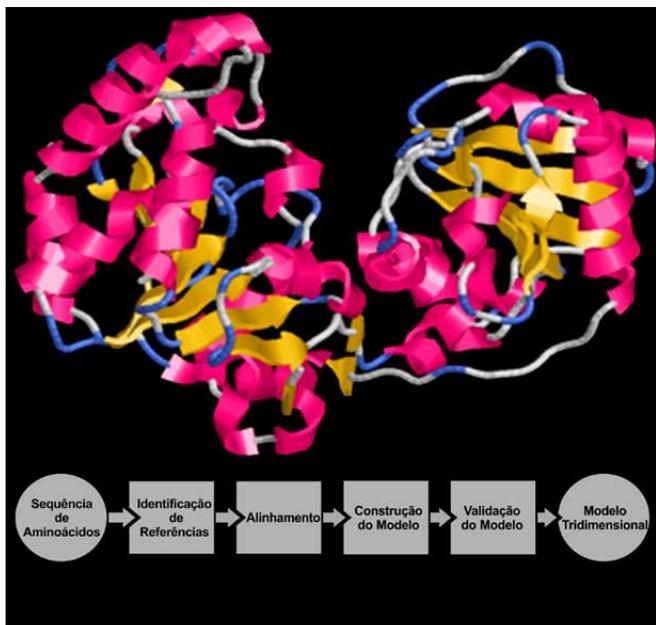


Foto: Márcia Soares Vidal



Modelagem molecular de proteínas: o caso de uma glucuronosiltransferase (*GumK*) de *Gluconacetobacter diazotrophicus* PAL5

Carlos Henrique Salvino Gadelha Meneses¹;
Luc Felicianus Marie Rouws²; Jean Luiz Simões Araújo²; Márcia Soares Vidal²;
José Ivo Baldani²

A Fixação Biológica de Nitrogênio (FBN) foi inicialmente descrita em bactérias diazotróficas da rizosfera e do rizoplane de uma grande variedade de plantas não leguminosas (DÖBEREINER, 1992), porém, evidências mostram que algumas bactérias, que colonizam o interior da planta, também são capazes de fixar nitrogênio e, por este fato, são conhecidas como bactérias endofíticas fixadoras de nitrogênio ou bactérias endofíticas diazotróficas (OLIVARES et al., 1996).

Gluconacetobacter diazotrophicus é uma bactéria diazotrófica, adaptada às condições de pressão osmótica e pH, que prevalecem no interior dos tecidos da cana-de-açúcar (GILLIS et al., 1989), já tendo sido isolada de raízes, colmos e folhas de cana-de-açúcar, capim-elefante e batata-doce (DÖBEREINER, 1992).

Os exopolissacarídeos (EPS), também conhecidos como biopolímeros ou gomas hidrossolúveis, são produzidos por uma grande variedade de microrganismos. Estes possuem um papel

importante tanto na sobrevivência da bactéria quanto na associação desta com a planta, auxiliando na adesão da célula a superfícies, na concentração de nutrientes, evitando a dessecação, o contato com o hospedeiro e protegendo contra moléculas tóxicas (SUTHERLAND, 2001).

O papel de moléculas exopolissacarídicas de alto e baixo peso molecular na associação entre bactérias fixadoras de nitrogênio e plantas tem sido muito estudado tanto em nível de caracterização química, como de expressão gênica (RAMEY et al., 2004; SERRATO et al., 2006; BALSANELLI et al., 2010; MENESES et al., 2011).

Considerando que a Rede de sequenciamento de genoma do Rio de Janeiro (RioGene) finalizou o sequenciamento do genoma da *G. diazotrophicus* (BERTALAN et al., 2009), e que uma das motivações para este sequenciamento foi o seu potencial biotecnológico, torna-se de extrema importância o estudo comparativo das estruturas terciárias de enzimas relacionadas com a síntese de

¹ Bolsista no Laboratório de Genética e Bioquímica, Embrapa Agrobiologia, Rodovia BR 465, km 7, CEP 23890-000, Seropédica, RJ, Brasil.

² Pesquisador da Embrapa Agrobiologia, Rodovia BR 465, km 7, CEP 23890-000, Seropédica, RJ, Brasil.

EPS, com diferentes especificidades pelo substrato, a fim de propor potenciais sítios para mutação nas enzimas relacionadas com a síntese de EPS de *G. diazotrophicus*, que implicariam na ampliação da utilização de substratos por estas enzimas. Informações cruciais sobre a função de proteínas podem ser obtidas a partir das suas estruturas tridimensionais (3D).

