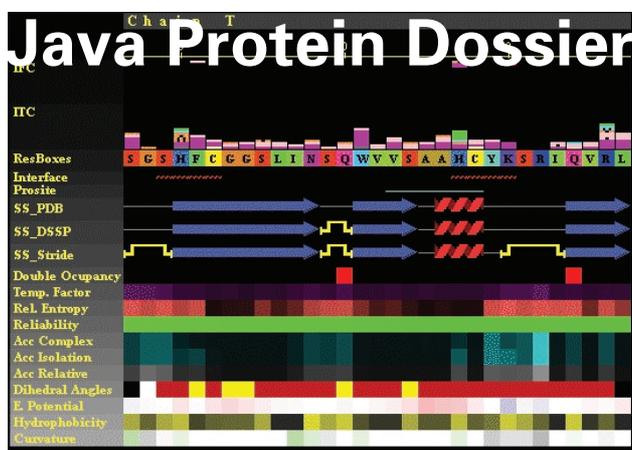


ISSN 1677-8464



Apresentação Gráfica de Parâmetros Protéicos Utilizando o Java Protein Dossier

Roberto Hiroshi Higa¹
Christian Baudet²
Paula Kuser Falcão³
Adauto Luiz Mancini⁴
Goran Neshich⁵

Para analisar estruturas de proteínas e sua relação com a função, biólogos moleculares utilizam diversos tipos de descritores ou parâmetros sobre a estrutura da proteína em estudo. Geralmente, para ter acesso a essas informações, é preciso acessar diversos *web sites* e utilizar programas que calculam os parâmetros. Ainda assim, a interpretação simultânea desses parâmetros é difícil, pois os resultados obtidos são apresentados de forma textual e não estão integrados.

Java Protein Dossier (JPD) é um programa integrado à *suite* de programas *Sting Millennium Suite* (SMS) (Structural Bioinformatics Group, 2002), desenvolvido pelo Núcleo de Bioinformática Estrutural (NBI) da Embrapa Informática Agropecuária, e disponibilizado a partir da versão 3.0, que visa exatamente atender a esta demanda de apresentação gráfica e integrada de uma grande quantidade de parâmetros estruturais e funcionais.

O JPD é um Java Applet (Sun Microsystems, 2002) que apresenta uma série de parâmetros de proteínas, cuja

estrutura tenha sido resolvida experimentalmente ou modelada utilizando-se programas de modelagem molecular. Estes parâmetros são apresentados graficamente e podem ser mapeados na estrutura tridimensional da proteína, que é apresentada utilizando-se o *Chime plugin*⁶ (MDL Information Systems, 2002). Para utilizá-lo, é preciso fornecer um arquivo que descreva a estrutura tridimensional da proteína em formato Protein Data Bank (PDB) (Berman et al., 2000). Pode-se fazer isso especificando-se o código correspondente a uma estrutura depositada no banco de dados PDB - PDB-ID, ou especificando um arquivo próprio, por exemplo, o arquivo resultante de um processo de modelagem molecular.

A apresentação dos parâmetros é organizada por resíduo e a apresentação gráfica é construída de forma a fornecer uma informação qualitativa sobre o valor de cada parâmetro, através de um código de cores. Além disso, posicionando-se o mouse sobre um determinado parâmetro para um determinado resíduo, o valor numérico do parâmetro é apresentado.

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 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 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 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)

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

⁶ O *Chime Plugin* é um programa integrável a *web browsers* que apresenta graficamente estruturas tridimensionais de macromoléculas em formato PDB (Berman et al., 2000).

O JPD funciona de maneira interativa de modo que, utilizando-se o *mouse*, os outros módulos do SMS podem ser acessados para verificar a posição do resíduo na estrutura tridimensional, como se comporta uma propriedade na vizinhança de um resíduo e listar todos os resíduos que possivelmente fazem contatos com um determinado resíduo, entre outras funcionalidades. O JPD permite ainda selecionar sítios específicos e comparar a proteína sob estudo com outra representada em um segundo arquivo PDB. Também é possível verificar como interagem duas cadeias de um mesmo complexo de proteínas, representados em um arquivo PDB.

O objetivo deste trabalho é apresentar como o JPD pode ser usado para analisar estruturas de proteínas. Na seção seguinte serão descritos os parâmetros protéicos apresentados pelo JPD. Em seguida, será apresentado um conjunto de funcionalidades do JPD utilizadas para análise mais detalhada dos parâmetros apresentados, como a seleção de resíduos baseados em propriedades e a integração com os outros módulos do SMS. Finalmente, são apresentadas as considerações finais sobre o trabalho desenvolvido.

Parâmetros apresentados pelo JPD

O JPD apresenta, através de uma representação gráfica, um leque variado de informações sobre a proteína. As informações são organizadas por resíduo, tal que cada resíduo possui um valor para cada parâmetro e os parâmetros são apresentados por linhas. A Fig. 1 ilustra a aparência básica da representação gráfica do JPD.

