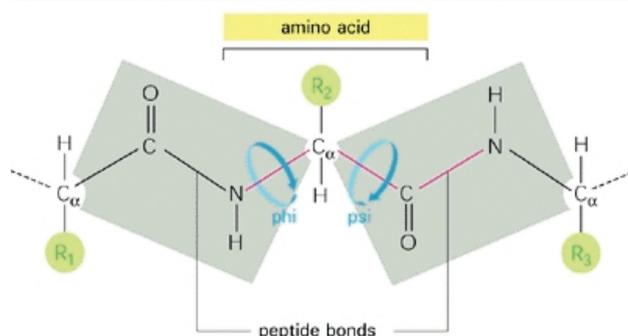


SMSLib**BIBLIOTECA C++**

ISSN 1677-8464

SMSLib – Biblioteca C++ do Sting Millennium Suite

Roberto Hiroshi Higa¹Christian Baudet²Adauto Luiz Mancini³Amanda Remmez Mattiuz⁴Esther Manoela de Freitas⁵Paula Kuser Falcão⁶Goran Neshich⁷

O *Sting Millennium Suite* (SMS) (Structural Bioinformatics Group, 2002) compreende um conjunto de programas para análise da estrutura de proteínas acessível através da Web. Ele permite a visualização simultânea da estrutura e da seqüência de proteínas e um conjunto de propriedades físico-químicas mapeadas para cada resíduo, entre outras funcionalidades. O seu desenvolvimento iniciou-se em 1998 na Columbia University e a partir do final do ano de 2000 passou a ser mantido e desenvolvido na Embrapa Informática Agropecuária, sendo o principal produto do Núcleo de Bioinformática Estrutural (NBI). A partir do segundo semestre de 2001, ele passou a ser espelhado pelo site do *Protein Data Bank* (PDB) (Berman et al., 2000).

A versão atual do SMS é a 2.2, sendo que a versão 3.0 encontra-se em estágio bastante avançado de desenvolvimento. As novas funcionalidades incorporadas a estas versões compreendem:

- redefinição dos critérios para cálculo de contatos, com inclusão de novos tipos como empilhamento aromático (*aromatic stacking*), pontes de hidrogênio intermediadas por até duas moléculas de água;
- versão Java do *Protein Dossier* – Java *Protein Dossier* que além da visualização da representação gráfica das propriedades reportadas, permite a seleção de aminoácidos de interesse (ex.: aminoácidos na interface de um complexo) e o acesso direto a ferramentas mais específicas para uma análise mais profunda de cada propriedade (ex: *Graphical Contacts* para análise de contatos);
- incorporação de uma série de novas propriedades apresentadas no Java *Protein Dossier* como potencial eletrostático, curvatura e hidrofobicidade.

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 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)

⁴ Estudante de Engenharia da Computação, Estagiário da Embrapa Informática Agropecuária, Caixa Postal 6041, Barão Geraldo - 13083-970 - Campinas, SP. (e-mail: amanda@cnptia.embrapa.br)

⁵ Estudante do Curso Técnico de Informática, Estagiário da Embrapa Informática Agropecuária, Caixa Postal 6041, Barão Geraldo - 13083-970 - Campinas, SP. (e-mail: esther@cnptia.embrapa.br)

⁶ 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)

⁷ Ph.D. em Biofísica, Pesquisador da Embrapa Informática Agropecuária, Caixa Postal 6041, Barão Geraldo - 13083-970 - Campinas, SP. (e-mail: neshich@cnptia.embrapa.br)

Durante este período ainda, o SMS passou a ser um produto multiplataforma, tendo seus programas sido portados, a partir da plataforma original SGI/IRIX, para Sun/Solaris e PC/Linux.

A necessidade de evolução constante do produto, agora multiplataforma, exigiu a realização de uma reestruturação parcial do SMS, no sentido de torná-lo mais robusto e de mais fácil manutenção e evolução. Este trabalho, refere-se a uma dessas alterações, especificamente à reestruturação do código dos programas utilizados pelo SMS e criação de uma biblioteca, denominada SMSLib, cuja funcionalidade compreende o núcleo de todos os programas do SMS como leitores de arquivos de bancos de dados de domínio público e cálculo de alguns dos parâmetros apresentados pelo SMS.

Nas seções seguintes serão apresentadas a arquitetura geral do SMS, uma descrição das principais classes que atualmente compõem a SMSLib e, finalmente, uma discussão sobre os resultados obtidos até o momento e os planos de desenvolvimento futuro.

Organização lógica do SMS

O SMS é organizado em duas camadas lógicas (Fig. 1): o *SMS server* e o *SMS client*. O *SMS server* é responsável pela atualização regular de todos os bancos de dados de domínio público utilizados pelo SMS, como o banco de dados de estrutura tridimensional de proteínas *Protein Data Bank* (PDB) (Berman et al., 2000), o banco de dados de proteínas com informações sobre estrutura secundária inferidas por homologia *Homology-derived Secondary Structure of Proteins* (HSSP) (Sander & Schneider, 1991; Schneider et al., 1997) e o banco de dados de padrões de sítios importantes em proteínas PROSITE (Hofmann et al., 1999). Utilizando esses bancos de dados, o *SMS server* também calcula uma série de propriedades macromoleculares para cada estrutura de proteína depositada no banco de dados PDB, como potencial eletrostático, curvatura, área acessível para solvente para cada cadeia e para todo o complexo, as estruturas secundárias, calculadas pelos métodos *Dictionary of Protein Secondary Structure* (DSSP) (Kabsch & Sander, 1983) e *Secondary STRuctural IDentification* (STRIDE) (Frishman & Argos, 1995), contatos intra e inter cadeia, interações proteína/dna, hidrofobicidade, *dihedral angles*, e padrões definidos pelo PROSITE. Ele também é responsável por prover acesso ao banco de dados do SMS através do protocolo HTTP (World Wide Web Consortium, 2002). O *SMS client* provê uma interface gráfica amigável, envia as requisições do usuário para o *SMS server* e apresenta as respostas formuladas para análise.

