disponível em: <http://www.cs.uic.edu/~tj/netsci_surf.html>. Acesso em: 20 dez. 2002.

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

SHRAKE, A.; RUPLEY, J. A. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. **J. Mol. Biol.**, v. 79, p. 351-371, 1973.

SRIDHARAN, S.; NICHOLLS, A.; HONIG, B. A new vertex algorithm to calculate solvent accessible surface areas. **Biophysics J.**, v. 61, p. A174, 1992.

STRUCTURAL BIOINFORMATICS GROUP. **Sting Millennium Suite**. Disponível em: <<http://www.nbi.cnpq.br/>>. Acesso em: 21 dez. 2002.

Comunicado Técnico, 36

**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@cnpqia.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