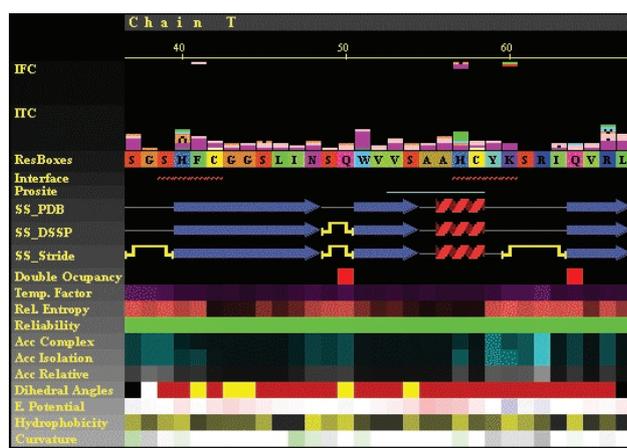


Fig. 1. Organização dos parâmetros no JPD
Ex.: cadeia T, PDB ID 1ldt.

• Sequência de Resíduos

Na linha ResBoxes a seqüência de resíduos relatada no arquivo PDB é apresentada utilizando-se o código

de uma letra para resíduos – *one letter code* (Branden & Tooze, 1999). Cabe observar que os resíduos para os quais a densidade eletrônica no experimento de difração por raios X não permitiu sua visualização são apresentados como *gaps* ou indeterminados.

• Contatos

Resíduos pertencentes a uma proteína podem interagir com outros, pertencentes a mesma proteína ou não, através de diversos tipos de interações químicas fracas ou não covalentes.

Os seguintes tipos de contatos são apresentados pelo JPD:

1. *Hidrofóbicos;*
2. *Eletrostática atrativas;*
3. *Eletrostática repulsivas;*
4. *Pontes de hidrogênio cadeia principal – cadeia principal;*
5. *Pontes de hidrogênio cadeia principal – água – cadeia principal;*
6. *Pontes de hidrogênio cadeia principal – água – água – cadeia principal;*
7. *Pontes de hidrogênio cadeia principal – cadeia lateral;*
8. *Pontes de hidrogênio cadeia principal – água – cadeia lateral;*
9. *Pontes de hidrogênio cadeia principal – água – água – cadeia lateral;*
10. *Pontes de hidrogênio cadeia lateral – cadeia lateral;*
11. *Pontes de hidrogênio cadeia lateral – água – cadeia lateral;*
12. *Pontes de hidrogênio cadeia lateral – água – água – cadeia lateral;*
13. *Interação aromática;*
14. *Pontes de sulfeto.*

O JPD apresenta, para cada resíduo, a quantidade de contatos potenciais em que ele está envolvido, para cada um dos tipos citados. Os contatos para cada resíduo são apresentados em forma de um gráfico de barras onde a altura corresponde à quantidade de contatos e um código de cores é utilizado para distinguir entre os diversos tipos de contatos apresentados.

⇒ **Contatos Internos:** contatos intramoleculares ou internos são aqueles entre resíduos de uma mesma cadeia protéica e são apresentados pelo JPD na linha ITC.

⇒ **Contatos na Interface:** nos casos em que o arquivo PDB possui mais de uma cadeia protéica, formando um complexo molecular, o JPD apresenta uma linha referente a contatos intermoleculares; na linha IFC são apresentados, para cada resíduos de cada cadeia, a quantidade de contatos intermoleculares em que o resíduo está envolvido.

• Interface

Em um complexo de proteínas, a região de interação entre duas cadeias é denominada *Interface* e os resíduos que dela participam são denominados resíduos de *Interface*. Estes resíduos são definidos como aqueles para os quais a área acessível por solvente (Richards, 1977) calculada para as cadeias inseridas no complexo é menor que aquela calculada para as cadeias isoladas. Esta perda de área indica que o resíduo se encontra muito próximo da outra cadeia.

O JPD identifica nesta linha os resíduos que estão na *Interface* grifando-os em vermelho. Ele ainda define a *Extended Interface*, identificada pelos resíduos grifados em azul. A *Extended Interface* é formada por resíduos que não perderam área acessível por solvente quando em complexo, mas possivelmente fazem contatos com resíduos que se encontram na outra cadeia.

• Prosite

O Prosite (Hofmann, 1999) é um banco de dados que identifica padrões de sítios que possuem funções importantes em proteínas conhecidas. Nesta linha, o JPD identifica se o resíduo pertence a algum padrão descrito no Prosite.

• Estrutura Secundária

A estrutura secundária de proteínas refere-se a estruturas tridimensionais locais que ocorrem ao longo da seqüência protéica. Essas estruturas locais se repetem em todas as estruturas protéicas, sendo que os principais tipos são: alfa-hélices (*alpha-helix*), folha-beta (*beta-sheet*) e alça (*turns*). As estruturas secundárias são caracterizadas por padrões de pontes de hidrogênio (*hydrogen bond*) entre o oxigênio do grupo CO de um resíduo com o nitrogênio do grupo NH de um segundo resíduo. Por exemplo, uma alfa hélice é caracterizada por um padrão de ligações por pontes de hidrogênio entre o grupo CO do resíduo **n** com o grupo NH do resíduo **n+4**. A importância dessas estruturas está no fato de que proteínas podem ser classificadas baseando-se no padrão de estruturas

secundárias apresentadas. Para proteínas sem estrutura tridimensional determinada, a predição da estrutura secundária pode ser um primeiro passo para prever a sua estrutura tridimensional. O JPD apresenta três diferentes fontes de determinação de estruturas secundárias:

⇒ **PDB:** na linha SS_PDB são exibidas as estruturas secundárias relatadas pelo autor no arquivo PDB.

