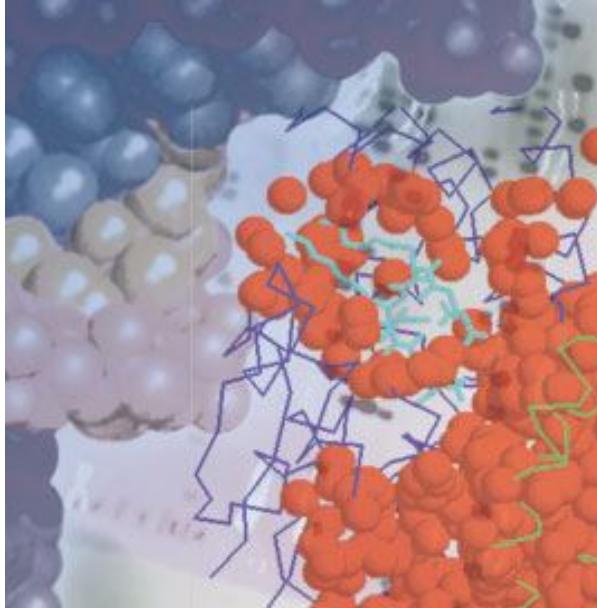


Comunicado 67

Técnico

Agosto, 2005
Campinas, SP

ISSN 1677-8464



Identificando *Pockets* na Superfície Protéica usando o Java Protein Dossier - JPD

Michel Eduardo Beleza Yamagishi¹
Paula Regina Kuser Falcão²
Roberto Hiroshi Higa³
Edgard Henrique dos Santos⁴
José Gilberto Jardine⁵
Ivan Mazoni⁶
Stanley Robson de Medeiros Oliveira⁷
Adauto Luiz Mancini⁸
Goran Neshich⁹

As proteínas desempenham importantes atividades nos organismos vivos. Elas podem atuar como enzimas, hormônios e, até mesmo, como anticorpos na defesa do organismo (Alberts et al., 2002). Para realizarem de forma satisfatória as suas funções, precisam ter uma conformação tridimensional específica (Branden & Tooze, 1999). Uma analogia útil para a compreensão é o paradigma da chave-fechadura, ou seja, uma "chave" deve possuir um formato tridimensional bem definido para abrir (função) uma determinada "fechadura". Embora nessa analogia haja a tendência de associar às proteínas apenas o papel de "chave", elas também podem ser "fechaduras".

Pensando-se em "fechaduras", é possível compreender a importância de se identificar bolsos ou grandes depressões, conhecidas como *pockets* (Edelsbrunner et al., 1998), na superfície protéica, pois, ao identificá-las, podem ser calculados o seu volume, área e forma tridimensional, e com base nessas informações, torna-se possível, por exemplo, a procura de possíveis candidatos a "chave", denominados ligantes (proteínas ou outras moléculas). Estes devem ter

volume e formato compatíveis com os do *pocket* em questão.

Pockets

Se por um lado o conceito de *pockets* é facilmente compreendido; por outro lado, a sua definição matemática não é trivial. São necessárias ferramentas avançadas de Geometria Computacional.

Uma proteína é uma cadeia de aminoácidos. Estes, por sua vez, são formados por um núcleo principal, comum a todos os aminoácidos, e por uma cadeia lateral, responsável pela diferença entre eles. Uma vez determinada a estrutura tridimensional de uma proteína, esta pode ser representada pelas coordenadas cartesianas (x,y,z) de cada átomo da proteína. Matematicamente, portanto, uma proteína pode ser pensada como um conjunto de esferas (Lee & Richards, 1971), que podem se interpenetrar, cujos centros são dados pelas coordenadas (x,y,z) e os raios são, em geral, os de van der Waals de cada átomo.

¹ Doutor em Matemática Aplicada, Pesquisador da Embrapa Informática Agropecuária, Caixa Postal 6041, Barão Geraldo - 13083-970 - Campinas, SP.

² Ph.D. em Física Aplicada - Cristalografia de Proteínas, Pesquisadora da Embrapa Informática Agropecuária, Caixa Postal 6041, Barão Geraldo - 13083-970 - Campinas, SP.

³ M.Sc. em Engenharia Elétrica, Pesquisador da Embrapa Informática Agropecuária, Caixa Postal 6041, Barão Geraldo - 13083-970 - Campinas, SP.

⁴ Bacharel em Ciência da Computação, Técnico de Nível Superior da Embrapa Informática Agropecuária, Caixa Postal 6041, Barão Geraldo - 13083-970 - Campinas, SP.

