

Utilização do Software GRASP para gerar Arquivo de Coordenadas com Valores de Potencial Eletrostático

Paula Kuser Falcão¹
Christian Baudet²
Adauto Luiz Mancini³
Roberto Hiroshi Higa⁴
Goran Neshich⁵

Muitas das atividades de um organismo estão baseadas nas atividades das proteínas mediadas pelas suas interações com o meio: água, sal, ácidos nucleicos, outras proteínas e inúmeras moléculas presentes nos sistemas. Essas interações surgem a partir de forças não covalentes (Creighton, 1993). Entre as forças não covalentes existentes está a interação eletrostática. Ligações eletrostáticas (iônicas) resultam da atração eletrostática entre dois grupos ionizados com cargas opostas, tais como os grupos carboxila ($-\text{COO}^-$) e amina ($-\text{NH}_3^+$). Na água, essas ligações são fracas, mas num ambiente hidrofóbico tal como o contato proteína-proteína, elas tornam-se mais fortes.

Por que as interações eletrostáticas em uma proteína são importantes? São importantes porque: a) interações eletrostáticas são forças atrativas de longo alcance; b) elas podem acelerar a velocidade de associação das moléculas; c) proteínas e moléculas de DNA têm muitas cargas.

Justamente por possuírem um longo alcance e serem interações fortes, as interações eletrostáticas são

criticamente importantes para entender o comportamento das moléculas. A interação eletrostática poderá guiar a proteína para a orientação correta fazendo com que uma reação ocorra.

A estrutura de uma proteína pode ser analisada de uma forma bastante exhaustiva com a utilização do STING (*Sequence To and with IN Graphics*) Millennium Suíte (SMS), uma suíte de programas com ferramentas para análise estrutural de proteínas. Estes programas estão concentrados em um pacote com o objetivo de oferecer um instrumento completo para estudos de macromoléculas. Informações como: posição dos aminoácidos na seqüência e na estrutura, busca de padrões, identificação de vizinhança, ligações de hidrogênio, ângulos e distâncias entre átomos, são facilmente obtidas. Além disso, o usuário pode obter: natureza e volume dos contatos atômicos inter e intracadeias, análise da qualidade da estrutura, conservação e relação entre os contatos intracadeia, parâmetros funcionais podem ser decifrados, etc.

O programa STING original (Neshich et al., 1998) foi

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

⁴ 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)

⁵ 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)

desenvolvido para permitir um acoplamento bidirecional das informações de sequência e estrutura de uma proteína, e também, para oferecer uma maneira simples e fácil de mapear um único aminoácido (ou nucleotídeo) na sua posição tridimensional, e vice-versa. STING Millenium está expandindo sua lista de propriedades e é agora uma ferramenta utilizada tanto para uso didático como para pesquisa na área da bioinformática estrutural.

Uma das propriedades que estão sendo incorporadas na nova versão do SMS (versão 3.0) é o valor do potencial eletrostático dos resíduos da superfície da molécula.

A determinação da distribuição de cargas é uma etapa essencial no estudo das interações inter e intramoleculares ou em comparações estruturais. As reações eletrostáticas certamente influenciam o encontro dinâmico entre moléculas. Por exemplo, a distribuição de cargas de um sistema enzima-substrato pode ajudar a atrair essas duas espécies e conduzi-las para a orientação necessária para que a reação ocorra.

Se duas moléculas se aproximam, os contatos iniciais ocorrem através de forças eletrostáticas. As interações eletrostáticas acontecem entre moléculas polares carregadas ou que tenham um momento de dipolo. Elas podem ser atrativas ou repulsivas; uma parte eletropositiva de uma molécula que está se aproximando de outra procurará se acomodar (dock) em uma região eletronegativa, enquanto partes com cargas similares vão se repelir.

Todas as estruturas de proteínas elucidadas são depositadas no *Protein Data Bank*-PDB (<http://www.rcsb.org/pdb/>), o maior depósito de distribuição para dados estruturais de moléculas biológicas. As estruturas tridimensionais de proteínas se tornaram uma parte importante do processo de descoberta de novos fármacos. Neste processo precisa-se obter o maior número de informações possíveis sobre as macromoléculas, suas características estruturais, propriedades químicas, físico-químicas e biológicas. O sítio ativo da proteína, local onde ocorre a reação pela qual a proteína é responsável, é um espaço que deve ser preenchido em termos de cargas e forma.

