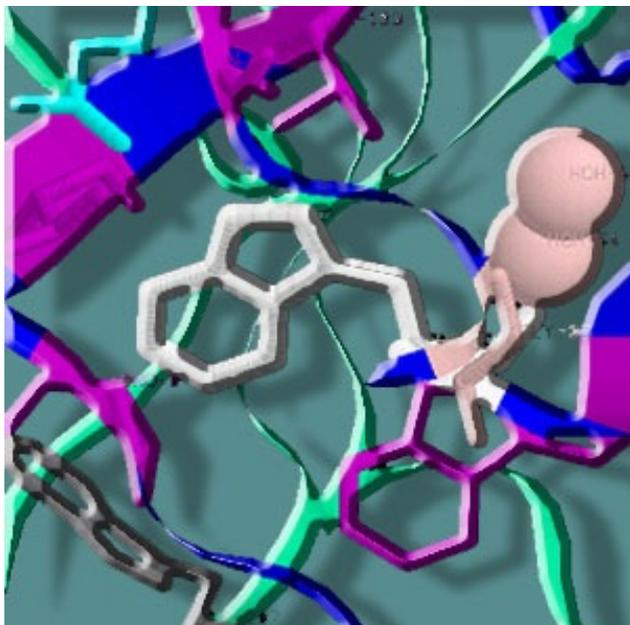


Comunicado 59

Técnico

Dezembro, 2003
Campinas, SP

ISSN 1677-8464



Cálculo de Possíveis Contatos Atômicos Internos a uma Proteína ou entre Proteínas no Software SMS

Adauto Luiz Mancini¹
Roberto Hiroshi Higa²
Paula Regina Kuser Falcão³
Michel Eduardo Beleza Yamagishi⁴
Goran Neshich⁵

Contatos interatômicos são definidos no contexto deste trabalho como as forças de atração ou de repulsão existentes entre átomos distintos. A estabilidade dos átomos nas moléculas é consequência do equilíbrio destas forças. As forças relevantes para o estudo de proteínas são as interações iônicas, pontes de hidrogênio, dipolares, van der Waals, hidrofóbicas, ponte de sulfeto e aromáticas (Fersht, 2000; Kyte, 1995; Lesk, 2001; Sternberg, 1996).

A estrutura espacial das proteínas é mantida pelo conjunto dos contatos, que são fracos se considerados individualmente (com exceção da ligação covalente, que é forte) mas tornam-se representativos à medida que sua quantidade aumenta (Day, 1996). A organização estrutural da proteína é estudada em quatro níveis: primário, secundário, terciário e quaternário. A estrutura primária, que mantém os resíduos dos aminoácidos enfileirados em um cordão polipeptídico, é conservada pelas ligações covalentes. A estrutura secundária, que consiste no enovelamento de trechos da estrutura primária em formatos tridimensionais conhecidos (hélices e folhas), é conservada pelas pontes de hidrogênio. Os principais formatos da estrutura secundária são as hélices alfa que contorcem a estrutura primária em formas helicoidais e as folhas beta

compostas de dobramentos em forma de fitas alinhadas ou cruzadas entre si (Branden & Tooze, 1999; Fersht, 2000). A estrutura terciária, formada por arranjos da estrutura secundária, é conservada pelas forças não covalentes (Dawson, 2003). A estrutura quaternária é o arranjo de mais de uma cadeia protéica para compor a proteína, e também é mantido pelas forças não covalentes (University of Paisley, 2003). Nas proteínas com formato globular, os contatos hidrofóbicos são importantes por manter os átomos num arranjo tal que átomos com carga ou com maior polaridade permaneçam na superfície externa da proteína, podendo interagir com outras moléculas, e átomos sem carga ou sem polaridade permaneçam no interior da proteína (Day, 1996; Dawson, 2003). As forças de van der Waals mantêm os átomos próximos entre si, conferindo densidade à estrutura protéica (Dawson, 2003).

Para a determinação dos contatos atômicos das proteínas é preciso considerar as características eletrostáticas e geométricas dos átomos envolvidos. A base de dados pública Protein Data Bank - PDB (Berman et al., 2000) fornece informações como os átomos constituintes e a posição espacial destes átomos de proteínas que já tenham sua estrutura tridimensional identificada. No PDB encontra-se:

¹Bacharel em Ciência da Computação, Pesquisador da Embrapa Informática Agropecuária, Caixa Postal 6041, Barão Geraldo - 13083-970 - Campinas, SP. (e-mail: adauto@cnptia.embrapa.br)

²Mestre em Engenharia Elétrica, Pesquisador da Embrapa Informática Agropecuária, Caixa Postal 6041, Barão Geraldo - 13083-970 - Campinas, SP. (e-mail: roberto@cnptia.embrapa.br)

³Ph.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)

⁴Doutor em Matemática, Pesquisador da Embrapa Informática Agropecuária, Caixa Postal 6041, Barão Geraldo - 13083-970 - Campinas, SP. (e-mail: michel@cnptia.embrapa.br)

⁵Ph.D. em Biofísica, Pesquisador da Embrapa Informática Agropecuária, Caixa Postal 6041, Barão Geraldo - 13083-970 - Campinas, SP. (E-mail: neshich@cnptia.embrapa.br)

estruturas de proteínas, podendo incluir cofatores, substratos, inibidores ou outros ligantes como ácidos nucleicos, estruturas de ácidos nucleicos ou oligonucleotídeos, estruturas de carboidratos e modelos hipotéticos de estruturas protéicas (Lesk, 2001).