⇒ **DSSP:** na linha SS_DSSP são exibidas as estruturas secundárias extraídas do banco de dados *Homology-Derived Secondary Structure of Proteins* (HSSP) (Sander & Schneider, 1991; Schneider et al., 1997) e calculadas de acordo com o método *Dictionary of Secondary Structure of Proteins* (DSSP) (Kabsch & Sander, 1983).

⇒ **Stride:** na linha SS_Stride são exibidas as estruturas secundárias calculadas de acordo com o método *secondary STRucture IDentification* (STRIDE) (Frishman & Argos, 1995).

• Dupla Ocupância

Uma das etapas do processo de determinação da estrutura tridimensional de uma proteína por difração por raios X é a interpretação da densidade eletrônica, obtida através do experimento de difração por raios X. Nesta etapa, o cristalógrafo interpreta o mapa de densidade eletrônica, tentando identificar os resíduos que deram origem àquele mapa. No final, são relatadas as coordenadas atômicas de cada átomo de cada resíduo identificado. Algumas vezes, o cristalógrafo depara-se com a situação em que os átomos de um resíduo parecem ocupar diferentes posições no espaço para as diferentes moléculas de proteína no cristal. Diz-se nestes casos que o resíduo apresenta conformação dupla e todas as coordenadas atômicas são relatadas. Na linha *Double Occupancy* são identificados os resíduos em que esse fenômeno acontece, conforme relatado no arquivo PDB.

• Fator de Temperatura

O Fator de Temperatura indica o grau de mobilidade dos átomos no cristal utilizado no experimento de difração por raios X. Este dado é extraído do arquivo PDB. Na linha *Temp. Factor*, apresentada pelo JPD, utiliza-se o valor médio do resíduo para determinação da cor. No entanto, posicionando-se o *mouse* sobre esta linha para um determinado resíduo, serão apresentados os valores de fator de temperatura médio, máximo e do Carbono Alfa para o resíduo escolhido.

• Entropia Relativa

A Entropia Relativa é uma medida do grau de conservação em uma determinada posição da seqüência protéica. Se a entropia relativa é baixa, há uma indicação de que o resíduo tende a permanecer o mesmo durante a evolução. Se a entropia relativa é alta, naquela posição a natureza promoveu diversas mudanças que não alteraram nem a estabilidade e nem a função da proteína. Este parâmetro é extraído do HSSP (Sander & Schneider, 1991; Schneider et al., 1997).

• Confiabilidade

Confiabilidade é um indicador de confiança para a entropia relativa, definido para cada posição da seqüência. Este parâmetro é extraído do HSSP e é definido como a razão entre o número de seqüências que participam do alinhamento utilizado para cálculo da entropia relativa e o número total de seqüências utilizadas no alinhamento. Se o valor de *reliability* é baixo, isto indica que a informação de entropia relativa para aquela posição é pouco confiável; se o valor de confiabilidade é alto, pode-se dizer que a informação sobre a entropia relativa para aquela posição é mais relevante.

• Acessibilidade de Resíduos

Este parâmetro indica a quantidade de área acessível por solvente. Os valores são calculados através do programa SurfV (Sridharan et al., 1992), que utiliza o algoritmo de Shrake & Rupley (1973):

⇒ **Acessibilidade de resíduos para cadeias em complexo:** este valor refere-se à área acessível por solvente para a situação em que a cadeia a que o resíduo pertence participa de um complexo de moléculas.

⇒ **Acessibilidade de resíduos para cadeia isolada:** este valor refere-se à área acessível por solvente para a situação em que a cadeia a que o resíduo pertence está isolada.

⇒ **Acessibilidade relativa de resíduos:** este valor é a razão entre o valor de área acessível por solvente para a situação em que a cadeia a que o resíduo pertence está isolada e o valor máximo possível para o resíduo. Isto indica o

percentual da área do resíduo exposta ao solvente.

• Ângulos de torsão

A coloração da linha *Dihedral Angles* indica a região em que o resíduo se encontra no *Ramachandran Plot* - gráfico Phi X Psi (Branden & Tooze, 1999). Posicionando-se o *mouse* sobre esta linha, os valores dos ângulos de torsão calculados para o resíduo são apresentados.

• Potencial Eletrostático

O Potencial eletrostático na superfície da proteína é calculado utilizando-se o programa Grasp (Nicholls et al., 1991). Ele demonstra a afinidade de determinadas regiões da superfície da proteína por determinados tipos de carga – positiva ou negativa, o que é importante para a compreensão dos mecanismos de interação intermolecular.