⁵ Doutor em Engenharia de Alimentos, Pesquisador da Embrapa Informática Agropecuária, Caixa Postal 6041, Barão Geraldo - 13083-970 - Campinas, SP.

⁶ Bacharel em Tecnologia em Processamento de Dados, Técnico de Nível Superior da Embrapa Informática Agropecuária, Caixa Postal 6041, Barão Geraldo - 13083-970 - Campinas, SP.

⁷ Doutor em Ciência da Computação, Pesquisador da Embrapa Informática Agropecuária, Caixa Postal 6041, Barão Geraldo - 13083-970 - Campinas, SP.

⁸ Bacharel em Ciência da Computação, Pesquisador da Embrapa Informática Agropecuária, Caixa Postal 6041, Barão Geraldo - 13083-970 - Campinas, SP.

⁹ Ph.D. em Biofísica, Pesquisador da Embrapa Informática Agropecuária, Caixa Postal 6041, Barão Geraldo - 13083-970 - Campinas, SP.

Há diferentes alternativas para o cálculo de volume e área de proteínas. Métodos numéricos para cálculos aproximativos, porém mais rápidos, como por exemplo o DCLM (Eisenhaber et al., 1995), e métodos analíticos (Connolly, 1983). Contudo, apenas em 1994, ao utilizar o *Alpha shapes*, que é um subcomplexo da triangulação de *Delaunay*, (Edelsbrunner & Mücke, 1994) para o cálculo de volume e área de proteínas, o trabalho pioneiro de Edelsbrunner et al., 1995, possibilitou a definição matemática dos *pockets*.

Matematicamente, utilizando-se *Alpha shapes*, as depressões na superfície protéica podem ser classificadas como: pequena depressão, *pocket* e cavidade (Fig. 1). Esta última trata-se de uma região dentro da proteína sem acesso ao exterior e não ocupada por nenhum átomo da proteína. Já a distinção entre "pequena depressão" e *pocket* se dá através do conceito de *Discrete Flow* (Edelsbrunner et al., 1998) que não será aqui descrito por estar fora do escopo deste trabalho. Entretanto, para dar uma idéia intuitiva do conceito ver a Fig. 2. Sabe-se que os triângulos vazios na triangulação de *Delaunay* que não fazem parte do *Alpha shape* podem ser classificados em obtusos e agudos. Um triângulo obtuso é aquele cujo seu maior ângulo interno é maior que 90 graus. Já num triângulo agudo, seu maior ângulo é menor que 90 graus. Um triângulo obtuso vazio pode ser considerado como uma "fonte" do espaço vazio que flui para os triângulos vizinhos. Em contrapartida, um triângulo agudo é um "ralo" para onde os triângulos obtusos vizinhos fluem. Uma depressão será classificada como *pocket* se existir pelo menos um triângulo agudo ("ralo") entre os triângulos vazios; caso todos os triângulos sejam obtusos ("fonte"), tem-se uma "pequena depressão".

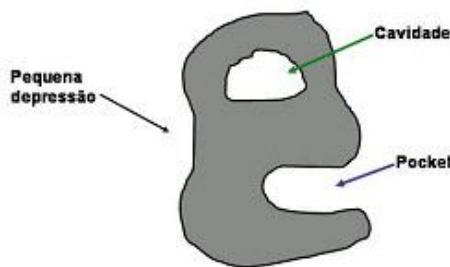


Fig. 1. A diferença entre pequena depressão, *pocket* e cavidade.

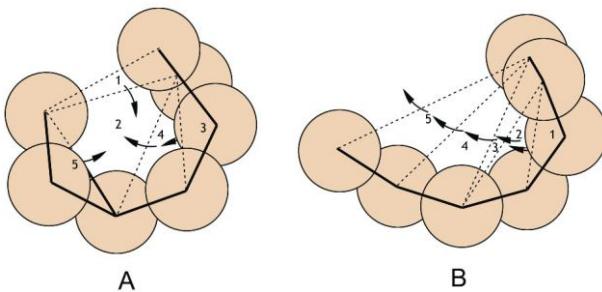


Fig. 2. Exemplo em 2D. À direita, tem-se um "ralo", isto é, um triângulo agudo (número 2) para onde os seus vizinhos fluem. À esquerda, todos os triângulos são obtusos e fluem para o infinito. No primeiro caso, tem-se um *pocket*, enquanto, no segundo, uma pequena depressão.