Os contatos atômicos são regidos por princípios químicos, mas não há uma regra precisa para determinar sua ocorrência. Falta uma compilação que dê diretrizes precisas para todos os tipos possíveis e que seja reconhecida como a mais correta. Há na literatura trabalhos pontuais sobre tipos específicos de contato, principalmente sobre pontes de hidrogênio, cuja evolução do conhecimento é descrita por Weiss et al. (2001). A precisão do cálculo de contatos atômicos é diretamente dependente da qualidade do conhecimento das posições espaciais dos átomos na estrutura protéica. Esta consideração aplica-se principalmente aos átomos de hidrogênio, que não são identificados em estruturas cristalizadas resolvidas pelo método de difração de raios x. Uma parcela importante de estruturas do PDB é determinada por este método, identificando-se o arranjo espacial dos átomos (posição em coordenadas tridimensionais). O segundo método mais usado, o Nuclear Magnetic Resonance - NMR, tem a limitação de funcionar adequadamente com proteínas pequenas que contenham 150 resíduos, aproximadamente (Fersht, 2000).

Nas próximas seções será explicado o método de inferência desenvolvido para a identificação dos contatos potenciais e exemplos dos resultados obtidos do cálculo e de sua visualização no software SMS.

Método Proposto

O método desenvolvido estabelece uma classificação de átomos que possuem comportamento eletrostático semelhante e pertencem a um mesmo tipo de cadeia no resíduo (cadeia principal, cadeia lateral) em grupos predeterminados. Para cada tipo de contato são estabelecidos pares de grupos (X, Y) tais que um átomo x do grupo X e um átomo y do grupo Y possuam afinidade para estabelecer o contato e o intervalo de distâncias mínima e máxima. Então, dada uma proteína, seus átomos são classificados e adicionados aos grupos de átomos. Para um dado tipo de contato todos os pares de átomos (x,y) pertencentes aos pares de grupos (X, Y) com a distância entre x e y pertencendo ao intervalo estabelecido de distâncias mínima e máxima são considerados como um contato. A Tabela 1 mostra para cada tipo de contato calculado pelo método os grupos cujos átomos são candidatos a estabelecer um contato do tipo considerado. Para cada primeira ocorrência de um grupo na tabela, uma descrição dos átomos que compõem o grupo é apresentada em forma de nota de rodapé. A nomenclatura usada para os átomos é a mesma utilizada para nomeá-los nos arquivos textos da base de dados PDB.

Para o caso de contatos internos de uma mesma cadeia os átomos x e y candidatos à avaliação devem ser distintos, não pertencerem a um mesmo resíduo e, se pertencerem a resíduos distintos, os resíduos não devem ser consecutivos. A primeira restrição garante que o contato é entre dois átomos. As duas últimas restrições desconsideram

interações entre átomos que sempre estarão próximos por causa da posição igual ou consecutiva do(s) resíduo(s) na cadeia polipeptídica.

Para os contatos do tipo ponte de hidrogênio, que envolvem 3 átomos (1 átomo hidrogênio e 2 átomos fortemente eletronegativos, como oxigênio ou nitrogênio), é considerada somente a distância entre os 2 átomos eletronegativos. Quanto mais o átomo de hidrogênio estiver alinhado entre os outros dois átomos, mais forte será a interação. Isto é feito porque nas estruturas resolvidas por difração de raios x a localização dos átomos de hidrogênio não é determinada. Outra particularidade das pontes de hidrogênio é que, se houver presença de água na proteína, pode-se formar redes de pontes de hidrogênio. Assim, decidiu-se estender o cálculo de pontes de hidrogênio considerando também a presença de uma ou duas moléculas intermediárias de água. Isto é feito da seguinte forma para uma molécula de água intermediária: seja w o átomo oxigênio de uma dada molécula de água e d(a,b) a distância entre dois átomos a e b. A água w intermedia a ponte de hidrogênio entre os átomos x e y se

$$\text{dist}_{\min} <= d(x, w) <= \text{dist}_{\max} \text{ e } \text{dist}_{\min} <= d(w, y) <= \text{dist}_{\max}$$

No caso de duas moléculas de água intermediárias w1 e w2 a restrição é:

$$\text{dist}_{\min} <= d(x, w1) <= \text{dist}_{\max} \text{ e } \text{dist}_{\min} <= d(w1, w2) <= \text{dist}_{\max} \text{ e } \text{dist}_{\min} <= d(w2, y) <= \text{dist}_{\max}$$

O fato dos átomos serem selecionados para os agrupamentos também por sua posição no resíduo (cadeia principal, cadeia lateral) não tem caráter funcional, mas foi aplicado para poder mostrar os resultados subdivididos por tipo de cadeia. A rigor, basta que os átomos satisfaçam as restrições eletrostáticas e de distância para a identificação de um dado tipo de contato.