A interação entre o *SMS client* e o *SMS server* acontece da seguinte forma. Um cliente HTTP (*browser*) permite que o usuário selecione um módulo SMS e especifique o código de identificação do arquivo PDB desejado –

PDB ID. No lado do servidor, um servidor HTTP atende às requisições do *SMS client* transmitidas através da Internet. Quando ele recebe uma requisição de um cliente, ele a redireciona para programas *Common Gateway Interface* (CGI) apropriados. Esses programas CGI realizam a busca pelo arquivo PDB apropriado e, acessando os bancos de dados do SMS, cria a resposta apropriada para a requisição. Essa resposta é, então, transmitida para o cliente para ser apresentada ao usuário.

A resposta é apresentada para o cliente utilizando uma série de recursos como HTML 3.2 —*Hyper Text Markup Language* (World Wide Web Consortium, 2002), Java Script 1.2 (Flanagan, 1998) e java 2 (Sun Microsystems, 2002) e o MDL *Chime Plugin*⁸ (MDL Information Systems, 2002), dependendo do módulo requisitado. Todos estes recursos constituem uma interface amigável, que permite ao usuário explorar e analisar os resultados de forma simples e rápida.

Dentro da organização do SMS, a SMSLib está basicamente vinculada à parte servidora do SMS, sendo utilizada para criação e leitura dos parâmetros que compõem o banco de dados SMS.

Descrição da SMSLib

As classes que compõem a SMSLib são basicamente de 2 tipos: a) classes que abstraem informações oriundas de bases de dados públicas, como o PDB e o HSSP, ou de arquivos gerados com a utilização de programas de terceiros, como o STRIDE (Frishman & Argos, 1995) para determinação de estruturas secundárias, o GRASP (Nicholls et al., 1991) para cálculo de potencial eletrostático na superfície da estrutura da proteína e o SurfRace (Tsodikov et al., 2002) para cálculo da curvatura na superfície da estrutura da proteína; e b) classes que além de abstraírem informações contidas em arquivos, também calculam os parâmetros nele contidos, como o Contacts (calcula as potenciais ligações não covalentes ou fracas entre átomos dos resíduos que constituem a proteína) e Dihedral Angles (calcula os ângulos de rotação em torno do carbono alfa para cada resíduo da proteína).

Leitura de arquivos PDB

O *Protein Data Bank* (PDB) é o banco de dados mundial onde são depositados dados estruturais de macromoléculas biológicas (Berman et al., 2000). Nele são armazenadas todas as estruturas de proteínas isoladas ou formando um complexo com outras moléculas como DNA, outras proteínas ou pequenas moléculas, resolvidas em geral por experimentos de difração de raios X ou ressonância magnética.

⁸ Plugin de domínio público, desenvolvido pela Molecular Design Limited Inc. (MDL) para visualização de estruturas de macromoléculas descritas em formato PDB.