Com esse ferramental matemático, tornou-se possível identificar automaticamente os *pockets* na superfície protéica.

Identificando e visualizando *pockets* através do JPD

O STING (Higa et al., 2003, 2004; Neshich et al. 2004a) é um *software* para análise estrutural de proteínas acessível gratuitamente via web (<http://www.cbi.cnptia.embrapa.br>), composto por diferentes módulos. O Java Protein Dossier (Neshich et al., 2004a), ou simplesmente JPD, é um dos módulos do STING e seu objetivo é apresentar diversos parâmetros estruturais para cada aminoácido da proteína de forma visual a fim de facilitar o trabalho de comparação e análise.

A identificação dos *pockets* no JPD é feita através de um *software* do pacote gratuito denominado *Proshape*, que é predominantemente escrito em FORTRAN (<http://www.cs.ucdavis.edu/~koehl/ProShape/download.html>) e utiliza a teoria de *Alpha shapes*. Este pacote é composto por três *softwares*: *Tetrafor* (cálculo da triangulação ponderada de *Delaunay* tridimensional), *Progeom* (cálculo da superfície e do volume das proteínas, e suas respectivas derivadas com respeito às coordenadas cartesianas) e *Pocket* (identifica *pockets* e cavidades nas proteínas). O autor do código é o Dr. Patrice Koehl membro do *Biogeometry Group* (<http://biogeometry.cs.duke.edu>).

Embora o *software* *Pocket* identifique os *pockets* e cavidades nas proteínas, há uma limitação com relação ao número máximo de aminoácidos permitidos. Na versão original, este número era da ordem de 500 resíduos. Como o STING calcula todos os parâmetros estruturais, a maioria por cadeia, para todas as estruturas depositadas no Protein Data Bank (PDB) (Berman et al., 2000), e há uma grande parte com mais de 500 resíduos, foi necessário, adaptar o código original para tratar proteínas com, pelo menos, 5.000 aminoácidos. Além disso, modificou-se o código para processar cada proteína por cadeia. Com essas alterações, o código resultante foi compilado e rodado em todos os arquivos do pdb, gerando a base de dados para a identificação dos *pockets* no JPD.

Aplicações

Há diferentes aplicações para a identificação de *pockets* em proteínas. Talvez a mais imediata seja descobrir quais átomos formam o *pocket* (Fig. 3 e 4). No caso de haver um ligante dentro do *pocket*, a lista de átomos que o forma possibilita analisar que forças físico-químicas estão contribuindo para a interação entre a proteína e o ligante (Fig. 5).

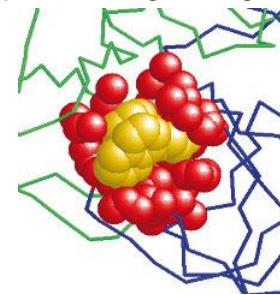


Fig. 3. Proteína 1hiv.pdb. Em vermelho os átomos dos resíduos que formam o pocket. Em amarelo, o ligante. Este *pocket* tem volume igual a 305,5 Å^3 e área igual a 203,12 Å^2 . Imagem gerada pelo JPD.

RESIDUE NUMBER	ATOM NUMBER	ATOM	RESIDUE	CHAIN
0008	67	CG	ARG	A
0008	68	CD	ARG	A
0008	69	NE	ARG	A
0008	71	NH1	ARG	A
0008	72	NH2	ARG	A
0023	181	CD1	LEU	A
0023	182	CD2	LEU	A
0025	197	OD1	ASP	A
0025	198	OD2	ASP	A
0027	207	CA	GLY	A
0027	208	C	GLY	A
0027	209	O	GLY	A
0029	215	N	ASP	A
0029	220	CG	ASP	A
0029	221	OD1	ASP	A
...

Fig. 4. As 15 primeiras linhas da lista de átomos que compõem o *pocket*. Cada átomo está associado a um aminoácido, e este a uma cadeia da proteína.