A predição da estrutura 3D de proteínas codificadas no genoma é o objetivo da genômica estrutural (BAKER e ŠALI, 2001), podendo ser realizada por métodos experimentais ou teóricos. As principais técnicas experimentais empregadas na predição estrutural de uma proteína são: a cristalografia por difração de raios-X (RX) e a espectroscopia de ressonância magnética nuclear (RMN). Mesmo sendo considerado o melhor processo para a obtenção da estrutura tridimensional de uma proteína, os métodos de predição experimental além de serem financeiramente custosos, podem levar anos e, em alguns casos, a estrutura final pode não chegar a ser obtida. Além disso, um número elevado de sequências protéicas encontram-se depositadas em bancos de dados como o "Swiss-Prot" (BAIROCH et al., 2004), contudo, poucas estruturas de proteínas foram resolvidas experimentalmente e depositadas em bancos de dados como o PDB (*Protein Data Bank*) (WESTBROOK et al., 2002). Diante disso, o emprego de métodos teóricos para a predição da estrutura tridimensional de proteínas seria uma saída para as limitações e entraves encontrados nos métodos de predição experimental.

Os principais métodos teóricos de predição de estrutura 3D de proteínas são: Modelagem Comparativa; Predição de Enovelamento de Proteínas (*Threading*) e Predição por Primeiros Princípios (*ab initio*). A modelagem comparativa, até pouco tempo conhecida como modelagem por homologia, usa como molde a estrutura 3D de outra proteína de referência, já resolvida experimentalmente e com coordenadas cartesianas depositadas em banco de dados de estruturas. Este tem sido o método mais frequentemente usado e seu limite de predição está intrinsecamente relacionado com o grau de similaridade entre as estruturas alvo e molde. A Predição de Enovelamento de Proteínas tenta ajustar a estrutura da proteína de interesse aos tipos de enovelamentos de proteínas conhecidas, ou seja, registradas e depositadas em bibliotecas de enovelamentos. Já

a Predição por Primeiros Princípios emprega as propriedades físico-químicas conhecidas de cada aminoácido para a construção de funções de energia, sendo que tais funções são minimizadas por algoritmos que realizam buscas no espaço de conformações que a proteína de interesse possa assumir (ŠALI e BLUNDELL, 1993; PETSKO e RINGE, 2004).

Diante do exposto acima, o presente trabalho teve como objetivo descrever simplificada e metodologicamente a modelagem comparativa empregando como modelo a proteína β -1,2-glucoronosiltransferase (*GumK*) de *Gluconacetobacter diazotrophicus*, que está envolvida na biossíntese de exopolissacarídeos. *GumK*, é uma proteína de 41 kDa, pertencente a família das glicosiltransferases 70, enzima responsável pela adição de carboidratos ativados (ácido UDP-glicurônico) a um resíduo de difosfoliprenol manosil (α -1,3) para produzir celobiose glucuronil-(β -1,2 difosfoliprenol)-manosil (α -1,3)-celobiose.

A metodologia de modelagem comparativa envolve quatro etapas sucessivas: identificação e seleção da proteína-molde; alinhamento das sequências alvo e molde; construção e otimização do modelo e, validação do modelo.

A sequência de aminoácidos da proteína *GumK* de *G. diazotrophicus* foi submetida a uma análise comparativa através do programa BLAST (Basic Local Alignment Search Tool) (www.ncbi.nih.gov/BLAST) (ALTSCHUL et al., 1990) na base de dados PDB (alinhamento local). As sequências de aminoácidos provenientes de alguns dos microrganismos encontrados no resultado da busca foram submetidas ao programa de alinhamento múltiplo de sequências, CLUSTALX (JEANMOUGIN et al., 1998). Nesta primeira etapa foi identificada uma proteína relacionada à sequência de aminoácidos de *GumK* de *G. diazotrophicus* (proteína alvo), tendo sido levados em conta durante o processo de escolha da proteína molde aspectos como: conhecimento estrutural, similaridade de sequência, similaridade de função, expressão pelo mesmo grupo de genes, e correlação evolutiva entre as proteínas.