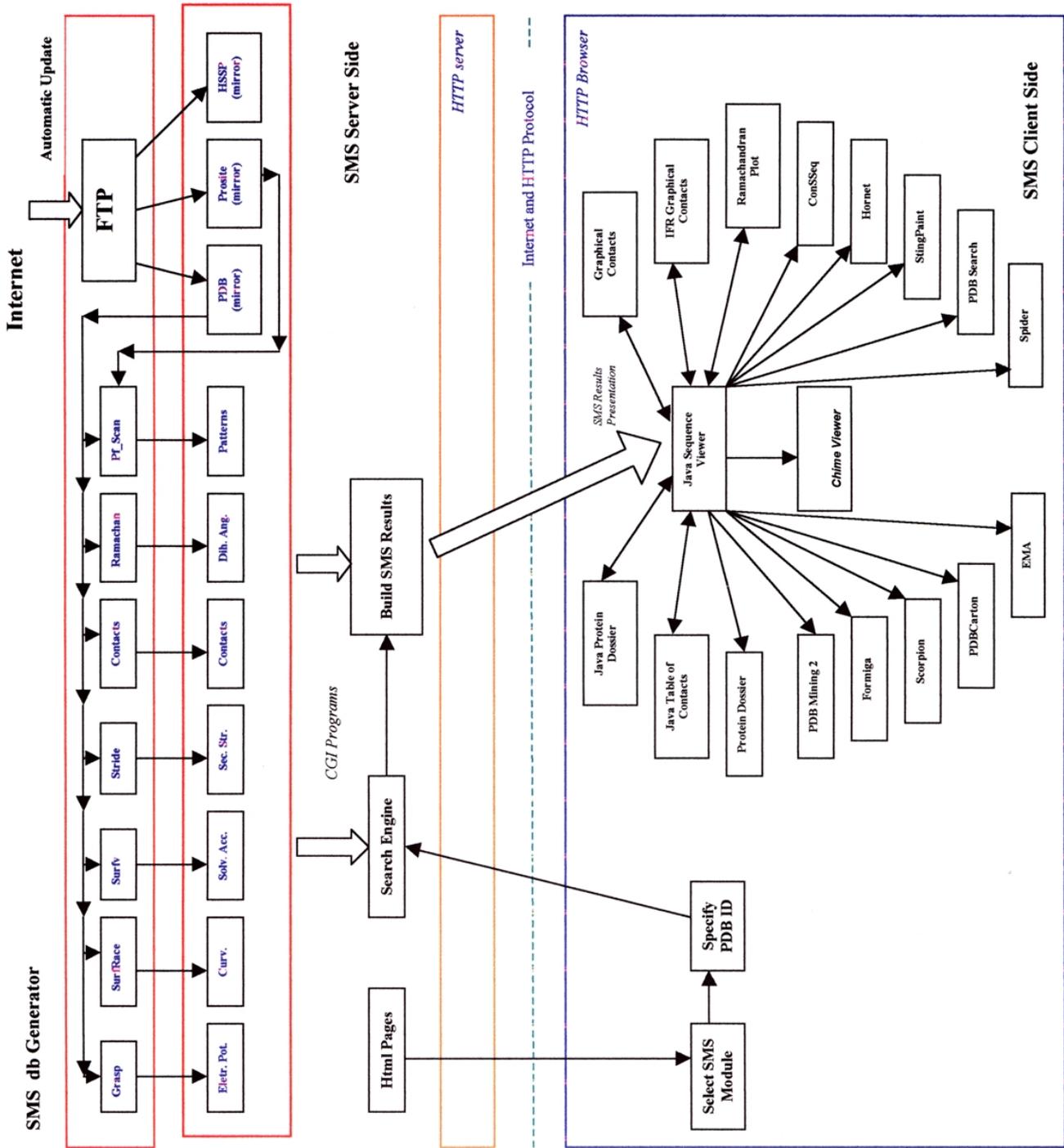


Fig. 1. Organização Lógica do SMS.

Cada estrutura depositada recebe um código de identificação composto por 4 caracteres (1 – 9; a - z), como por exemplo –“1cho”, e está disponível para acesso público através da Web ou via ftp. Dentre os formatos utilizados para distribuição dos dados, o mais utilizado é o chamado formato PDB, um arquivo texto formatado padrão (Research Collaboratory for Structural Bioinformatics, 2002).

Cada arquivo PDB, contém informações sobre a estrutura de uma macromolécula, bem como sobre o experimento utilizado para sua determinação. Ele é composto pelas seguintes seções:

- **Title Section**, que descreve a(s) macromolécula(s). Também inclui o depositante, as publicações relacionadas e palavras-chave;
- **Primary Structure Section**, que descreve a sequência de resíduos em cada cadeia de macromolécula;
- **Heterogen Section**, que descreve resíduos não padrão (resíduos que não sejam nem protéicos, nem de DNA ou RNA, alterados ou não, e nem um resíduo denominado como desconhecido – UNK);

- **Secondary Structure Section**, que descreve as estruturas secundárias – TURN, SHEET e HELIX encontradas na(s) macromolécula(s);
- **Connectivity Annotation Section**, que relata as ligações observadas, como pontes de hidrogênio e de dissulfeto;
- **Crystallographic and Coordinate Transformation Section**, que descreve as transformações do experimento cristalográfico e as transformações do sistema de coordenadas;
- **Coordinate Section**, que contém as coordenadas atômicas dos átomos da(s) macromolécula(s), inclusive para os diversos modelos obtidos através de experimentos de ressonância nuclear magnética; e
- **Connectivity Section**, que fornece informações de conectividade para heteroátomos.

A Fig. 2 apresenta o diagrama de classes simplificado com as classes da SMSLib responsáveis pela leitura do arquivo PDB e abstração do seu conteúdo.

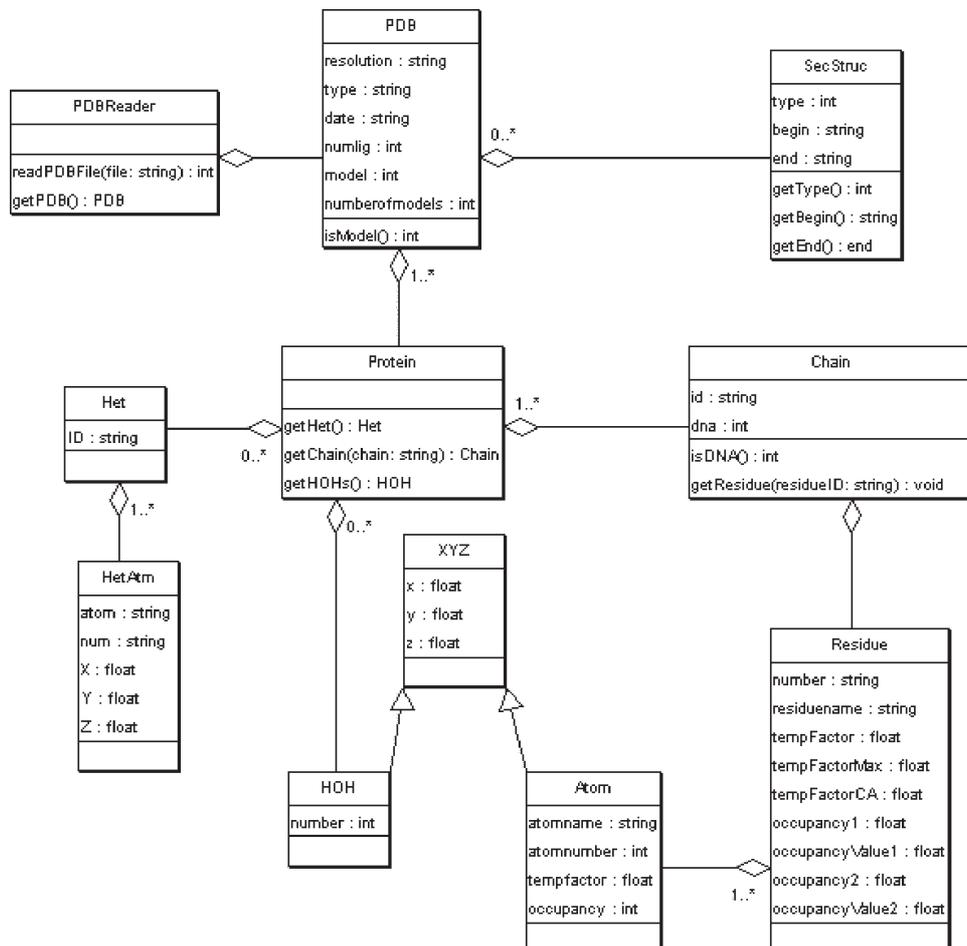


Fig. 2. Leitura de arquivo PDB - diagrama de classes simplificado.

