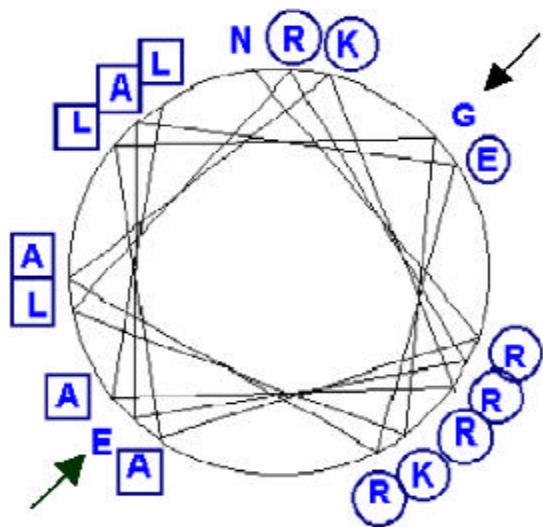




Análise Computacional para Estudo de Genômica Funcional

I - Identificação do Domínio de Ligação de Calmodulina em Proteínas HSP70 e BiP

José Edson Fontes Figueiredo¹
Wanderson Duarte da Rocha²
Vinício Tadeu da Silva Coelho²

O conhecimento nas áreas da Biologia Molecular e da Computação cresceram de forma vertiginosa nos últimos 20 anos. O mapeamento e o seqüenciamento de genomas de vários organismos geram diariamente um volume elevado de informações. A impossibilidade de trabalhar manualmente esses dados criou uma demanda em informática que desenvolveu programas especiais para armazenamento de seqüências gênicas em bancos de dados (Sander, 2002; Smith, 1990; Waterman, 1990). Programas para identificação de “motifs” estruturais e funcionais de proteínas a partir de seqüências primárias de aminoácidos como EMOTIF, EMATRIX E 3DMOTIF, LOCK, 3DSEARCH, desenvolvidos pelo grupo de bioinformática da *Stanford University*, CA, USA, ajudam a elucidar funções de proteínas desconhecidas, geradas a partir de seqüências de DNA dos projetos genoma, e auxiliam o entendimento

sobre questões aparentemente ambíguas, como a ocorrência de apenas uma cópia de determinado gene vital em alguns organismos e múltiplas cópias do mesmo gene em outros. Este estudo foi realizado com o objetivo de demonstrar alguns aspectos da importância da análise computacional de seqüências de DNA, utilizando-se como exemplo a proteína BiP1 da soja, para esclarecimento da funcionalidade de genes.

A proteína BiP está presente em todos os organismos eucariotos e desempenha funções celulares vitais, como mediadora no dobramento e montagem de proteínas multiméricas no interior do retículo endoplasmático (RE). Na ausência de BiP, muitas proteínas da via secretora não assumem a conformação nativa e precipitam no RE. O gene BiP ocorre como cópia única em diferentes organismos, como as leveduras, humanos e várias espécies de plantas. Em

¹Biólogo, Doutor, Embrapa Milho e Sorgo. Caixa Postal 151, CEP 35701-970 Sete Lagoas, MG

²Eng. Agr., Doutor, Departamento de Bioquímica e Imunologia, UFMG, CEP 30000-000, Belo Horizonte, MG

³Estudante do Curso de Ciências Biológicas/Centro Universitário Metodista Izabela Hendrix, Caixa Postal 46, CEP 30.160-012 Belo Horizonte, MG

humanos, foram identificadas duas cópias do gene BiP, mas uma delas era um pseudogene (Ting & Lee, 1988) e, em leveduras, foram identificadas várias mutações letais do gene BiP (Normington, *et al.*, 1989). Em algumas espécies de plantas, foram descritas mais de uma cópia do gene BiP por genoma haplóide. Em soja (*Glycine max*), foram identificadas três cópias de BiP em biblioteca de cDNA de folhas (Kalinski *et al.*, 1995) e uma cópia em biblioteca de cDNA de sementes (Figueiredo, *et al.*, 1997). Em tabaco (Denecke *et al.*, 1991) e milho (Boston, comunicação pessoal), BiP foi verificado como sendo uma família multigênica, enquanto em espinafre existia apenas uma cópia por genoma haplóide (Anderson *et al.*, 1994). O significado exato de algumas espécies de plantas possuírem várias cópias de genes BiP, enquanto outras apresentam apenas uma cópia, não é compreendido. Tem sido especulado que os diferentes membros da família multigênica codificam isoformas contendo pequenas diferenças de especificidade com o substrato e, por isso, desempenhariam funções diferenciadas (Denecke *et al.*, 1991). Outra hipótese é a de que diferentes cópias de BiP apresentam expressão especializada para responder aos diferentes estímulos ambientais (Carolino *et al.*, 2003). Por outro lado, o número de cópias de BiP em tabaco e soja poderia ser parcialmente explicado pelo caráter autotetraplóide dessas espécies.