Identificada a proteína molde, métodos de alinhamento global entre as sequências foram aplicados, o que consiste na segunda etapa da modelagem comparativa. O alinhamento entre as sequências foi realizado com o emprego do

2HY7. O arquivo gerado pelo programa consiste em um modelo tridimensional para a sequência alvo que contém todos os átomos da cadeia principal e da cadeia lateral da proteína. No alinhamento das sequências da proteína 2HY7 (Molde) e *GumK* de *G. diazotrophicus* (Alvo) mostrado na Fig. 1, é possível verificar quais aminoácidos são idênticos e quais são conservados. Estas características encontradas no alinhamento indicam que a estrutura cristalográfica de 2HY7 pode ser usada como estrutura molde.

Desta modelagem inicial foram obtidos três modelos, que em seguida foram avaliados de acordo com a função objetiva e com os parâmetros stereoquímicos. A Fig. 2 mostra o modelo preliminar obtido, onde o critério inicial para seleção do melhor modelo foi o de menor energia fornecido pelo próprio programa. Após essa seleção um modelo mais representativo foi selecionado levando-se em consideração a porcentagem de identidade e o tamanho do segmento da proteína modelada. Em seguida foram geradas as coordenadas da estrutura através do programa MODELLER, e estas foram visualizadas com o programa de visualização de estruturas RASMOL.

Os resultados mostram que a proteína possui dois domínios bem definidos, denominados domínios Rossmann, com uma fenda catalítica entre eles (Fig. 2), o que é uma característica típica da superfamília das glicosiltransferase do tipo B.

A Fig. 3 mostra a sobreposição do modelo gerado com a estrutura da molécula molde (Alinhamento estrutural). Nota-se que, além da alta porcentagem de identidade de sequência, a estrutura terciária é bastante conservada devido ao tipo de alinhamento global/estrutural que a sequência de aminoácidos de *GumK* de *G. diazotrophicus* foi submetida.

O gráfico de Ramachandran pode ser observado na Fig. 4. A estrutura deve apresentar acima dos 90% de seus resíduos nas regiões favoráveis (Região azul), ou seja, nas regiões a, b e l. Devem-se desconsiderar os resíduos de glicina (apresentados na figura como triângulos), já que eles não possuem cadeia lateral, os de prolina, pois nestes o carbono alfa está ligado à cadeia lateral e os resíduos das extremidades (C-terminal e N-terminal), que apresentam padrões stereoquímicos diferentes dos outros resíduos.

A porcentagem de resíduos encontrados nas regiões favoráveis foi de 95% (Em azul escuro),

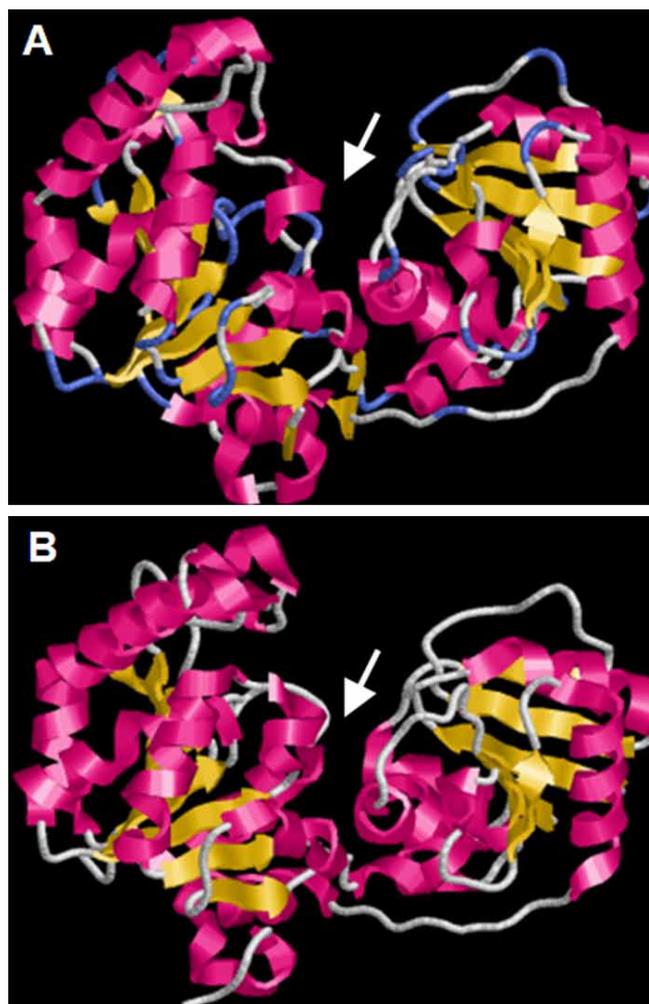


Fig. 2. Visualização do modelo da estrutura de tridimensional da proteína *GumK* de *G. diazotrophicus* pelo programa RASMOL. **(A)** Modelo obtido para *GumK* de *G. diazotrophicus*. **(B)** Estrutura para *GumK* de *X. campestris*, utilizada como molde (BARRERAS et al., 2008). As setas em cada uma das figuras indicam suas respectivas fendas catalíticas.