A classe *PDBReader* realiza a leitura de um arquivo PDB criando um objeto da classe *PDB* com as informações sobre o experimento para determinação da estrutura da(s) macromolécula(s) que o arquivo descreve. A classe *Protein* abstrai o conjunto de macromoléculas descritas no arquivo PDB e identificadas como cadeias (classe *Chain*), que é constituído por um conjunto de resíduos (classe *Residue*). Um resíduo, por sua vez, é constituído por um conjunto de átomos (classe *Atom*). O conteúdo do arquivo PDB, agora representado por um conjunto de objetos, é acessado navegando-se através dos objetos criados pelo *PDBReader* durante a leitura do arquivo PDB e acessando os valores de seus atributos.

Leitura de arquivos HSSP

O HSSP é construído agregando-se informações oriundas de bancos de dados de estrutura protéica com informações de bancos de dados de seqüências protéicas. Para cada proteína com estrutura depositada no PDB, é realizada uma busca por seqüências homólogas no banco de dados de seqüência protéicas SWISS-PROT (Bairoch & Apweiler, 2000) e construído um alinhamento múltiplo com estas seqüências (Sander & Schneider, 1991; Schneider et al., 1997). O arquivo HSSP relata ainda a definição de estruturas secundárias, conforme calculado pelo DSSP (Kabash & Sander, 1983), a freqüência relativa de ocorrência dos diferentes tipos de aminoácidos para cada posição da seqüência do PDB, e a entropia relativa⁹ para cada posição da proteína original do PDB considerando o alinhamento obtido. Ele é dividido nas seguintes seções:

- **HEADERS**, que indica a data de liberação do arquivo HSSP, a proteína do PDB a que se refere o arquivo, os parâmetros utilizados pelos programas na geração do alinhamento;

- **PROTEINS**, que lista todas as proteínas alinhadas, indicando os resíduos de início e final do alinhamento na seqüência do PDB e do Swiss-Prot.
- **ALIGNMENTS**, que contém o alinhamento propriamente dito. Além disso, para cada posição do alinhamento é indicado o correspondente tipo de estrutura secundária, conforme calculado pelo DSSP; e
- **SEQUENCE PROFILE**, que contém a matriz (*profile*) com a freqüência relativa de ocorrência de tipos de aminoácidos para cada posição da seqüência do PDB, um peso de conservação, a entropia relativa e a ocupância (número de seqüências no alinhamento em que a posição não é ocupada por *gap*¹⁰) para cada posição do alinhamento.

A Fig. 3 apresenta o diagrama de classes simplificado com as classes da SMSLib responsáveis pela leitura do arquivo HSSP e abstração do seu conteúdo. A classe *HSSPReader* realiza a leitura de um arquivo HSSP criando objetos das classes *MSAHandler* e *HSSPSequenceProp*, que abstraem as informações contidas no arquivo. As informações do arquivo HSSP são abstraídas basicamente como um alinhamento, representado pela classe *MSAHandler* e o conjunto de propriedades relativo a cada resíduo das macromoléculas do arquivo PDB pela classe *HSSPSequenceProp*. Para facilidade de manipulação, um método da classe *HSSPReader* cria objetos da classe *DSSP* que representam as estruturas calculadas de acordo com o DSSP. O conteúdo do arquivo HSSP, agora representado por um conjunto de objetos, é acessado navegando-se através dos objetos criados pelo *HSSPReader* durante a leitura do arquivo HSSP e acessando os valores de seus atributos. Tanto a classe *MSAHandler* quanto a classe *HSSPSequenceProp* possuem métodos que retornam um objeto da mesma classe, mas com dados referentes apenas a uma cadeia determinada.

⁹ A entropia relativa é uma medida de variabilidade, sendo calculada da seguinte forma:

$$relent(i) = \frac{S(i)}{\ln 20} \quad \text{onde:}$$

$$S(i) = \sum_R f_{Ri} \ln f_{Ri}$$

f_{Ri} é a freqüência de ocorrência do tipo de resíduo R na posição i do alinhamento.

¹⁰ Ao fazer o alinhamento entre duas ou mais seqüências, em algumas posições do alinhamento algumas seqüências não apresentam o elemento correspondente. Esta posição, então, é ocupada por um *gap*. Por exemplo, no melhor alinhamento entre as seqüências AVTP e AVP, a segunda seqüência terá um *gap* inserido na terceira posição e aparecerá no alinhamento como AV-P.