A distância $\text{dist}(A, B)$ entre dois pontos $A = (x_A, y_A, z_A)$ e $B = (x_B, y_B, z_B)$ no espaço tridimensional é dada pela fórmula:

$$\text{dist}(A, B) = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2 + (z_A - z_B)^2}$$

Para acelerar a verificação da distância entre os dois átomos no intervalo de dist_{\min} a dist_{\max} é usado o seguinte (pseudo)código para redução do processamento numérico, com o símbolo ← representando o operador de atribuição, isto é, $x ← 10$ significa atribuir o valor 10 à variável x.

```

distAB2 ← (xA - xB)2
if (distAB2 > distMAX2)
  then reject;
else {distAB2 ← distAB2 + (yA - yB)2
      if (distAB2 > distMAX2)
        then reject;
      else {distAB2 ← distAB2 + (zA - zB)2
            if (distAB2 > distMAX2)
              then reject
            else if (distAB2 < distMIN2)
              then reject
            else accept
          }
    }
  
```

Usando o código anterior, ao invés de executar-se sempre todos os cálculos da fórmula da distância, faz-se a soma das diferenças ao quadrado progressivamente, uma vez que, se esta soma for maior que o quadrado da distância máxima permitida, pode-se rejeitar o resultado, porque a próxima diferença ao quadrado irá manter ou aumentar o valor da soma.

Tabela 1. Pares de grupos de átomos.

<i>Interação hidrofóbica</i>	
Hidrofóbico ⁶	Hidrofóbico
<i>Interação eletrostática atrativa</i>	
Lateral negativo ⁷	Lateral positivo ⁸
<i>Interação eletrostática repulsiva</i>	
Lateral positivo	Lateral positivo
Lateral negativo	Lateral negativo
<i>Cadeia principal X Cadeia principal</i>	
Doador principal ⁹	Receptor principal ¹⁰
<i>Cadeia principal X Cadeia lateral</i>	
Doador principal	Lateral quase doador ¹¹
Doador principal	Receptor lateral ¹²
Doador principal	Lateral negativo
Receptor principal	Lateral positivo
Receptor principal	Doador lateral ¹³
Receptor principal	Lateral quase doador
<i>Cadeia lateral X Cadeia lateral</i>	
Lateral positivo	Lateral quase doador
Lateral positivo	Receptor lateral
Doador lateral	Lateral quase doador
Doador lateral	Receptor lateral
Doador lateral	Lateral negativo
Lateral quase doador	Receptor lateral
Lateral quase doador	Lateral negativo
<i>Cadeia principal X DNA</i>	
Doador principal	DNA receptor ¹⁴
Doador principal	DNA quase doador
Receptor principal	DNA doador ¹⁵
Receptor principal	DNA quase doador
<i>Cadeia lateral X DNA</i>	
Lateral positivo	DNA receptor
Doador lateral	DNA receptor
Lateral quase doador	DNA receptor
Lateral negativo	DNA doador
Receptor lateral	DNA doador
Lateral quase doador	DNA doador
Receptor lateral	DNA quase doador ¹⁶
Doador lateral	DNA quase doador
Lateral positivo	DNA quase doador
<i>Cadeia principal X fosfato</i>	
Doador principal	DNA fosfato receptor
<i>Cadeia lateral X fosfato</i>	
Lateral positivo	DNA fosfato receptor
Doador lateral	DNA fosfato receptor
Lateral quase doador	DNA fosfato receptor
<i>Pontes de Disulfeto</i>	
Enxofre de cisteína	Enxofre de cisteína

⁶ Hidrofóbico: átomo carbono de resíduo de cadeia protéica.⁷ Lateral negativo: átomo negativamente carregado de cadeia lateral de resíduo protéico (OD1,OD2 de ASP; OE1,OE2 de GLU).⁸ Lateral positivo: átomo positivamente carregado de cadeia lateral de resíduo protéico(NZ de LYS;ND1,NE2 de HIS; NH1,NH2 de ARG).⁹ Doador principal: átomo nitrogênio da cadeia principal de resíduo protéico.¹⁰ Receptor principal: átomo oxigênio da cadeia principal de resíduo protéico.¹¹ Lateral quase doador: átomo de cadeia lateral de resíduo protéico que em geral comporta-se como doador, mas age como receptor ao encontrar um doador (OG da SER; OG1 da THR; OH da TYR).¹² Receptor lateral: átomo receptor de cadeia lateral de resíduo protéico (OD1 de ASN; OE1 de GLN).¹³ Doador lateral: átomo doador de cadeia lateral de resíduo protéico (NE1 de TRP; ND2 de ASN; NE2 de GLN).¹⁴ DNA receptor: átomo receptor de nucleotídeo de DNA (O4 de (+)T,(+)U; O6 de (+)G; O2 de (+)T,(+)C,(+)U; N1 de (+)A; N3 de (+)C,(+)G,(+)A; N7 de (+)G,(+)A; O'3 de (+)A, (+)T(+)C, (+)G, (+)U; O'4 de (+)A, (+)T(+)C, (+)G, (+)U; O'5 de (+)A, (+)T(+)C, (+)G, (+)U).¹⁵ DNA doador: átomo doador de base de cadeia de DNA (N1,N2 de (+)G; N4 de (+)C; N3 de (+)T, (+)U).¹⁶ DNA quase doador: átomo de nucleotídeo de DNA que em geral comporta-se como doador, mas age como receptor ao encontrar um doador (O'2 de (+)A, (+)C, (+)G, (+)U).