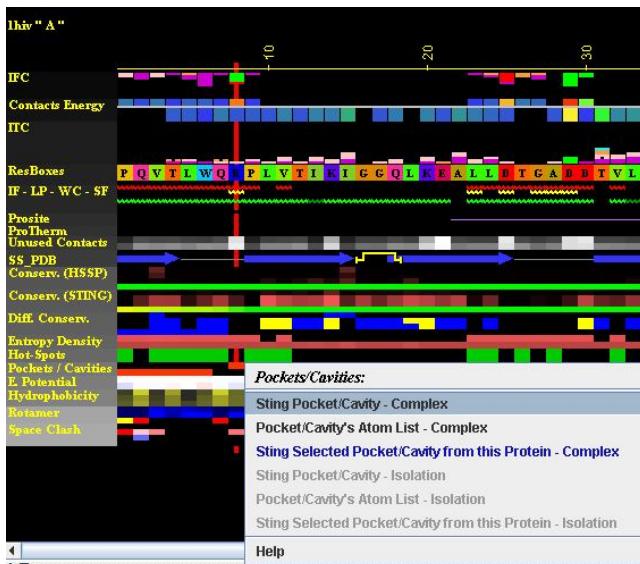


Fig. 5. Módulo JPD mostrando parte da sequência de aminoácidos da proteínas HIV-1 Protease (código pdb 1hiv) e vários parâmetros que são utilizados para a análise estrutural das proteínas. Usando o JPD é fácil visualizar o *pocket*, obter a lista de átomos que o formam e os valores analíticos do volume e área.

Uma outra aplicação interessante da identificação de *pockets* na superfície protéica foi apresentada no ISMB 2004 (Neshich et al., 2004b). Neste trabalho, procurou-se responder a seguinte pergunta: existe um conjunto de parâmetros que podem definir os resíduos principais do sítio ativo da proteína? Usando-se o *Select Tool* do JPD foi possível responder afirmativamente a esse questionamento. Como exemplo, pode-se citar o caso da enzima HIV-1 integrase (cuja estrutura no PDB recebeu o código 1biu) que é essencial para o ciclo vital do vírus HIV. Esta enzima é responsável por catalisar a inserção do genoma viral no cromossomo da célula hospedeira, sendo

um possível alvo para o *design* de drogas antivirais. Estudos mostraram que mesmo as mutações mais conservadoras nos resíduos D64, D116 e E152 aboliram completamente a atividade catalisadora da enzima. Filtrando-se os resíduos mais conservados, com potencial eletrostático negativo e que pertenciam a um *pocket* obtém-se exatamente os três resíduos principais D64, D116 e E152 (Fig. 6 e Fig. 7).

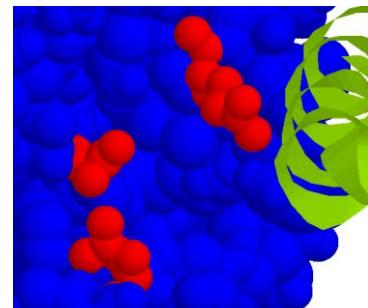


Fig. 6. Representação CPK dos átomos da proteína 1biu que estão na vizinhança dos resíduos principais para a atividade da enzima: D64, D116 e E152. Estes resíduos estão destacados em vermelho. Imagem gerada pelo JPD.

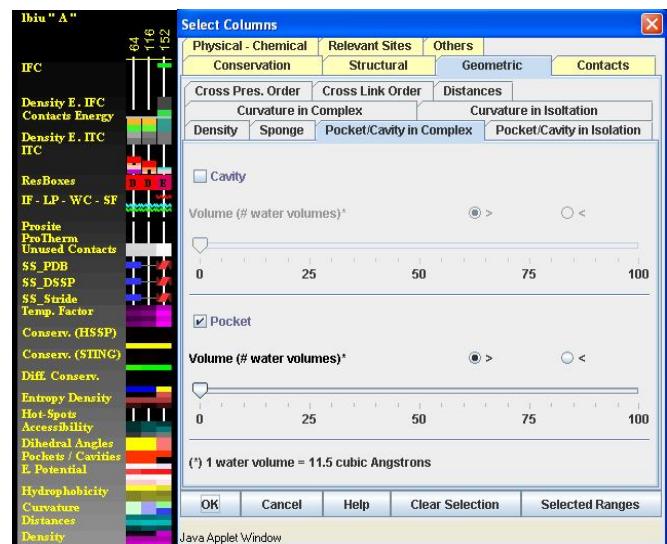


Fig. 7. Utilizando-se o *Select Tool* do JPD é possível filtrar apenas os resíduos principais para a atividade da enzima.

Conclusões

A inserção da identificação de *pockets* no JPD abriu novas possibilidades na análise da relação entre estrutura e função das proteínas. Além disso com os dados sobre volume e área dos *pockets* é possível identificar possíveis ligantes em um banco de dados de pequenas moléculas.