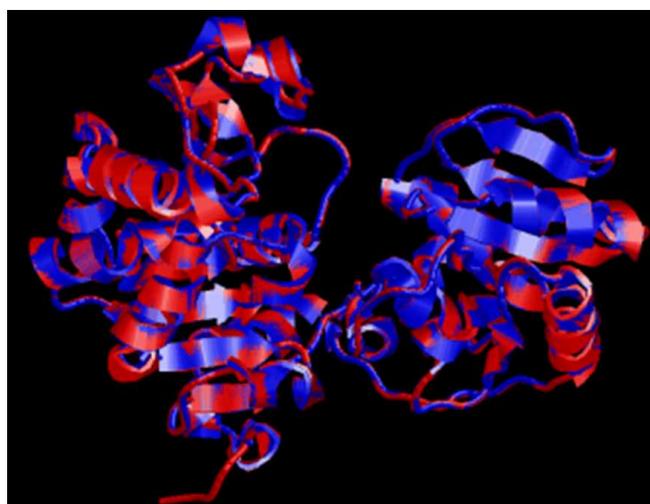


Fig. 3. Visualização do alinhamento estrutural da proteína *GumK* de *G. diazotrophicus* (em azul) e *GumK* de *X. campestris* (em vermelho) pelo programa RASMOL.

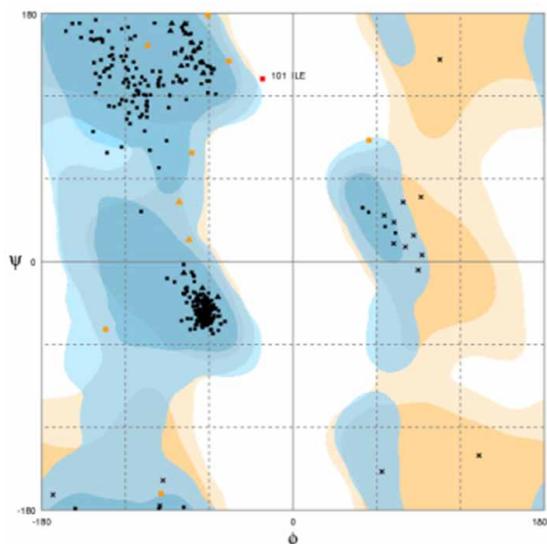


Fig. 4. Gráfico de Ramachandran do modelo final obtido para *GumK* de *G. diazotrophicus*, gerado pelo programa Procheck. Os ângulos Phi e Psi estão em graus. Em azul escuro estão as regiões favoráveis; em azul claro as regiões adicionalmente permitidas; em amarelo claro as regiões generosamente permitidas e em branco as regiões desfavoráveis.

nas regiões adicionalmente permitidas (Regiões a, b, l, p, ou seja, azul claro) de 3,5%, nas regiões generosamente permitidas (Regiões ~a, ~b, ~l, ~p, ou seja, amarelo claro) foi de 1,5% e nas regiões desfavoráveis (Região em branco) foi de 0,0%. Os resíduos que se encontram em regiões desfavoráveis são um reflexo das estruturas usadas como molde, e alguns se encontram nas extremidades da proteína. Sendo assim, os dados do gráfico de Ramachandran suportam o modelo obtido, em outras palavras a última etapa do processo de modelagem comparativa, validação do modelo, foi finalizada com sucesso.

Os dados biológicos e estruturais relatados acima abrem as portas para estudos de base molecular entre esta proteína e seu respectivo ligante, podendo assim mostrar uma possível seletividade nesta família de glicosiltransferases. Este é o primeiro modelo construído para representar a estrutura tridimensional para *GumK* de *G. diazotrophicus*, uma das enzimas envolvidas na síntese de EPS em *G. diazotrophicus*.