Um ponto importante a ser observado é que os contatos calculados não devem ser considerados como definitivos, mas como contatos potenciais. Isto significa que produz-se uma indicação de possíveis contatos, não uma afirmação de sua existência. Os motivos desta incerteza são a inexistência de uma regra precisa e aceita para a determinação dos contatos, a possibilidade de parametrização dos intervalos de distância mínima e máxima, a não determinação dos átomos de hidrogênio em estruturas resolvidas por difração de raios x e a não consideração de possíveis interferências causadas pelos átomos da vizinhança dos átomos envolvidos nos contatos.

Resultados

A seleção de átomos em função de similaridade de comportamento eletrostático em grupos antes da verificação dos contatos permite que os cálculos sejam feitos apenas com os átomos de interesse para um dado tipo de contato. Sem esta seleção seria necessário analisar individualmente, para cada combinação possível entre dois átomos, se os átomos têm afinidade para fazer um contato de dado tipo e aplicar as restrições de distância.

O método proposto está implementado no software STING Millennium Suite (Neshich et al., 2003) que pode ser publicamente acessado via web (Structural Bioinformatics Group, 2003). O programa foi codificado em C++ e é orientado a objetos, e está inserido na biblioteca (*lib*) do SMS, juntamente com outros programas de funcionalidades distintas, como o leitor de arquivos textos da base de dados do PDB. O programa de contatos utiliza-se do leitor de pdb para obter como entrada a composição e as informações espaciais da proteína (cadeias da proteína, resíduos das cadeias, átomos dos resíduos e suas posições). Uma vez obtidos os dados estruturais da proteína, utiliza-se o método para cálculo dos contatos. O programa também dispõe de métodos para leitura e gravação dos contatos em arquivo texto em um formato estabelecido pelos desenvolvedores do SMS.

Comparando-se os resultados obtidos pelo método proposto, implementado no SMS, com os obtidos pela ferramenta Contacts of Structural Units (2003) observa-se que o método calcula apropriadamente contatos hidrofóbicos e pontes de hidrogênio tradicionais (NH-O, OH-N). Interações com átomos de carbono formando pontes de hidrogênio mais fracas que as tradicionais, podendo ser doador de hidrogênio ou receptor quando o carbono pertencer a um anel aromático, não foram consideradas no método. O método deverá ser ampliado para incluir o cálculo destas pontes de hidrogênio com carbonos, porque apesar de mais fracas que as pontes tradicionais, podem ter papel importante na estabilidade da estrutura protéica porque ocorrem com frequência. Optou-se por não calcular contatos entre resíduos consecutivos da cadeia no método apresentado porque os resíduos consecutivos já são fortemente conectados pela ligação peptídica, mas pode-se alterar esta implementação facilmente caso desejado. A ferramenta Contacts of Structural Units (2003) calcula contatos desestabilizadores da estrutura que não são previstos no método proposto com exceção de contatos repulsivos entre átomos com carga eletrostática igual. A abordagem usada no método proposto é apontar apenas os contatos estabilizadores da estrutura. A

seguir são mostrados os contatos obtidos pelo Contacts of Structural Units e a comparação com o método proposto para os resíduos 27 e 261 da cadeia A da estrutura código 1mxm da base de dados PDB. As entradas da coluna *Dist* marcadas com * indicam contatos desestabilizadores. As entradas da coluna *Método proposto* mostram se o contato obtido pelo Contacts of Structural Units foi também obtido pelo método proposto. Entradas da coluna *Método proposto* com valor *não previsto* indicam que o tipo de contato não é especificado pelo método proposto, e as entradas com valor *consecutivos* indicam que o contato é entre dois resíduos consecutivos, não sendo calculados pela definição do método proposto.

Atom from TYR 27 A		Contacting atom				
Name	Class	Residue	Name	Class	Dist	Método proposto
N	III	ALA 28A	N	III	3.1*	não previsto
N	III	ILE 31A	CG2	IV	4.1*	não previsto
N	III	ILE 31A	CB	IV	4.4*	não previsto
N	III	ILE 31A	CD1	IV	5.2*	não previsto
CA	VII	ALA 28A	N	III	2.4*	não previsto
C	VI	ALA 28A	N	III	1.3	não previsto
C	VI	ALA 28A	CA	VII	2.5	consecutivos
C	VI	VAL 29A	N	III	2.8	não previsto
C	VI	ILE 31A	CG2	IV	3.6	ok
O	II	ALA 28A	N	III	2.3	consecutivos
O	II	VAL 29A	N	III	2.4	ok
O	II	ALA 28A	C	VI	2.7	não previsto
O	II	ILE 31A	CG2	IV	2.9*	não previsto
O	II	ILE 31A	N	III	3.1	ok
CB	IV	ASN 30A	CB	IV	3.8	ok
CB	IV	ILE 31A	N	III	3.9	não previsto
CB	IV	ASN 30A	N	III	4.1*	não previsto
CB	IV	ILE 31A	CB	IV	4.4	ok
CG	V	ASN 30A	CB	IV	4.4	ok
CG	V	ASN 30A	OD1	II	5.1	não previsto
CG	V	ASN 30A	CG	VI	5.2	ok
CD1	V	ILE 31A	CB	IV	5.4	ok
CD1	V	ILE 31A	CD1	IV	6.3	ok
CD2	V	ASN 30A	CB	IV	4.3	ok
CD2	V	ASN 30A	OD1	II	4.7	não previsto
CE2	V	ASN 30A	OD1	II	5.2	não previsto