• Curvatura na Superfície

A superfície de uma proteína não é plana, ela possui regiões côncavas e convexas. Esta característica é mapeada no JPD através do parâmetro curvatura.

O cálculo da curvatura é feito utilizando-se o programa SurfRace (Tsodikov et al., 2002) e considerando-se os LHAs (*Last Heavy Atoms*) de cada aminoácido. Esta propriedade pode ser utilizada para determinação de regiões com cavidades ou saliências na superfície protéica, e que são um primeiro indicativo de que estas regiões constituem-se em sítios importantes para interação com outras moléculas.

• Hidrofobicidade

A hidrofobicidade indica o grau de repulsão à água que um resíduo possui. Os resíduos hidrofóbicos, por não terem afinidade com a água, tendem a se agrupar no interior da molécula protéica, assim como os resíduos hidrofílicos, que têm afinidade com água, tendem a ficar na parte exterior. Os valores de hidrofobicidade para cada aminoácido, de acordo com a escala de hidrofobicidade, determinada experimentalmente por Radzicka & Wolfenden (1988), são apresentados pelo JPD segundo uma escala de cores, oferecendo uma visão geral de como esta propriedade se comporta ao longo da seqüência.

Analisando com Maior Detalhes os Parâmetros Apresentados

O JPD oferece uma visão global e qualitativa de um conjunto de parâmetros protéicos, apresentados através de um gráfico, mas também oferece funcionalidades que permitem a análise mais detalhada desses parâmetros.

Posicionando o *mouse* na linha correspondente à propriedade de interesse e apertando o botão direito, um *pop-up menu* apresenta as opções de invocação de módulos do SMS para análise específica do parâmetro (Fig. 2).

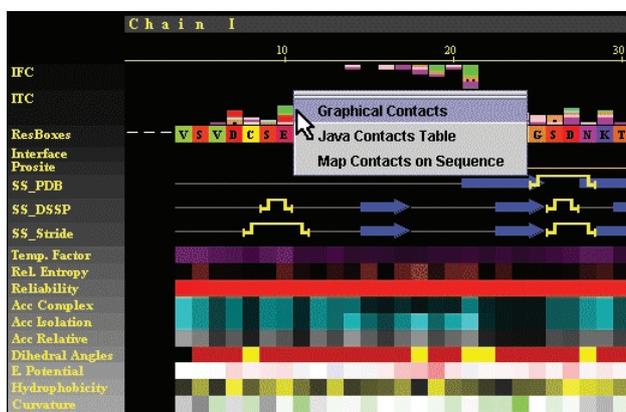


Fig. 2. Pop-up Menu da linha ITC —InternalContacts.

Além disso, posicionando-se o *mouse* sobre a linha correspondente ao parâmetro, o valor exato correspondente à propriedade é apresentado (Fig. 3).

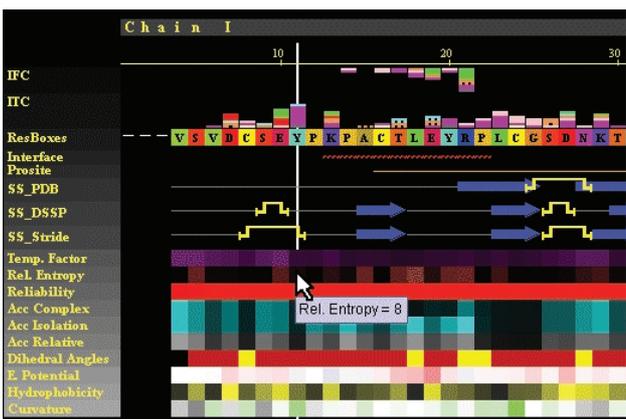


Fig. 3. Hint informando o valor de entropia relativa para a Tirosina número 11 da cadeia I.

Dentre as principais funcionalidades oferecidas pelo JPD, pode-se citar:

• Função “Sting It”

Esta função permite o mapeamento e a visualização de resíduos ou parâmetros apresentados na estrutura

tridimensional da proteína. Isso permite, por exemplo, a identificação na estrutura das áreas mais conservadas na proteína (Fig. 4).