Referências Bibliográficas

ALTSCHUL, S. F.; GISH, W.; MILLER, W.; MEYERS, E. W.; LIPMAN, D. J. Basic local alignment search tool. *Journal of Molecular Biology*, v. 215, p. 403-410, 1990.

ARNOLD, K.; BORDOLI, L.; KOPP, J.; SCHWEDE, T. The swiss-model workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*, v. 22, n. 2, p. 195-201, 2006.

BAIROCH, A.; BOECKMANN, B.; FERRO, S.; GASTEIGER, E. Swiss-prot: juggling between evolution and stability. *Briefings in Bioinformatics*, v. 5, n. 1, p. 390-455, 2004.

BAKER, D.; ŠALI, A. Protein structure prediction and structural genomics. *Nature*, v. 294, p.93-96, 2001.

BARRERAS, M.; SALINAS, S. R.; ABDIAN, P. L.; KAMPEL, M. A.; IELPI, L. Structure and mechanism of *GumK*, a membrane-associated glucuronosyltransferase. *Journal of Biological Chemistry*, v. 283, p. 25027-25035, 2008.

BALSANELLI, E.; SERRATO, R. V.; BAURA, V. A.; SASSAKI, G.; YATES, M. G.; RIGO, L. U.; PEDROSA, F. O.; SOUZA, E. M. de; MONTEIRO, R. A. *Herbaspirillum seropedicae* rfbB and rfbC genes are required for maize colonization. *Environmental Microbiology*, v. 12, n. 8, p. 2233-2244, 2010.

BERTALAN, M.; ALBANO, R.; PÁDUA, V.; ROUWS, L.; ROJAS, C.; HEMERLY, A.; TEIXEIRA, K.; SCHWAB, S.; ARAUJO, J.; OLIVEIRA, A.; FRANÇA, L.; MAGALHÃES, V.; ALQUÉRES, S.; CARDOSO, A.; ALMEIDA, W.; LOUREIRO, M.; NOGUEIRA, E.; CIDADE, D.; OLIVERIA, D.; SIMÃO, T.; MACEDO, J.; VALADÃO, A.; DRESCHSEL, M.; FREITAS, F.; VIDAL, M.; GUEDES, H.; RODRIGUES, E.; MENESES, C.; BRIOSO, P.; POZZER, L.; FIGUEIREDO, D.; MONTANO, H.; CUNHA JÚNIOR, J.; SOUZA FILHO, G. de; FLORES, V.; FERREIRA, B.; BRANCO, A.; GONZALEZ, P.; GUILLOBEL, H.; LEMOS, M.; SEIBEL, L.; MACEDO, J.; FERREIRA, A.; MARTINS, G.; COELHO, A.; SANTOS, E.; AMARAL, G.; NEVES, A.; PACHECO, B.; CARVALHO, D.; LERY, L.; BISCH, P.; ROSSLE, C.; URMÉNYI, T.; PEREIRA, R.; SILVA, R.; RONDINELLI, E.; KRUGER, W.; MARTINS, O.; BALDANI, J.; FERREIRA, P. Complete genome sequence of the sugarcane nitrogen-fixing endophyte *Gluconacetobacter diazotrophicus* Pal5. *BMC Genomics*, London, v. 10, n. 450, p. 1-17, 23 set. 2009.

DÖBEREINER, J. Recent changes in concepts of plant bacteria interaction: endophytic N_2 fixing bacteria. *Ciência e Cultura*, v. 44, p. 310-313, 1992.

ESWAR, N.; WEBB, B.; MARTI-RENOM, M. A.; MADHUSUDHAN, M.; ERAMIAN, D.; SHEN, M.-Y.; PIEPER, U.; SALI, A. **Comparative Protein Structure Modeling Using MODELLER**. In: Current Protocols in Protein Science. London: John Wiley & Sons, 2007. v. 50, p. 1-31.

GILLIS, M.; KERTERS, B.; HOSTE, D. J.; KROOPENSTEDT, R. M.; STEPHAN, M. P.; TEIXEIRA, K. R. S.; DÖBEREINER, J.; LEY, J. *Acetobacter diazotrophicus* sp. nv. a nitrogen fixing acetic acid bacterium associated with sugar cane. **International Journal of Systematic Bacteriology**, v. 39, p. 361-364, 1989.

GUERX, N.; PEITSCH, M. C. Swiss-model and swiss-pdb viewer: an environment for comparative protein modeling. **Electrophoresis**, v. 18, p. 2714, 1997.