Atom from PHE 261		Contacting atom				
Name	Class	Residue	Name	Class	Dist	Método proposto
N	III	GLU 260A	C	VI	1.3	não previsto
N	III	GLU 260A	CA	VII	2.4*	não previsto
N	III	GLU 260A	N	III	2.8*	não previsto
N	III	ARG 259A	O	II	2.9	ok
N	III	ARG 259A	C	VI	3.0	nao prev
N	III	ILE 257A	O	II	3.2	ok
CA	VII	GLU 260A	C	VI	2.4	consecutivos
CA	VII	GLU 260A	O	II	2.8	não previsto
CA	VII	ILE 266A	N	III	3.5	não previsto
CA	VII	GLY 265A	N	III	3.9	não previsto
CA	VII	GLY 265A	CA	VI	4.0	ok
CA	VII	ILE 204A	CG2	IV	5.5	ok
C	VI	ASP 262A	N	III	1.3	não previsto
C	VI	ASP 262A	CA	VII	2.4	consecutivos
C	VI	ALA 263A	N	III	2.8	não previsto
C	VI	ILE 266A	N	III	3.0	não previsto
O	II	ASP 262A	N	III	2.3	consecutivos
O	II	ALA 263A	N	III	2.5	ok
O	II	ILE 266A	N	III	2.6	ok
O	II	ASP 262A	C	VI	2.6	não previsto
O	II	ALA 264A	N	III	2.6	ok

O	II	GLY265A	N	III	2.8	ok
CB	IV	ASP 262A	N	III	2.8*	não previsto
CB	IV	ILE 257A	O	II	3.1*	não previsto
CB	IV	PHE 268A	CZ	V	3.8	ok
CB	IV	ILE 266A	N	III	3.8*	não previsto
CB	IV	PHE 268A	CE1	V	4.0	ok
CB	IV	ILE 266A	CA	VII	4.3	ok
CG	V	ILE 257A	O	II	3.3	não previsto
CG	V	ILE 266A	N	III	4.6	não previsto
CG	V	GLY 265A	C	VI	5.0	ok
CG	V	GLY 265A	CA	VI	5.0	ok
CD1	V	ILE 257A	O	II	3.1	não previsto
CD1	V	ILE 204A	CG2	IV	3.9	ok
CD1	V	ILE 257A	C	VI	4.0	ok
CD1	V	ILE 257A	CA	VII	4.2	ok
CD1	V	ILE 204A	CB	IV	4.6	ok
CD1	V	ILE 204A	CG1	IV	5.2	ok
CD1	V	GLY 265A	CA	VI	5.5	ok
CD2	V	PHE 268A	CZ	V	3.8	ok
CD2	V	PHE 268A	CE2	V	3.8	ok
CD2	V	ILE 266A	O	II	4.1	não previsto
CD2	V	ILE 266A	N	III	4.8	não previsto
CD2	V	ILE 190A	O	II	4.8	não previsto
CD2	V	GLY 265A	C	VI	5.0	ok
CD2	V	GLY 265A	CA	VI	5.3	ok
CD2	V	VAL 201A	CG2	IV	5.7	ok
CD2	V	GLY 191A	CA	VI	6.4	ok
CE1	V	ILE 257A	CG2	IV	3.5	ok
CE1	V	ILE 204A	CG2	IV	3.6	ok
CE1	V	LEU 205A	CG	IV	3.9	ok
CE1	V	ILE 204A	CB	IV	4.1	ok
CE1	V	LEU 205A	CD2	IV	4.3	ok
CE1	V	ILE 208A	CD1	IV	4.3	ok
CE2	V	VAL 201A	CG1	IV	4.0	ok
CE2	V	PHE 268A	CE2	V	4.3	ok
CE2	V	ILE 257A	CG2	IV	4.4	ok
CE2	V	LEU 205A	CD1	IV	4.6	ok
CE2	V	ILE 189A	CG2	IV	4.6	ok
CE2	V	ILE 190A	O	II	4.7	não previsto
CE2	V	VAL 201A	CG2	IV	4.9	ok
CE2	V	ILE 266A	O	II	4.9	não previsto
CE2	V	VAL 237A	CG2	IV	5.0	ok
CE2	V	VAL 201A	CA	VII	5.3	ok
CE2	V	ILE 190A	C	VI	5.6	ok
CE2	V	GLY 191A	CA	VI	6.1	ok
CZ	V	LEU 205A	CD1	IV	3.6	ok
CZ	V	LEU 205A	CG	IV	3.6	ok
CZ	V	VAL 201A	CG1	IV	3.8	ok
CZ	V	ILE 257A	CG2	IV	3.9	ok
CZ	V	LEU 205A	CD2	IV	4.1	ok
CZ	V	ILE 189A	CG2	IV	4.5	ok
CZ	V	VAL 201A	O	II	4.6	não previsto
CZ	V	ILE 204A	CB	IV	4.8	ok
CZ	V	VAL 201A	CA	VII	4.9	ok

O Sting Millennium Suite - SMS (Neshich et al., 2003) é uma suite de programas desenvolvida pelo Núcleo de Bioinformática da Embrapa Informática Agropecuária para análise de estruturas de proteínas e a relação com a sua função. Dentre eles, existem ferramentas para análise de contatos que apresentam para o usuário, de forma gráfica, a quantidade e os tipos de contatos em que cada resíduo da proteína está envolvido. Além disso, elas também permitem que os resíduos envolvidos sejam mapeados na estrutura tridimensional.