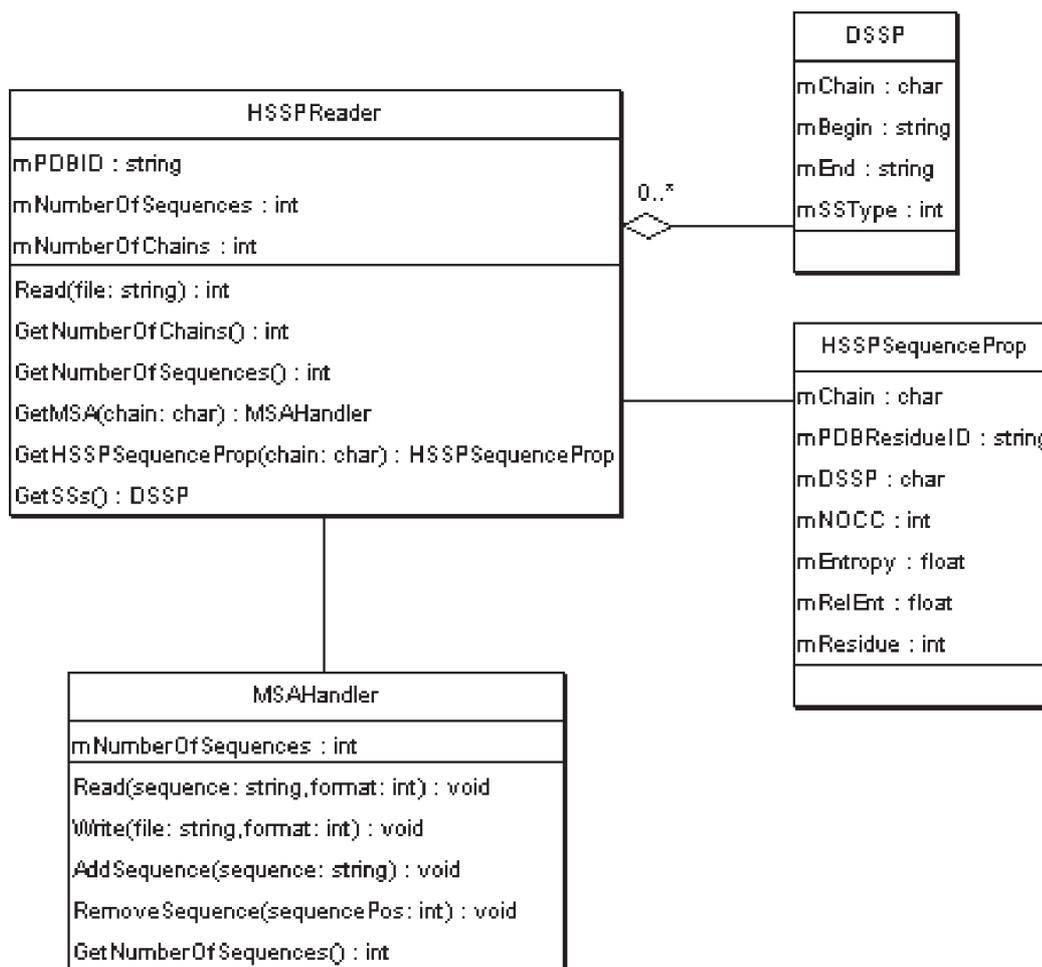


Fig. 3. Leitura de arquivo HSSP – diagrama de classes simplificado.

Leitura de Arquivos com Parâmetros Simples

O SMS utiliza também o resultado da execução de uma série de softwares de terceiros que calculam parâmetros relevantes para a análise da estrutura de proteínas. Os resultados obtidos da execução destes programas são armazenados em arquivos específicos para cada parâmetro, sendo que o seu formato depende do software utilizado. Este formato, em geral, é organizado por colunas, uma representando o código de identificação do aminoácido e outra representando o valor do parâmetro de interesse. Esses dados são acessados através de classes de leitura específica, cuja função é ler o conteúdo do arquivo e criar uma estrutura de dados em memória para acesso ao parâmetro de interesse.

Atualmente, os seguintes parâmetros são obtidos e manipulados conforme a descrição acima:

- definição de estruturas secundárias calculadas pelo programa STRIDE (Frishman & Argos, 1995);
- identificação de sítios do PROSITE identificados pelo programa pfsan (Hofmann, 1999);
- o cálculo do potencial eletrostático calculado pelo programa GRASP (Nicholls et al., 1991);
- o cálculo da curvatura calculada pelo programa SurfRace (Tsodikov et al., 2002); e
- área acessível para solvente calculada pelo programa SURFV (Sridharan et al., 1992).

A Fig. 4 apresenta o diagrama de classes simplificado das classes de leitura para arquivos com área acessível para solvente - AIRReader, estrutura secundária calculada pelo STRIDE - Stride e sítios do PROSITE - prositeReader, respectivamente. Os demais parâmetros deste tipo são lidos e manipulados por classes com funcionalidade similar.