Com o intuito de disponibilizar um banco de dados de valores de potencial eletrostático para todas as estruturas de proteínas depositadas no PDB, foi utilizado o programa GRASP (Graphical Representation and Analysis of Structural Properties) (Nicholls et al., 1991) para geração deste banco de dados. GRASP é um programa de análise e visualização de moléculas particularmente útil para a visualização e manipulação das superfícies das moléculas e suas propriedades eletrostáticas. O programa contém algoritmos para a resolução da equação Poisson-Boltzmann que permite uma descrição do campo eletrostático gerado pelas cargas existentes na molécula.

GRASP é um programa gráfico escrito para máquinas Silicon Graphics e tem uma linguagem própria (*grasp script language* - gsl). Um *script* (Fig. 1) foi desenvolvido em gsl para gerar a superfície da proteína (Fig. 2), calcular o potencial eletrostático desta e escrever um arquivo com os valores de potencial eletrostático calculados para cada átomo. Os átomos de hidrogênio foram adicionados aos arquivos do PDB utilizando o programa *hgen* da suíte de programas CCP4 (*Collaborative Computational Project Number 4*, 1994). *Hgen* gera átomos de hidrogênio para um arquivo com as coordenadas de uma proteína usando um comprimento de ligação de 1,0 Å.

Para poder viabilizar o cálculo do valor de potencial eletrostático em larga escala para aproximadamente 19.000 estruturas do PDB, foi escrito um *script* na linguagem Perl (Fig. 3) que chama o programa GRASP.

```
Macro Name: Electrostatic_Potential
Menu Cascade Begin
menu :GRASP | Read
menu :Read Submenu | Radius/Charge File (+Assign)
menu :Charge and Size File List | full.crg
Menu Cascade End
Menu Cascade Begin
menu :GRASP | Calculate
menu :Calculate Submenu | New Potential Map
Menu Cascade EndMenu Cascade Begin
menu :GRASP | Build
menu :This Structure.. | Molecular Surface
menu :Which Atoms | All Atoms
Menu Cascade End
DRAW
Menu Cascade Begin
menu :GRASP | Calculate
menu :Calculate Submenu | Pot. via Map at Surfaces/
Atoms
menu :Which Atoms | All Atoms
menu :What Surface | All Surfaces
Menu Cascade End
DRAW
Menu Cascade Begin
menu :GRASP | Write menu :Write Submenu | Atom
(PDB) File
line :pdb.ep
menu :Coordinate Type | Absolute Centering
menu :Format of Atom File | 2 Prop. PDB, extended
fields
menu :Atom Property | Potential
menu :Atom Property | Charge
menu :Which Atoms | All Atoms
Menu Cascade End
DRAW
Key :Control Z
```

Fig. 1. Script em GRASP script language para gerar a superfície de uma proteína e os valores de potencial eletrostático associados a ela.

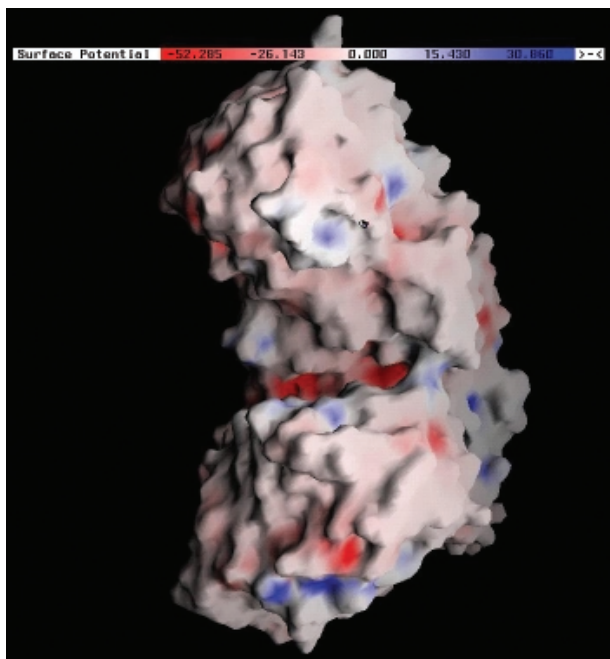


Fig. 2. Exemplo de mapa de superfície de uma Molécula gerada com *GRASP* colorida de acordo com o potencial eletrostático - vermelho: carga negativa, azul: carga positiva.