Também em milho, uma espécie considerada por muitos anos como diplóide, a existência de várias cópias de BiP pode ser decorrente de poliploidia, pois foram encontradas muitas evidências recentes sobre sua origem alotetraplóide (Gaut & Doebley, 1997).

Contudo, em espinafre, uma espécie triplóide, a existência de apenas uma cópia de BiP por genoma (Anderson *et al.*, 1994) indica que o número de cópias de BiP em uma espécie não se deve apenas a ocorrência de poliploidia. Além disso, no caso da soja, uma espécie autotetraplóide e de autofecundação, o isolamento de três isoformas de BiP em uma

única variedade mostra claramente a existência de mais de dois *loci* para os genes BiP no genoma dessa espécie. Portanto, se a existência de várias cópias de BiP em algumas espécies de plantas está relacionada com função diferencial ou consiste em redundância, é uma questão que ainda não foi resolvida. Análise computacional das seqüências primárias de proteínas deduzidas de seqüências de DNA depositadas no GenBank podem gerar informações úteis que auxiliem a esclarecer esse aspecto aparentemente ambíguo do genoma.

Programas computacionais foram previamente utilizados para identificar “motifs” estruturais e funcionais na proteína deduzida da seqüência de DNA de BiP da soja (Figueiredo *et al.*, 1997). Vários domínios foram identificados, entre eles, o possível domínio de ligação à calmodulina, que é o objeto deste estudo.

Diferentes proteínas celulares como Ca-Atpase, NAD cinase, fosfolipases, fosfatases e uma variedade de proteínas cinases, cujas atividades são reguladas por calmodulina, possuem seqüências específicas que definem o domínio ao qual a calmodulina liga e exerce atividade regulatória (Lu & Harrington, 1994; Roberts *et al.*, 1985). Esses domínios são definidos por uma série de características estruturais comuns a todas essas proteínas, como a justaposição de um conjunto de resíduos de aminoácidos básicos e hidrofóbicos, que, ao serem projetados em um anel de eixo axial, ocupam os lados opostos de uma hélice anfifílica. Proteínas da família “Heat Shock Protein” (HSP) como BiP, são reguladas por cálcio, mas os mecanismos bioquímicos envolvidos nesse controle ainda não foram totalmente esclarecidos.

A fim de verificar se o domínio de ligação à calmodulina da proteína BiP1 e outras regiões da molécula poderiam apresentar características de alfa-hélice anfifílica, procedeu-se à predição da estrutura secundária da proteína utilizando os métodos SOPMA (“Self Optimized Prediction Method

from Alignments”) (Geourjon & Deleage, 1994, 1995) e o método de predição de alfa-hélice de Chou & Fasman (1978). Várias regiões de alfa-hélice foram identificadas na proteína BiP1 (dados não mostrados). Essas alfa-hélices foram posteriormente analisadas, para verificar a existência de regiões contendo resíduos hidrofílicos e hidrofóbicos adjacentes e, portanto, potencialmente capazes de formar hélices anfifílicas que constituiriam prováveis domínios de ligação à calmodulina na proteína BiP1.

Apenas a sequência NRALGKLRREAERAKRALSSQ, na proteína BiP1, apresentou características compatíveis com a formação de alfa-hélice anfifílica e, portanto, é uma sequência candidata para domínio de ligação à calmodulina. A comparação do peptídeo com outras sequências de proteínas HSP e BiP, tanto de animais quanto de plantas, revelou algumas características que são peculiares a cada grupo de organismo (Figura 1). Dos 21 resíduos de aminoácidos da proteína BiP1 da soja, que foram identificados como sendo o sítio provável de calmodulina, 12 (A-RRL—ERAKR-LSS) são conservados em todas as proteínas HSP de plantas e de animais. A análise dos domínios de ligação à calmodulina das proteínas HSP70 e BiP, nos diferentes organismos, possibilitou estabelecer dois consensos para ligação de calmodulina em BiP: um para plantas (NRALGKLRREAERAKRALSSQ) e outro para animais (NRAVQKLRREVEKAKRALSSQ). É interessante destacar que, em praticamente todas as proteínas BiP e HSP, pode ser observada a presença do resíduo N na exata vizinhança amino-terminal do domínio de ligação à calmodulina, sugerindo significado biológico para esse resíduo. As exceções foram observadas em HSP70 de *Daucus carota* (Figura 1) e na proteína DnaK de *E. coli* (dado não mostrado). Estudos de mutagênese sítio dirigida juntamente com ensaios de ligação à calmodulina poderiam esclarecer a funcionalidade ou não desse resíduo em

proteínas da família HSP70. Ao mesmo tempo, a alta conservação de sequência de aminoácidos, no domínio de ligação à calmodulina nas proteínas HSP70, em organismos evolutivamente divergentes, sugere que essa região da proteína exerce função essencial, estrutural ou regulatória.