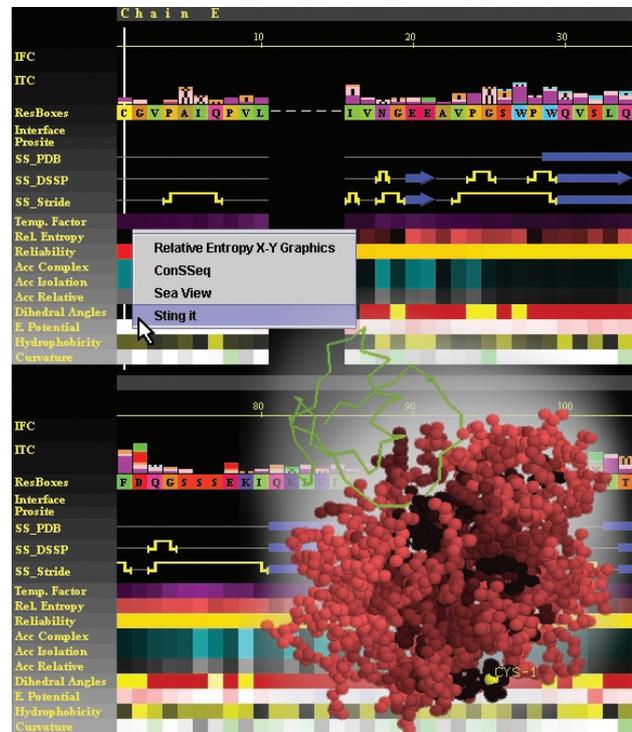


Fig. 4. Pop-up menu para a linha Rel. Entropy com a função Sting it e o resultado apresentado no Chime. As regiões mais escuras indicam a predominância de resíduos mais preservados.

• Funções “Split On” e “Two PDBs”

No seu modo de operação normal, o JPD exibe todas as cadeias da proteína em um único painel e uma barra de rolagem vertical permite que os parâmetros para todos os resíduos de cada cadeia sejam visualizados. Já o modo *Split On*, permite que um arquivo PDB que descreva um complexo molecular (duas ou mais moléculas interagindo) possam ser comparadas. Neste modo, a tela é dividida em dois painéis, cada um apresentando os parâmetros referentes a uma proteína. Barras de rolagem horizontal permitem a visualização dos parâmetros para todos os resíduos de cada cadeia.

O processo de modelagem da estrutura tridimensional de uma proteína por homologia é realizado de forma interativa, tal que a cada iteração o modelo é refinado. Esse refinamento é realizado comparando-se o modelo criado com a estrutura utilizada como *template*. Com a função *Two PDBs*, é possível comparar a estrutura modelada com a estrutura do *template*. A apresentação é a mesma do modo *Split On*, mas em um painel é apresentado o modelo e no outro o *template*. Na Fig. 5 o JPD é apresentado em modo *Split On*.

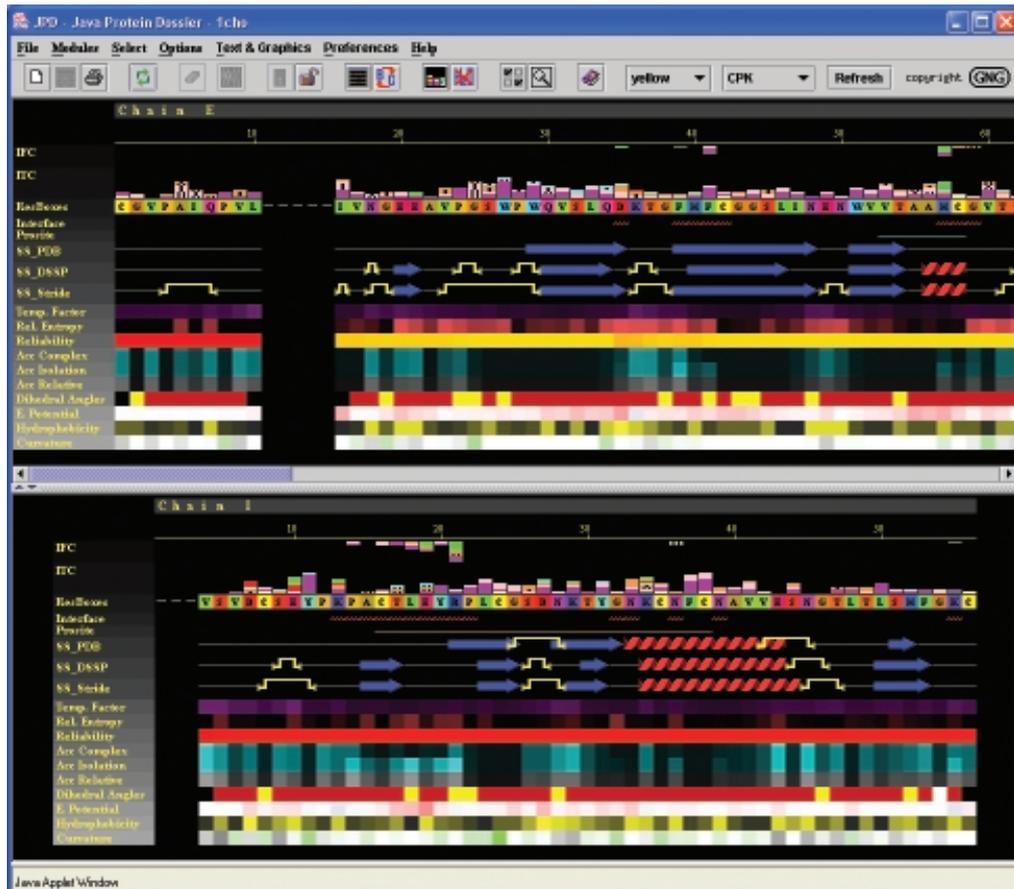


Fig.5. JPD em modo *Split On*.