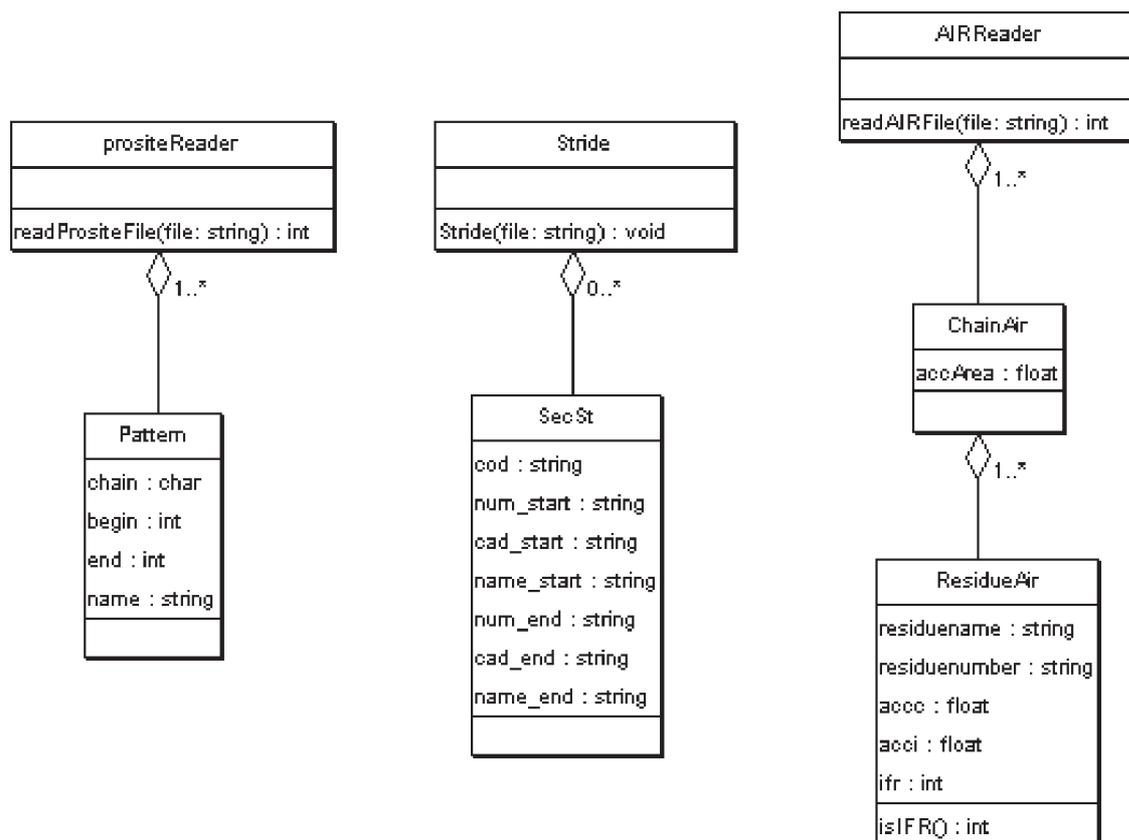


Fig. 4. Leitura de arquivos produzidos por programas externos – diagrama de classes simplificado.

Cálculo e Leitura de Contatos

A estrutura tridimensional de uma proteína é estabilizada através de ligações ditas fracas ou não covalentes, também denominadas contatos (Branden & Tooze, 1999). O cálculo desses contatos procura pares de átomos com afinidade química (*donnor* e *acceptor*) e que, ao mesmo tempo, se encontram a uma distância apropriada para formarem ligações não covalentes. Os tipos de contatos considerados são (Structural Bioinformatics Group, 2002):

- interação hidrofóbica;
- pontes de hidrogênio simples ou com uma ou duas moléculas de água intermediárias;
- eletrostática atrativa;
- eletrostática repulsiva;

- pontes de sulfeto; e
- interação aromática (*aromatic ring stacking*).

A Fig. 5 apresenta o diagrama simplificado de classes para manipulação de contatos. A classe *Contacts* é responsável pela geração dos diversos tipos de contatos, bem como pela leitura e gravação dos dados em arquivo. Após o seu cálculo, os contatos são organizados em memória por tipo de contato. A classe *Contacts* cria um objeto da classe *Section* para cada cadeia do arquivo PDB (contatos intramoleculares) e, no caso de existir mais de uma cadeia, também cria um objeto da classe *Section* para cada par de cadeias distintas (contatos intermoleculares). Dentro da classe *Section*, esses contatos são armazenados por classes específicas, correspondente aos diversos tipos de contatos (Tabela 1).