A projeção da sequência NRALGKLRREAERAKRALSSQ em anel de eixo axial (Figura 2) mostrou que o domínio de ligação à calmodulina da proteína BiP1 possui dois pólos opostos, um hidrofóbico, com sete resíduos interrompidos por L (hidrofílico) e outro pólo hidrofílico, com oito resíduos interrompidos por G (hidrofóbico) na posição 266 na proteína madura. É interessante destacar que o resíduo triptofano (W), identificado em proteínas que não pertencem à família HSP70 como sendo importante para estabilizar a ligação da calmodulina com a proteína alvo (O’Neil & DeGrado, 1990), está ausente na sequência do sítio de calmodulina das proteínas HSP70 e de BiP1. Contudo, a ausência de triptofano não seria suficiente para impedir a ligação de calmodulina às proteínas HSP70, conforme ficou demonstrado por meio de experimentos *in vitro* usando peptídeos sintéticos, que não possuíam o aminoácido triptofano em suas sequências, e, mesmo assim, foram capazes de se ligar à calmodulina (Stevenson & Calderwood, 1990). Nesse caso, é possível que o caráter anfipático da alfa-hélice e/ou as relações do domínio de calmodulina com outras regiões da própria proteína sejam mais importantes para estabilizar as interações entre a calmodulina e a proteína alvo. Contudo, a presença de um resíduo hidrofílico (E) em BiP1, no polo hidrofóbico da alfa-hélice (Figura 2) é suficiente para impedir que o domínio de calmodulina assumo o caráter anfipático, impedindo, portanto, que ocorra a associação com a proteína calmodulina (Hanley *et al.*, 1988). Essa situação observada no domínio de ligação à calmodulina na proteína BiP sugere que esse domínio não é funcional.

BiP (Plantas)

cUFVBiP1 (*Glycine max*)
Nicotiana tabacum
Lycopersicon esculentum
Arabidopsis thaliana
Oryza sativa
Spinacia oleracea
Consenso

NRALGKLRREAERAKRALSSQ (262-282)
NRALGKLRREAERAKRALSSQ (262-282)
NRALGKLRREAERAKRSLSSQ (262-282)
NKALGKLRRECERAKRALSSQ (262-282)
NRALGKLRREAERAKRALSNQ (262-282)
NRALGKLRRECERAKRALSSQ (262-282)
NRALGKLRREAERAKRALSSQ

BiP (Animais)

Homo sapiens
Mus musculus
Rattus norvegicus
Xenopus laevis
Eimeria tenella
Schistosoma japonicum
Consenso

NRAVQKLRREVEKAK-ALSSQ (262-281)
NRAVQKLRREVEKAKRALSSQ (262-282)
NRAVQKLRREVEKAKRALSSQ (262-282)
KRAVQKLRREVEKAKRALSAQ (261-281)
KSALQRLRREVERAKRALSSS (indeter)
KRAVQKLRREVEKAKRMLSTT (261-281)
NRAVQKLRREVEKAKRALSSQ

HSP70 (Plantas)

Glycine max
Zea mays
Petunia hybrida
Pisum sativum
Triticum aestivum
Daucus carota
Consenso

NARALRRLRTACERAKRTLSS (261-282)
NPRALRRLRTACERAKRTLSS (261-282)
NPRALRRLRTACERAKRTLSS (261-282)
NARALRRLRTACERAKRTLSS (261-282)
NPRALRRLRTACERAKRTLSS (260-281)
ECEALRRLRTACERAKRTLSSS (259-280)
NPRALRRLRTACERAKRTLSS

HSP70 (Animais)

Homo sapiens
Mus musculus
Tilapia mossambica
Drosophila melanogaster
Biomphalaria glabrata
Eunicella cavolini
Asbestopluma hypogea
Consenso

NKRAVRRLRTACERAKRTLSSS (256-277)
NKRAVRRLRTACERAKRTLSSS (256-277)
NKRAVRRLRTAFERAKRTLSSS (256-277)
NPRALRRLRTAAERAKRTLSSS (254-275)
NPRALRRLRTACERAKRTLSSS (259-280)
NKRAVRRLRTACERAKRTLSSA (indeter)
NKRAVRRLRTACERAKRTLSS (indeter)
NKRAVRRLRTACERAKRTLSSS