• Função "Map Contacts on Sequence"

A estrutura tridimensional das proteínas é estabilizada através de interações fracas ou não covalentes, os contatos. Essas forças também são as responsáveis por manter duas moléculas unidas num processo de interação intermolecular. Daí a importância deste parâmetro.

A funcionalidade *Map Contacts on Sequence* permite que se verifique rapidamente todos os potenciais contatos

para um determinado resíduo. Quando ativada, esta função marca o resíduo de interesse com a cor branca e todos os resíduos que potencialmente fazem contato com ele com a cor azul. Além disso, nos gráficos de barras que indicam a quantidade de contatos, as barras correspondentes são apresentadas de forma intermitente. Pode-se então ter uma idéia sobre os tipos e a quantidade de contatos que cada resíduo faz com o resíduo de interesse (Fig. 6). Também é possível mapear os contatos de *Interface* entre duas cadeias de maneira similar.

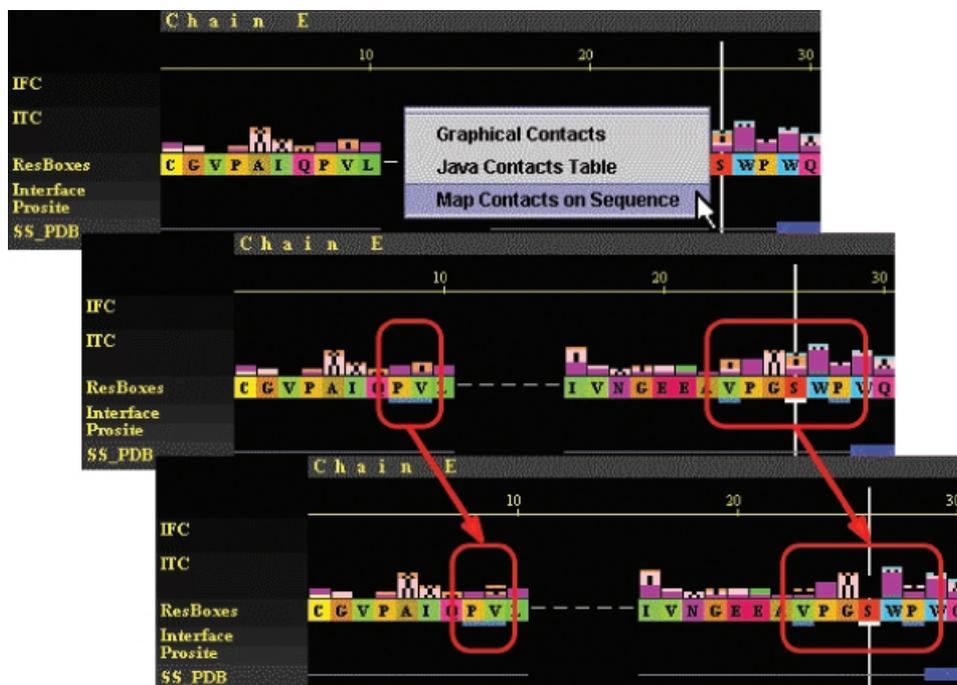


Fig. 6. Utilização da função "Map Contacts on Sequence".

• Função “Select Residues”

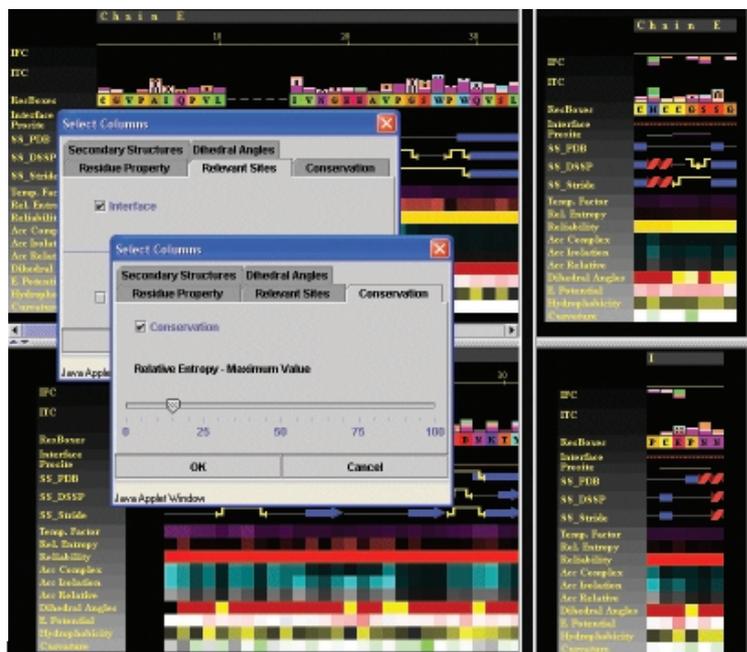
Em geral, as proteínas contêm alguns resíduos que tem uma importância maior para a estabilidade de sua estrutura ou a realização de sua função. Um dos objetivos do JPD é exatamente apresentar um conjunto de parâmetros capaz de fornecer subsídios para a identificação destes resíduos. Por exemplo, resíduos conservados (baixa entropia relativa) na superfície pode ser indicativo de um sítio ativo.