O Protein Dossier é um módulo do SMS, cuja funcionalidade é produzir e apresentar um cartoon, onde uma série de parâmetros físico-químicos utilizados para análise da estrutura tridimensional da proteína e sua relação com a função são resumidas (Fig. 1).

Tanto os contatos internos - ICT quanto os contatos na interface - IFR são apresentados em forma de gráfico de barras. Para cada resíduo, é apresentada uma barra com altura proporcional ao número de contatos nos quais o resíduo está envolvido. Um código de cores identifica cada tipo de contato, de acordo com a tabela apresentada na Fig. 2.

Para uma análise mais detalhada dos contatos de uma proteína utilizando o SMS, as ferramentas indicadas são o Graphical Contacts para análise de contatos internos e IFR Graphical Contacts para análise dos contatos na interface. Estas ferramentas têm apresentação e funcionalidades similares, sendo que a diferença entre elas reside no fato de uma apresentar os contatos internos de uma cadeia da proteína - *Internal Contacts* e a outra os contatos na interface - *IFR contacts*.

O intervalo das distâncias mínima e máxima para cada um dos tipos de contatos e a cadeia da proteína que deseja-se analisar os contatos podem ser configurados pelo usuário (Fig. 3), nos módulos Graphical Contacts ou IFR Contacts.

A Fig. 4 ilustra os recursos gráficos disponíveis para análise de contatos através do Graphical Contacts. Na janela principal do Graphical Contacts a seqüência de resíduos que compõe a proteína é apresentada em conjunto com um gráfico de barras que indica a quantidade de contatos em que cada resíduo está envolvido. Na posição central da janela, é exibido um leque representando todos os contatos internos nos quais o resíduo posicionado no centro da janela está envolvido. Na extremidade do leque são apresentados os resíduos fazendo contato com o resíduo posicionado no centro. Através de uma barra de rolagem, é possível visualizar o leque de contatos para cada resíduo da seqüência de resíduos da proteína. Uma janela auxiliar informa a quantidade exata de cada tipo de contato representada pelo leque. Também são oferecidas opções para visualização ou não dos contatos com distância mais curta e/ou pontes de hidrogênio intermediadas por água. O usuário pode ainda visualizar uma tabela com todos os contatos e, posicionando o cursor no resíduo central na janela principal, mapear o leque de contatos na estrutura tridimensional da proteína para análise utilizando o Chime plugin (MDL Informatics Systems, 2002).

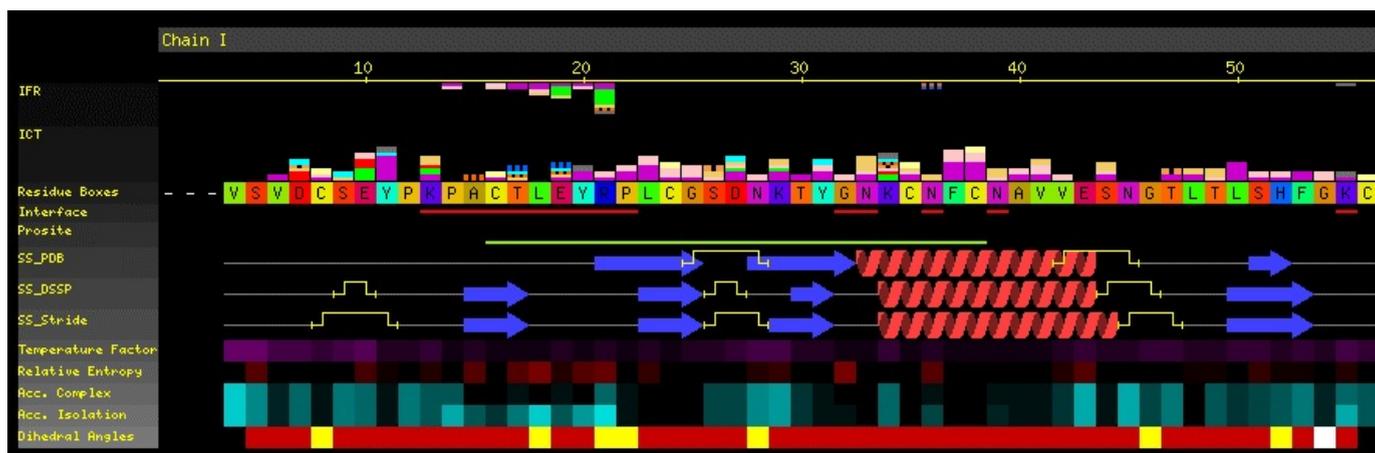


Fig. 1. Apresentação de Contatos no Protein Dossier.

Internal Contacts (ICT) and Interface Contacts (IFR):

	Hydrophobic
	Charged Attractive
	Charged Repulsive
	Hydrogen Bond : Main Chain - Main Chain
	Hydrogen Bond : Main Chain - Water - Main Chain
	Hydrogen Bond : Main Chain - Water - Water - Main Chain
	Hydrogen Bond : Main Chain - Side Chain
	Hydrogen Bond : Main Chain - Water - Side Chain
	Hydrogen Bond : Main Chain - Water - Water - Side Chain
	Hydrogen Bond : Side Chain - Side Chain
	Hydrogen Bond : Side Chain - Water - Side Chain
	Hydrogen Bond : Side Chain - Water - Water - Side Chain
	Aromatic Stacking
	Cysteine Bridge

Fig. 2. Tabela com o código de cores utilizado para apresentação de contatos.