O *script* mostrado na Fig. 3 chama o programa *GRASP*, e manda executar o *script* mostrado na Fig.1. Este por sua vez gera uma figura de potencial

```
#!/usr/local/bin/perl
#
# GRASP-ht.
#
# Chama e executa o GRASP: GRASP -rx potential_full.mac
# 1acb.pdb
# para cada entrada dos arquivos *.pdb
#
@files = `ls -c1 *.pdb`;
for ($i=0;$i<=$#files;$i++) {
    $pdb=$files[$i];
    $code = substr($pdb,0,4);
    system ("echo | GRASP -rx potential_full.mac $pdb");
    print "$pdb,\n";
    $a="$pdb.ep";
    $b=$code.".ep";
    system ("mv $a $b");
}
```

Fig. 3. *Script* em *Perl* chamando o programa *GRASP*.

eletrostático como aquela ilustrada na Fig. 2 e um arquivo "nome do pdb".ep (Fig. 4) com os valores de potencial eletrostático.

	at	res	c	n	x	y	z	ep
ATOM	N	LYS	A	58	40.023	67.761	16.791	0.11993E+01
ATOM	H1	LYS	A	58	40.976	67.493	16.934	0.97139E+00
ATOM	CA	LYS	A	58	38.969	66.989	17.435	0.23117E+01
ATOM	H1A	LYS	A	58	38.406	66.588	16.712	0.11866E+01
ATOM	C	LYS	A	58	38.091	67.863	18.331	0.28893E+01
ATOM	O	LYS	A	58	36.877	67.650	18.407	0.14945E+01
ATOM	CB	LYS	A	58	39.507	65.875	18.329	0.23733E+01
ATOM	H1B	LYS	A	58	38.723	65.336	18.637	0.50510E+00
ATOM	H2B	LYS	A	58	39.943	66.310	19.117	0.31292E+01
ATOM	CG	LYS	A	58	40.509	64.913	17.734	0.16905E+00
ATOM	H1G	LYS	A	58	40.041	64.306	17.092	-0.10621E+01
ATOM	H2G	LYS	A	58	41.214	65.430	17.249	0.17333E+01
ATOM	CD	LYS	A	58	41.163	64.078	18.829	0.35404E+01
ATOM	H1D	LYS	A	58	41.384	64.676	19.599	0.52219E+01
ATOM	H2D	LYS	A	58	40.516	63.377	19.127	0.52401E+00
ATOM	CE	LYS	A	58	42.437	63.397	18.359	0.11341E+02
ATOM	H1E	LYS	A	58	42.565	62.560	18.891	0.22885E+01
ATOM	H2E	LYS	A	58	42.335	63.165	17.392	0.46236E+01
ATOM	NZ	LYS	A	58	43.651	64.249	18.517	0.34877E+02
ATOM	H1Z	LYS	A	58	43.542	65.087	17.983	0.46322E+02
ATOM	H2Z	LYS	A	58	43.771	64.482	19.482	0.25900E+02
ATOM	H3Z	LYS	A	58	44.454	63.749	18.192	0.20712E+02

Fig. 4. Parte de arquivo de saída do programa *GRASP* com os valores de potencial eletrostático inseridos na última coluna. at: tipo de átomo; res: tipo de resíduo; c: cadeia; n: número do resíduo; x,y,z: coordenadas; ep: valor do potencial eletrostático.

As informações geradas com o cálculo do potencial eletrostático estão incorporadas ao aplicativo *Java Protein Dossier* (Fig. 5) da versão 3.0 do SMS.

O *Java Protein Dossier* é uma atualização do já existente *Protein Dossier* (versão 2.2 do SMS), onde foram acrescentados, além do potencial eletrostático, os parâmetros de curvatura, hidrofobicidade, dupla ocupância e acessibilidade relativa, que são também parâmetros importantes na análise das estruturas tridimensionais de proteínas.

A maior contribuição do JPD (comparado com o *Protein Dossier*) é a possibilidade de consultar qualquer informação dos parâmetros mapeados através do clique do mouse. A inserção do potencial

eletrostática em um único arquivo, juntamente com as informações de estrutura secundária, acessibilidade, entropia entre outros oferece uma riqueza de informações e facilita a visualização de sítios importantes da proteína, potenciais sítios para ligantes, sítios de atividade, etc.

O programa SMS desenvolvido no Núcleo de Bioinformática Estrutural (NBI) da Embrapa Informática Agropecuária, pode ser utilizado por qualquer pessoa através do site <http://www.cbi.cnptia.embrapa.br>. A versão atual disponível no site é a versão 2.2, e a nova versão que inclui o resultado exposto neste trabalho está disponível no beta site do NBI e será instalada para acesso externo no início de 2003.