A função *Select Residues* permite que resíduos sejam selecionados especificando um conjunto de restrições baseadas nos parâmetros apresentados pelo JPD. Uma vez realizado este processo de seleção, pode-se analisar os parâmetros correspondentes e sua localização na

estrutura tridimensional da proteína. A Fig. 7 ilustra esta funcionalidade.

• Integração com o *Sting Millennium Suite*

Sendo parte do SMS, o JPD permite que seus módulos sejam invocados para análise mais detalhada dos parâmetros apresentados. Assim, por exemplo, para analisar mais detalhadamente os contatos, pode-se invocar o módulo *Graphical Contacts*; para avaliar a qualidade da estrutura tridimensional, através da análise dos *Dihedral Angles*, pode-se invocar o módulo *Ramachandran Plot* (Fig. 8); e para analisar a entropia relativa, pode-se invocar o módulo ConSSeq.



tententes à *Interface* e com de baixa entropia relativa. O resultado da operação é apresentado do lado direito.

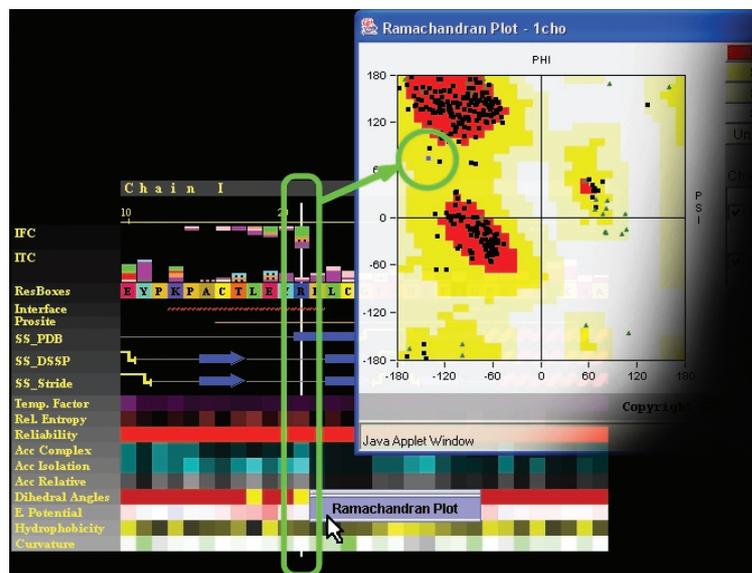


Fig. 8. Invocação do módulo do SMS *Ramachandran Plot* através do JPD.

Discussão e Trabalhos Futuros

Existem diversos programas, disponíveis através da Web ou não, que oferecem os tipos de dados apresentado pelo JPD. Muitos deles são de domínio público, sendo que alguns são utilizados pelo próprio JPD para geração dos dados apresentados.

O *web site* OCA (2002) oferece um serviço para determinação de contatos, via Web (Sobolev et al., 1999). Eles são classificados e calculados de forma diferente que os contatos calculados pelo SMS, não havendo uma correspondência bi-unívoca. Os tipos de contatos calculados pela OCA são: pontes de hidrogênio, interação aromático-aromático, hidrofóbico e hidrofóbico-hidrofílico (contato desestabilizador). As maiores restrições para utilização deste serviço são o fato da saída ser fornecida como texto e de ser necessário solicitar o cálculo dos contatos para um número máximo de dez resíduos. Não há uma opção para apresentação de todos os contatos, como no caso do JPD.

O Prosite (Hofmann et al., 1999), o HSSP (Sander & Schneider, 1991; Schneider et al., 1997) e o DSSP (Kabsch & Sander, 1983) e mesmo o PDB (Berman et al., 2000) têm seus dados disponíveis através de *web sites*. Novamente, a maior restrição relativo ao acesso direto a estes *web sites* é que as informações estão disponíveis como texto, cabendo aos seus usuários mapear as informações tanto na estrutura quanto na seqüência. O mesmo é válido para o STRIDE (Frishman & Argos, 1995), um programa de domínio público para determinação de estruturas secundárias, que apresenta seus resultados como um arquivo textual.

Para o cálculo da área acessível por solvente, existem uma série de algoritmos e implementações disponíveis (Connolly, 2002). O JPD utiliza o programa proprietário SURFV (Sridharan et al., 1992) que, como seus similares, também apresenta seus resultados em formato texto.

O programa gráfico GRASP (Nicholls et al., 1992) é utilizado para cálculo de propriedades como potencial eletrostático e hidrofobicidade na superfície da proteína. Ele permite que essas propriedades, mapeadas na superfície da proteína, sejam estudadas de forma gráfica e interativa. Entretanto, apenas uma propriedade por vez pode ser mapeada na superfície da proteína. Além disso, ele é um programa proprietário e está disponível apenas para a plataforma SGI/IRIX. Uma interface Web para algumas das funcionalidades do GRASP está disponível através do GRASS (Honig Lab at Columbia University, 2002).