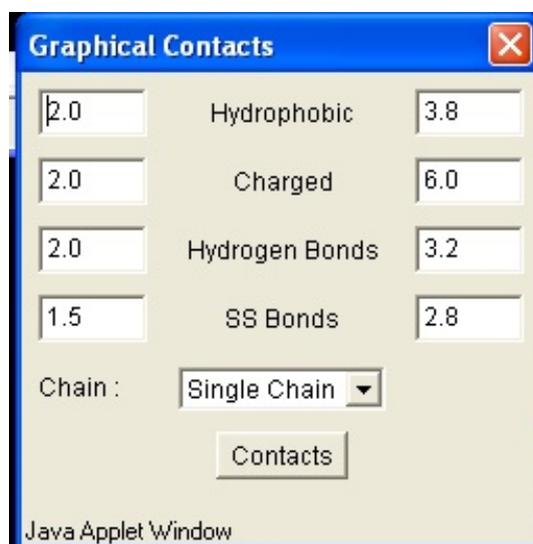


Fig. 3. Configuração de intervalos de distância e seleção de cadeia.

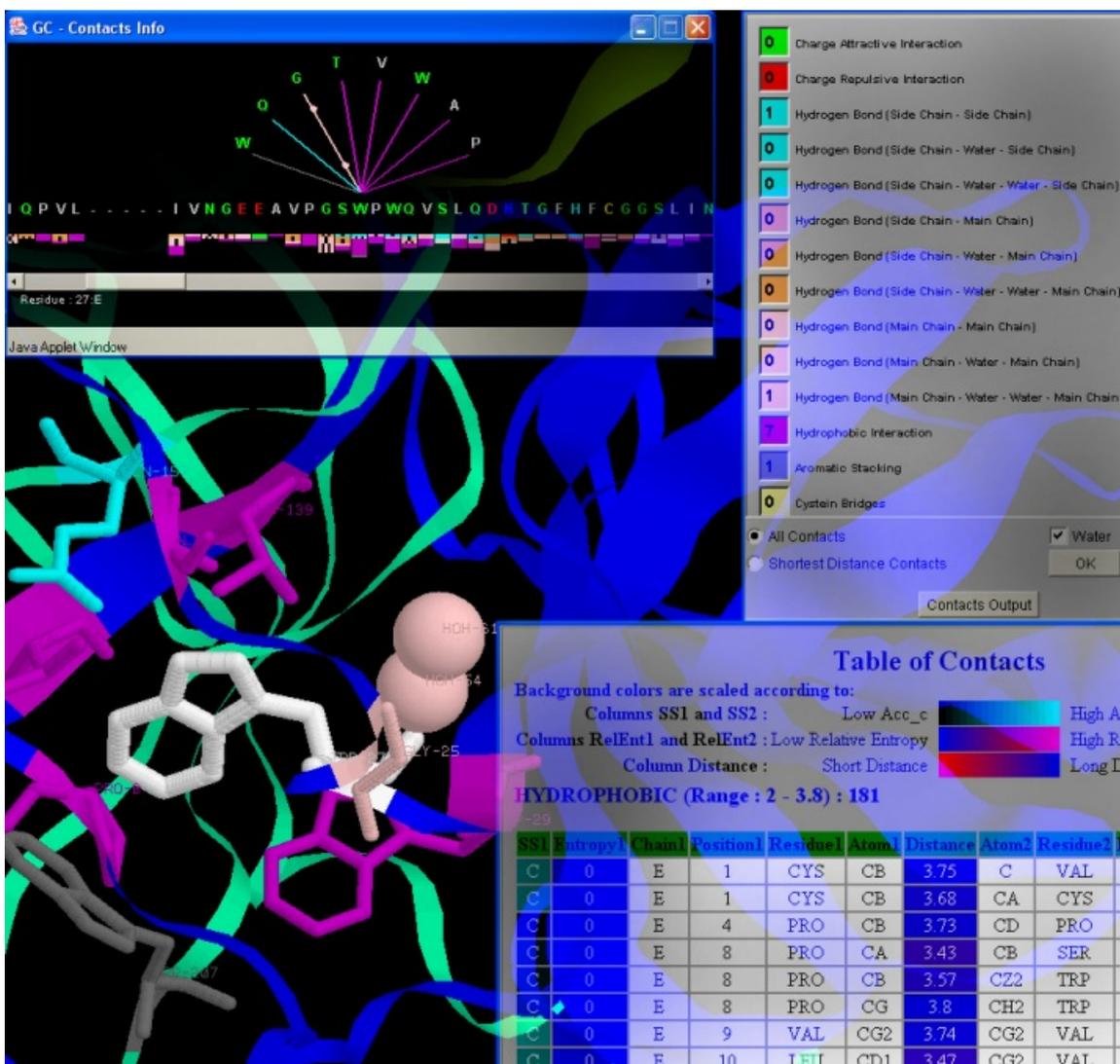


Fig. 4. Análise de Contatos utilizando o Graphical Contacts.