Referências bibliográficas

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

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.

CONNOLLY, T. H. Analytical molecular surface calculation. *J. Appl. Cryst.*, v. 16, p. 548-558, 1983.

EDELSBRUNNER, H.; FACELLO, M.; FU, P.; LIANG J. Measuring proteins and voids in proteins. Annual Hawaii Inter. Conf. on System Sciences, 28., 1995, Maui, Hawaii. **Proceedings of the 28th...** Maui: IEEE Computer Society, 1995. p. 256-264.

EDELSBRUNNER, H.; FACELLO, M.; LIANG, J. On the definition and the construction of pockets in macromolecules. *Discrete Applied Mathematics*, v. 8, p. 83-102, 1998.

EISENHABER, F.; LIJNZAAD, P.; ARGOS, P.; SANDER, C.; SCHAFER, M. The double cubic lattice method: efficient approaches to numerical Integration of surface area and volume and to dot surface contouring of molecular assemblies. *J. Comp. Chem.*, v. 16, n. 3, p. 273-284, 1995.

EDELSBRUNNER, H.; MÜCKE, E. P. Three-dimensional alpha shapes. *ACM Trans. Graphics*, v. 13, p. 43-72, 1994.

HIGA, R. H.; TOGAWA, R. C.; MANCINI, A.; FALCÃO, P. R. K.; YAMAGISHI, M. E. B.; PAPPAR JUNIOR, G.; TORRES, V. W.; CAMPOS, T. F. E.; BAUDET, C.; NESHICH, G. STING Millennium: a web-based suite of programs for comprehensive and simultaneous analysis of protein structure and sequence. *Nucleic Acids Research*, v. 31, n. 13, p. 3386-3392, 2003.

HIGA, R. H.; TOGAWA, R. C.; MONTAGNER, A. J.; PALANDRANI, J. C. F.; OKIMOTO, I. K. S.; KUSER, P. R.; YAMAGISHI, M. E. B.; MANCINI, A. L.; NESHICH, G. STING Millennium Suite: integrated software for extensive analyses of 3D structures of proteins and their complexes. *BMC Bioinformatics*, v. 5, n. 1, p. 107, 2004. Disponível em: <<http://www.biomedcentral.com/1471-2105/5/107>> Acesso em: 25 ago. 2005.

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

NESHICH, G.; MANCINI, A.; BAUDET, C.; YAMAGISHI, M. E. B.; FALCÃO, P. R. K.; FILETO, R.; ROCCHIA, W.; PINTO, I.; MONTAGNER, A.; PALANDRANI, J.; KRAUCHENKO, J.; TORRES, R. C.; SOUZA, S.; TOGAWA, R. C.; HIGA, R. H. Java Protein Dossier: a novel Web based data visualization tool for comprehensive analysis of protein structure. *Nucleic Acid Research*, v. 32, p. W595-W601, 2004a.

NESHICH, G.; MANCINI, A.; FALCÃO, P. R. K.; FILETO, R.; YAMAGISHI, M. E. B.; HIGA, R. H. Gold STING: studying protein stability and folding by looking at an extensive DB of the structure descriptors. In: INTERNATIONAL CONFERENCE ON INTELLIGENT SYSTEMS FOR MOLECULAR BIOLOGY, 12th, Glasgow, 2004. [Proceedings...]. [Glasgow: s. n., 2004b].

Comunicado Técnico, 67

Ministério da Agricultura, Pecuária e Abastecimento



Embrapa Informática Agropecuária
Área de Comunicação e Negócios (ACN)
Endereço: Caixa Postal 6041 - Barão Geraldo
13083-970 - Campinas, SP
Fone: (19) 3789-5743
Fax: (19) 3289-9594
e-mail: sac@cnptia.embrapa.com.br

1^a edição on-line - 2005

© Todos os direitos reservados.

Comitê de Publicações Presidente: Kleber Xavier Sampaio de Souza.
Membros Efetivos: Adriana Farah (secretária), Ivanilde Dispato, Luciana Alvim Santos Romani, Marcia Izabel Fugisawa Souza, Renato Fileto.
Suplentes: José Iguelmar Miranda, Laurimar Gonçalves Vendrúsculo, Maria Goretti Gurgel Praxedes, Silvio Roberto Medeiros Evangelista.

Expediente Supervisor editorial: Ivanilde Dispato
Normalização bibliográfica: Marcia Izabel Fugisawa Souza
Editoração eletrônica: Área de Comunicação e Negócios