O *web site*, cuja funcionalidade mais se aproxima da que o JPD oferece é o PDBSum (Laskowski et al., 1997). Ele apresenta um sumário de informações sobre as estruturas depositadas no PDB. As estruturas secundárias relatadas no PDB, sítios importantes e um

gráfico do tipo *Ramachandran Plot* (Branden & Tooze, 1999) são disponibilizadas. Entretanto, a apresentação é estática, não havendo a possibilidade de se verificar maiores detalhes sobre os parâmetros apresentados.

O mérito maior do JPD é o fato dele reunir uma grande quantidade de informações para análise de estruturas protéicas e apresentá-las de forma gráfica e interativa. Embora, essas informações possam ser obtidas de outras fontes que não o JPD, elas estão dispersas e não são apresentadas de forma integrada o que, muitas vezes, é fundamental para a análise.

Para o futuro, pretende-se expandir o conjunto de parâmetros apresentados, incluindo-se informações sobre contatos entre proteína e DNA — *Desoxyribonucleic acid*, marcação de resíduos que participam de cavidades, identificação de famílias protéicas como definidas por alinhamento como Pfam (Washington University in St. Louis, 2002) e ProDom (Corpet et al., 2000) e inclusão de bibliotecas de rotâmeros⁷.

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, n. 1, p. 235-242, 2000.
- BRANDEN, C.-U.; TOOZE, J. **Introduction to protein structure**. 2nd ed. New York: Garland Science, 1999. 410 p.
- CONNOLLY, M. L. **Molecular surface**: a review. Disponível em: <<http://www.netsci.org/Science/Compchem/feature14.html>>. Acesso em: 20 dez. 2002.
- CORPET, F.; SERVANT, F.; GOUZY, J.; KAHN, D. ProDom and ProDom-CG: tools for protein domain analysis and whole genome comparisons. **Nucleic Acids Research**, v. 28, n. 1, p. 267-269, 2000.
- FRISHMAN, D.; ARGOS, P. Knowledge-based protein secondary structure assignment. **Proteins: Struct., Func. Genetics**, v. 23, n. 4, p. 566-579, 1995.
- HOFFMANN, K.; BUCHER, P.; FALQUET, L.; BAIROCH, A. The PROSITE database, its status in 1999. **Nucleic Acids Research**, v. 27, n. 1, p. 215-219, 1999.

⁷ Rotâmeros são conformações preferenciais assumidas por resíduos de aminoácidos. Cada um dos 20 aminoácidos possui um conjunto de conformações preferenciais ou rotâmeros.

HONIG LAB AT COLUMBIA UNIVERSITY. **GRASS browser – Graphical Representation and Analysis of Structure Server**. Disponível em: <http://honiglab.cpmc.columbia.edu/cgi-bin/GRASS/surfserv_enter.cgi>. Acesso em: 20 dez. 2002.

KABSCH, W.; SANDER, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. **Biopolymers**, v. 22, n. 12, p. 2577-2637, 1983.

LASKOWSKI, R. A.; HUTCHINSON, E. G.; MICHIE, A. D.; WALLACE, A. C.; JONES, M. L.; THORNTON, J. M. PDBsum: a web-based database of summaries and analyses of all PDB structures. **Trends Biochemistry Science**, v. 22, n. 12, p. 488-490, 1997.

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

NICHOLLS, A.; SHARP, K. A.; HONIG, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons.

Proteins: Struct., Func. and Genet., v. 11, n. 4, p. 281-296, 1991.

OCA. **OCA browser & database**. Disponível em: <<http://bioinfo.weizmann.ac.il:8500/oca-docs/>>. Acesso em: 20 dez. 2002.

RADZICKA, A.; WOLFENDEN, R. Comparing the polarities of the amino-acids - side-chain distribution coefficients between the vapor-phase, cyclohexane, 1-octanol, and neutral aqueous-solution. **Biochemistry**, v. 27, p. 1664-1670, 1988.

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

SANDER, C.; SCHNEIDER, R. Database of homology-derived protein structures and the structural meaning of sequence alignment. **Proteins: Struct. Function and Genetics**, v. 9, p. 56-68, 1991.

SCHNEIDER, R.; DE DARUVAR, A.; SANDER, C. The HSSP database of protein structure sequence alignments. **Nucleic Acids Research**, v. 25, n. 1, p. 226-230, 1997.

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.

SOBOLEV, V.; SOROKINE, A.; PRILUSKY, J.; ABOLA, E. E.; EDELMAN, M. Automated analysis of interatomic contacts in proteins. **Bioinformatics**, v. 15, p. 327-332, 1999.

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.cnptia.embrapa.br/>>. Acesso em: 21 dez. 2002.

SUN MICROSYSTEMS. **The source for Java technology**. Disponível em: <<http://java.sun.com/>>. Acesso em: 21 dez. 2002.

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.

WASHINGTON UNIVERSITY IN ST. LOUIS. **The Pfam database of protein families and HMMs**. Disponível em: <<http://pfam.wustl.edu/>>. Acesso em: 20 dez. 2002.

Comunicado Técnico, 40

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