Conclusões

O método proposto consegue abranger vários tipos de contatos usando cálculos simples de restrições de distância, que podem ser parametrizados e recalculados rapidamente para o usuário em função da simplicidade dos cálculos. Uma rica interface gráfica permite visualizar os dados em diversas formas de apresentação, inclusive enfatizando os contatos em uma janela de visualização espacial da proteína. Estes recursos permitem que o usuário possa de forma interativa, analisar rapidamente os contatos nos locais da proteína em que tiver interesse.

Contatos são importantes para a análise da estrutura de uma proteína, pois eles são responsáveis tanto pela estabilidade da estrutura da proteína - *Internal Contacts*, quanto pelas interações que ela realiza com outras moléculas - *IFR Contacts*.

A análise deste importante parâmetro estrutural utilizando o SMS é realizada através de um conjunto de ferramentas gráficas que oferece ao usuário de forma amigável um rico conjunto de informações. Em particular, o Graphical Contacts e o IFR Contacts apresentam os contatos de forma gráfica, mas também oferecem ao usuário os dados exatos através de tabelas, além da possibilidade de análise dos contatos mapeados na estrutura tridimensional da proteína.

No momento, os contatos calculados são internos a uma cadeia protéica, na interface entre cadeias protéicas e entre cadeias protéicas e cadeias nucleicas (DNA/RNA). Os contatos entre cadeias protéicas e moléculas não protéicas (como fármacos, ions, herbicidas, etc.) ainda não foram incluídos no método.

Desta forma, desenvolvimentos futuros dessas ferramentas devem priorizar a incorporação deste tipo de contato, bem como o enriquecimento da capacidade de interação com o usuário, por exemplo, tornando a tabela de contatos interativa.

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.

CONTACTS OF STRUCTURAL UNITS. Ligand-Protein Contacts & Contacts of Structural Units. Disponível em: < <http://bip.weizmann.ac.il/oca-bin/lpccsu> >. Acesso em: 13 dez. 2003.

DAWSON, J. F. Chem 730 proteins and nucleic acids fall semester 2002. Disponível em: < <http://www.chembio.uoguelph.ca/educmat/chm730/CHM730.htm> >. Acesso em: 15 jul. 2003.

DAY, A. The source of stability in proteins. 1996. Disponível em: < <http://www.cryst.bbk.ac.uk/PPS2/projects/day/TDayDiss/index.html> >. Acesso em: 15 jul. 2003.

FERSHT, A. Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding. 3rd ed. New York: W. H. Freeman, 2000.

KYTE, J. Structure in protein chemistry. New York: Garland, 1995.

LESK, A. M. Introduction to protein architecture: the structural biology of proteins. New York: Oxford University Press, 2001. 360 p.

MDL INFORMATION SYSTEMS. MDLThe MDL Chime site. Disponível em: < <http://www.mdlichime.com/chime/> >. Acesso em: 21 dez. 2002.

NESHICH, G.; TOGAWA, R. C.; MANCINI, A. L.; KUSER, P. R.; YAMAGISHI, M. E. B.; PAPPAS JUNIOR, G.; TORRES, W. V.; CAMPOS, T. F. e; FERREIRA, L. L.; LUNA, F. M.; OLIVEIRA, A. G.; MIURA, R. T.; INOUE, M. K.; HORITA, L. G.; SOUZA, D. F. de; DOMINQUINI, F.; ÁLVARO, A.; LIMA, C. S.; OGAWA, F. O.; GOMES, G. B.; PALANDRANI, J. F.; SANTOS, G. F. dos; FREITAS, E. M. de; MATTIUZ, A. R.; COSTA, I. C.; ALMEIDA, C. L. de; SOUZA, S.; BAUDET, C.; HIGA, R. H. STING Millennium: a Web based suite of programs for comprehensive and simultaneous analysis of protein structure and sequence. *Nucleic Acids Res.*, v. 31, n. 13, p. 3386-3392, 2003.

STERNBERG, M. J. E. (Ed.). Protein structure prediction: a practical approach. Oxford: IRL Press at Oxford University Press, 1996.

STRUCTURAL BIOINFORMATICS GROUP. Sting Millennium Suite. Disponível em: < <http://www.nbi.cnptia.embrapa.br/> >. Acesso em: 8 dez. 2003.

UNIVERSITY OF PAISLEY. Department of Biological Sciences. Structure and function of macromolecules. Disponível em: < <http://www-biol.paisley.ac.uk/courses/stfunmac/home.html> >. Acesso em: 11 dez. 2003.

WEISS, M. S.; BRANDL, M.; SÜHNEL, J.; PAL, D.; HILGENFELD, R.; More hydrogen bonds for the (structural) biologist. *Trends Biochem. Sci.*, v. 26, n. 9, p. 521-523, Sep. 2001.

Comunicado
Técnico, 59

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

Ministério da Agricultura,
Pecuária e Abastecimento

Governo
Federal

1ª edição on-line - 2003

ÓTodos os direitos reservados.

Comitê de
Publicações

Presidente: *Luciana Alvim Santos Romani*
Membros Efetivos: *Carla Geovana Macário, José Ruy Porto de Carvalho, Marcia Izabel Fugisawa Souza, Marcos Lordello Chaim, Suzilei Almeida Carneiro.*
Suplentes: *Carlos Alberto Alves Meira, Eduardo Delgado Assad, Maria Angelica Andrade Leite, Maria Fernanda Moura, Maria Goretti Gurgel Praxedes.*

Expediente

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