JEANMOUGIN, F.; THOMPSON, J. D.; GOUY, M.; HIGGINS, D. G.; GIBSON, T. J. Multiple sequence alignment with Clustal X. **Trends in Biochemical Sciences**, v. 23, p. 403-405, 1998.

LASKOWSKI, R. A.; MACARTHUR, M. W.; MOSS D. S.; THORNTON, J. M. PROCHECK: A program to check the stereochemical quality of protein structures. **Journal of Applied Crystallography**, v.26, p.283-291, 1998.

MENESES, C. H. S. G.; ROUWS, L. F. M.; SIMÕES-ARAÚJO, J. L.; VIDAL, M. S.; BALDANI, J. I. Exopolysaccharide production is required for biofilm formation and plant colonization by the nitrogen-fixing endophyte *Gluconacetobacter diazotrophicus*. **Molecular Plant-Microbe Interactions**, v. 24, n.12, p.1448-1458, 2011.

OLIVARES, F. L.; BALDANI, V. L. D.; REIS, V. M.; BALDANI, J. I.; DÖBEREINER, J. Occurrence of the endophytic diazotroph *Herbaspirillum* spp. in roots,

stems and leaves predominantly of Gramineae. **Biology and Fertility of Soils**, v. 2, p.197-200, 1996.

PETSKO, G. A.; RINGE, D. **Protein Structure and Function**. Sunderland, MA: Sinauer Association, 62004, 195p.

RAMACHANDRAN, G. N.; RAMAKRISHNAN, C.; SASISEKHARAN, V. Stereochemistry of polypeptide chain conformations. **Journal of Molecular Biology**, v.7, p.95-99, 1963.

RAMEY, B. E.; KOUTSOUDIS, M.; VON BODMAN, S. B.; FUQUA, C. Biofilm formation in plant-microbe associations. **Current Opinion in Microbiology**, v. 7, p. 602-609, 2004.

ŠALI, A.; BLUNDELL, T. L. Comparative protein modeling by satisfaction of spatial restraints. **Journal of Molecular Biology**, v. 234, p. 779-815, 1993.

SAYLE, R. Biomolecular graphics for all. **Trends in Biochemical Sciences**. v. 20, n. 9, p.374-3796, 1995.

SERRATO, R. V.; SASSAKI, G. L.; CRUZ, L. M.; PEDROSA, F. O.; GORIN, P. A. J.; IACOMINI, M. Culture conditions for the production of an acidic exopolysaccharide by the nitrogen-fixing bacterium *Burkholderia tropica*. **Canadian Journal of Microbiology**, v. 52, p. 489-493, 2006.

SUTHERLAND, I. W. Biofilm exopolysaccharides: a strong and sticky framework. **Microbiology**, v. 147, p. 3-9, 2001.

WESTBROOK, J.; FENG, Z.; JAIN, S.; BHAT, T. N.; THANKI, N., RAVICHANDRAN, V., GILLILAND, G. L.; BLUHM, W. F.; WEISSIG, H.; GREER, D. S.; BOURNE, P. E.; BERMAN, H. M. The protein data bank: unifying the archive. **Nucleic Acids Research**, v. 30, n. 1, p. 245-248, 2002.

Comunicado Técnico, 141

Exemplares desta edição podem ser adquiridos na:
Embrapa Agrobiologia
 Endereço: BR465, km7 - Caixa Postal 74505
 CEP 23890-000 - Seropédica/RJ, Brasil
 Fone: (21) 3441-1500
 Fax: (21) 2682-1230
 Home page: www.cnpab.embrapa.br
 E-mail: sac@cnpab.embrapa.br
 1ª edição

1ª impressão (2011): 50 exemplares

Ministério da
 Agricultura, Pecuária
 e Abastecimento



Comitê de Publicações

Presidente: Norma Gouvêa Rumjanek
Secretária-Executiva: Carmelita do Espírito Santo
Membros: Bruno José Rodrigues Alves,
 Ednaldo da Silva Araújo, Guilherme Montandon Chaer,
 José Ivo Baldani, Luís Henrique de Barros Soares.

Expediente

Supervisão editorial: Norma Gouvêa Rumjanek
Normalização bibliográfica: Carmelita do Espírito Santo
Tratamento das ilustrações: Maria Christine Saraiva Barbosa
Editoração eletrônica: Marta Maria Gonçalves Bahia