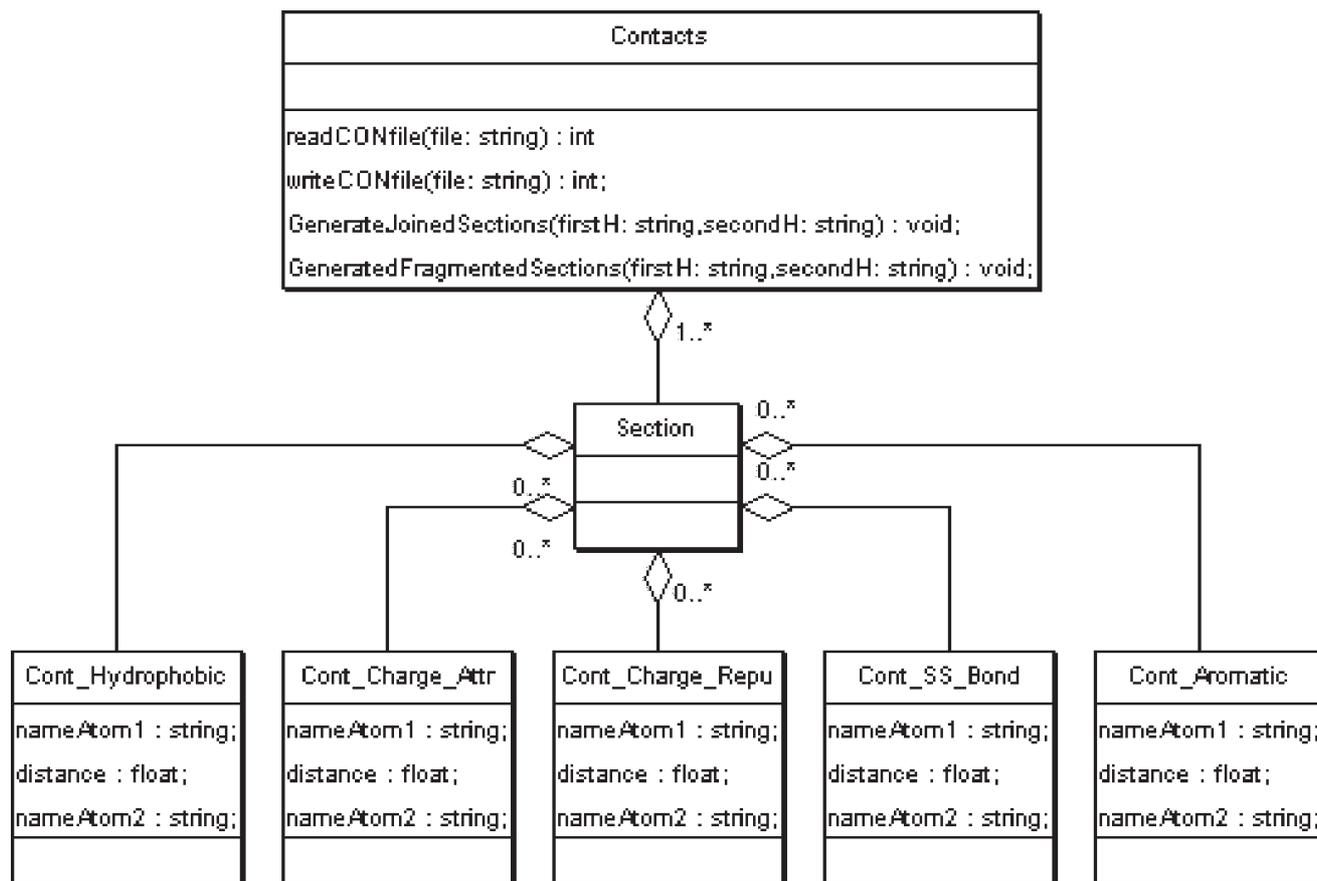


Fig. 5. Contacts – diagrama de classes simplificado.

Tabela 1. Classes para manipulação dos diversos tipos de contatos.

Classe	Função
Cont_Hydrophobic	Classe que representa interação do tipo hidrofóbico.
Cont_Charge_Atr	Classe que representa interação do tipo eletrostático atrativo.
Cont_Charge_Repu	Classe que representa interação do tipo eletrostático repulsivo.
Cont_SS_Bond	Classe que representa interação do tipo ponte de sulfeto.
Cont_Aromatic	Classe que representa interação do tipo aromática.
Cont_HB_MM	Classe que representa interação do tipo ponte de hidrogênio entre átomos da cadeia principal.
Cont_HB_MS	Classe que representa interação do tipo ponte de hidrogênio entre um átomo da cadeia principal e um da cadeia lateral.
Cont_HB_SS	Classe que representa interação do tipo ponte de hidrogênio entre átomos da cadeia lateral.
Cont_HB_MWM_MM	Classe que representa interação do tipo ponte de hidrogênio entre átomos da cadeia principal intermediado por uma molécula de água.
Cont_HB_MWS	Classe que representa interação do tipo ponte de hidrogênio entre um átomo da cadeia principal e um da cadeia lateral intermediado por uma molécula de água.
Cont_HB_SWS	Classe que representa interação do tipo ponte de hidrogênio entre átomos da cadeia lateral intermediados por uma molécula de água.
Cont_HB_MWWW_MM	Classe que representa interação do tipo ponte de hidrogênio entre átomos da cadeia principal intermediado por duas moléculas de água.
Cont_HB_MWWS	Classe que representa interação do tipo ponte de hidrogênio entre um átomo da cadeia principal e um da cadeia lateral intermediado por duas moléculas de água.
Cont_HB_SWWS	Classe que representa interação do tipo ponte de hidrogênio entre átomos da cadeia lateral intermediados por duas moléculas de água.

Cálculo e Leitura de Dihedral Angles

Devido à natureza parcial de ligação dupla da ligação peptídica (ligação CO – NH), a rotação em torno destas ligações é restrita. Assim, as unidades peptídicas de uma proteína são formadas por estruturas planares e a rotação na cadeia principal é restrita às ligações em torno do carbono alfa - C_{α} (Branden & Tooze, 1999). Os dois ângulos de rotação em torno do carbono C_{α} são denominados *dihedral angles* para a cadeia principal e especificados por *phi* (ϕ) e *psi* (ψ) (Fig. 6). Estes ângulos caracterizam a conformação dos átomos da cadeia principal de uma proteína e, devido a possibilidade de colisões estéricas, seus valores são restritos.

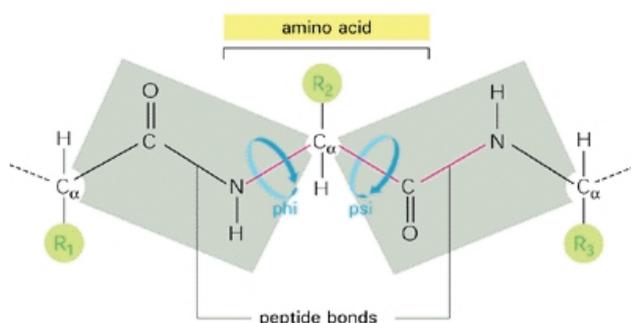


Fig. 6. Dihedral angles – phi e psi.

Para o cálculo dos *dihedral angles* para um resíduo de aminoácido A_n , considera-se o plano formado pelo carbono alfa deste ($C_{\alpha n}$), o nitrogênio (N_n) e o carbono (C_n) deste resíduo como referência. O ângulo ψ_n corresponde ao ângulo entre as normais ao plano de referência e o plano formado pelo carbono alfa ($C_{\alpha n}$) e carbono (C_n) do resíduo A_n e o nitrogênio (N_{n+1}) do resíduo A_{n+1} . O ângulo ϕ_n corresponde ao ângulo entre as normais ao plano de referência e o plano formado pelo carbono alfa ($C_{\alpha n}$) e o nitrogênio (N_n) do resíduo A_n e carbono (C_{n-1}) do resíduo A_{n-1} .