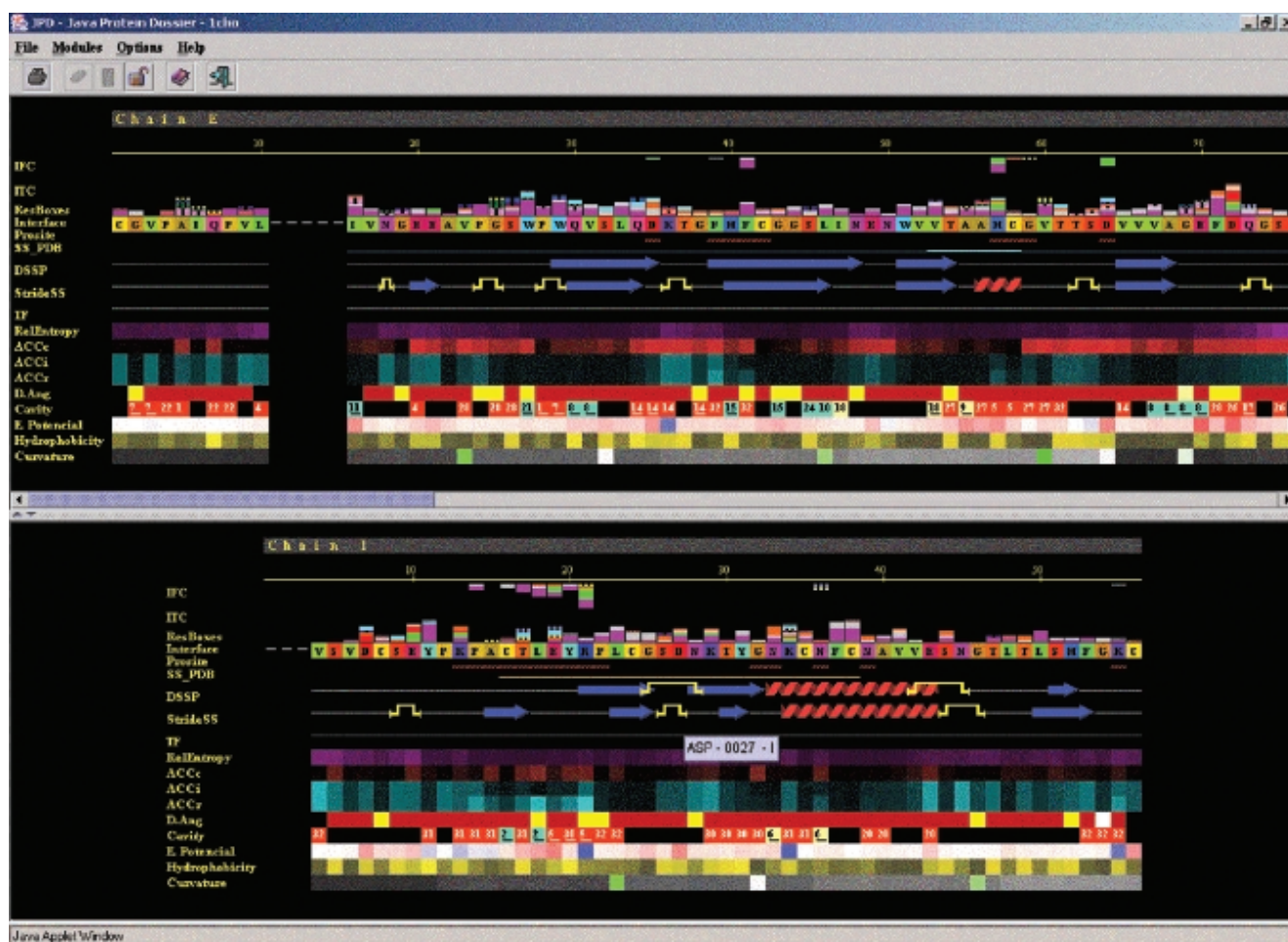


Fig. 5. *Java Protein Dossier*, com informações de potencial eletrostático incorporados na linha *E. Potential*.

Referências Bibliográficas

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

CREIGHTON, T. **Proteins: structures and molecular properties**. 2nd ed. New York: W. H. Freeman, 1993. 507 p.

NESHICH, G.; TOGAWA, R.; VILELLA, W.; HONIG, B. STING - Sequence To and withIN Graphics - PDB_viewer. **Protein Data Bank Quarterly Newsletter**, v. 84, p. 7, July, 1998.

NICHOLLS, A.; SHARP, K.; HONIG, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. **Proteins, Structure, Function and Genetics**, v. 11, n. 4, p. 281, 1991.

Comunicado Técnico, 24

**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