A Fig. 7 apresenta o diagrama de classes simplificado com as classes da SMSLib que lêem e calculam os *dihedral angles*.

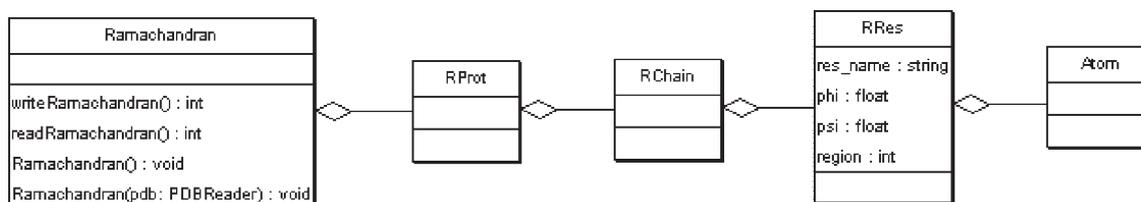


Fig. 7. Dihedral angles – diagrama de classes simplificado.

A classe *Ramachandran* utiliza a classe *PDBReader* para acessar os dados de coordenadas atômicas de uma estrutura protéica, calcula os *dihedral angles* para cada resíduo e determina a correspondente região no **Ramachandran Plot**¹¹ (Branden & Tooze, 1999). Para isso, são criados um conjunto de objetos das classes *RProt*, *RChain*, *RRes* e *RAtom* representando a estrutura protéica e o conjunto de dados necessários para cálculo dos *dihedral angles*. A classe *Ramachandran* pode ainda gravar esses dados em um arquivo e recuperá-los posteriormente, ao invés de recalculá-los. O dados de *dihedral angles*, calculados *on-line* ou lidos de arquivo, podem ser acessados e manipulados navegando-se através dos objetos criados pelo *Ramachandran* e acessando os valores de seus atributos.

Discussão e Trabalhos Futuros

A implementação presente da SMSLib compreende uma parte considerável do núcleo de funcionalidades do SMS e que deve continuar a ser expandido à medida que novas funcionalidades sejam incorporadas.

O fato mais marcante observado até o momento com a criação da SMSLib refere-se aos já esperados benefícios da reutilização de código. A manutenção dos componentes do SMS foi facilitada na medida em que erros corrigidos na SMSLib são automaticamente incorporados por todos os componentes que a utilizam e o desenvolvimento de novos módulos é agilizado na medida em que mais e mais funcionalidades já se encontram disponíveis e não precisam ser reimplementadas.

Das funcionalidades atuais do SMS, apenas o cálculo de superfície acessível por solvente e a determinação de microambiente de aminoácidos ainda não estão

¹¹ Ramachandran Plot é um gráfico phiXpsi onde esses valores são plotados para todos os resíduos da proteína. Este gráfico possui regiões pré-determinadas que representam valores de (phi, psi) fisicamente factíveis. Muitos pontos fora destas regiões indicam uma estrutura de qualidade duvidosa, que apresenta problemas. Este gráfico serve, portanto, para avaliar a qualidade da estrutura de proteínas.

incorporadas à SMSLib. Estas funcionalidades devem ser incorporadas no futuro próximo. Atualmente, o cálculo de superfície acessível por solvente é realizado pelo programa SURFV (Sridharan et al., 1992), que baseia-se no algoritmo desenvolvido por Shrake & Rupley (1973). Para a implementação desta funcionalidade, pretende-se também avaliar a possibilidade de utilização de outros algoritmos. Com referência à determinação de microambientes de aminoácidos, pretende-se incorporar funcionalidades para exportação de dados em formato textual tabulado para permitir a leitura desses dados por aplicações de análise estatística e a incorporação de diferentes critérios para definição de ambiente, como, por exemplo, contatos.

Finalmente, um benefício vislumbrado a partir da definição de uma biblioteca reutilizável com as principais funcionalidades do SMS é a flexibilidade para mudança tanto de plataforma quanto de padrão de interface utilizados.

Referências Bibliográficas

- BAIROCH, A.; APWEILER, R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. **Nucleic Acids Research**, v. 28, n.1, p. 45-48, 2000.
- 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.
- FLANAGAN, D. **JavaScript: the definitive guide**. 3rd ed. Beijing: O'Reilly, 1998. 776 p.
- FRISHMAN, D.; ARGOS, P. Knowledge-based protein secondary structure assignment. **Proteins: Structure, Function and 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.
- 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.
- MDL INFORMATION SYSTEMS. **MDLChime MDLChime 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.
- RESEARCH COLLABORATORY FOR STRUCTURAL BIOINFORMATICS. **Protein Data Bank contents guide**: atomic coordinate entry format description, version 2.2: PDB guide. PDB. PDB Format Description Version 2.2: PDB guide. Disponível em: <http://www.rcsb.org/pdb/docs/format/pdbguide2.2/guide2.2_frame.html>. Acesso em: 21 dez. 2002.
- 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.
- 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.cnpia.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.
- WORLD WIDE WEB CONSORTIUM. **HTTP - hypertext transfer protocol overview**. Disponível em: <<http://www.w3.org/Protocols/>>. Acesso em: 21 dez.2002.

Comunicado Técnico, 39

**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@cnpia.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 Pedrosa Júnior

Expediente

Supervisor editorial: Ivanilde Dispatto
Normalização bibliográfica: Marcia Izabel Fugisawa Souza
Capa: Intermídia Publicações Científicas
Editoração Eletrônica: Intermídia Publicações Científicas