L

Figura 1. Domínio de ligação à calmodulina em proteínas BiP e HSP70. O domínio de ligação à calmodulina de BiP e HSP70 de plantas e animais foi separado em grupos diferentes após o alinhamento das seqüências (Clustal W). As cores destacam resíduos idênticos em todos os grupos (negro); resíduos idênticos em cada grupo (vermelho); resíduos mais frequentes do grupo (amarelo). Verde e azul são resíduos presentes em todas as proteínas BiP de mamíferos e demais animais, respectivamente, e violeta, resíduo próximo à seqüência definida como domínio de ligação à calmodulina. Os números indicam a posição do domínio na proteína madura, exceto para as seqüências incompletas, sem a região amino-terminal (indeter). Na última linha de cada grupo, foi ressaltada a seqüência consenso do domínio de ligação à calmodulina e os resíduos conservados foram grifados.

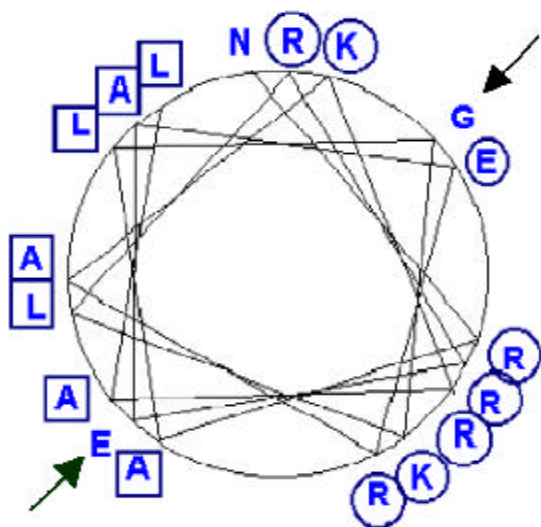


Figura 2. Projeção em anel de eixo axial do domínio de ligação à calmodulina da proteína BiP1. A projeção inicia com N (posição 262) e termina em L (posição 279). Os grupos de resíduos hidrofóbicos e hidrofílicos nos pólos da projeção em anel foram inscritos nos quadrados e nos círculos, respectivamente. Os resíduos SSQ que fazem parte do domínio de ligação à calmodulina não foram projetados, pois não definem a estrutura da hélice (Stevenson & Calderwood, 1990). As setas indicam os resíduos hidrofílico (L) e hidrofóbico (G) que interrompem o caráter anfipático da hélice. A projeção foi feita utilizando o programa SOPM (Geourjon & Deleage, 1994).

O cálcio é relevante para modular a atividade de proteínas da classe HSP70. Em situações de estresse, como choque térmico, por exemplo, ocorre um aumento significativo na concentração de Ca^{2+} intracelular. Acredita-se que o aumento da concentração de Ca^{2+} promova a ligação de calmodulina à HSP70 (Stevenson & Calderwood, 1990). A dissociação do complexo calmodulina/HSP70 ocorreria após o restabelecimento dos níveis normais de cálcio, possibilitando à proteína se ligar com as proteínas desnaturadas ou recém-sintetizadas. A ligação reversível de calmodulina com HSP70 impõe restrições estruturais, espaciais e temporais à HSP70, em função dos níveis de cálcio intracelular que são importantes para a regulação das interações de HSP70 com outras proteínas.

Embora a ocorrência de calmodulina no citoplasma de células animais e vegetais tenha sido demonstrada de maneira inequívoca, até o momento não existe nenhuma evidência da presença dessa proteína no retículo

endoplasmático (Espindola, comunicação pessoal; Ling & Assman, 1992; Nelson, comunicação pessoal). A presença de um domínio conservado de ligação à calmodulina em BiP de plantas e animais pode ser um indicativo de co-localização de calmodulina no RE. É possível, também, que o domínio de calmodulina em proteínas BiP, embora tenha sido conservado durante o processo evolutivo, não seja funcional. De fato, tem sido demonstrado que a presença de um resíduo hidrofílico, no pólo hidrofóbico da alfa-hélice, é suficiente para impedir que o peptídeo assumo o caráter anfipático e que se ligue à calmodulina (Hanley *et al.*, 1988).

Considerando que essa condição é observada em BiP de plantas e animais (resíduo E), provavelmente esse domínio não seja funcional. Ensaios *in vitro* de ligação de calmodulina com peptídeos sintéticos poderão esclarecer sobre a funcionalidade ou não dos domínios de ligação de calmodulina identificados em BiP. Contudo, não pode ser descartada a hipótese sobre a existência de uma proteína endoplasmática, análoga à calmodulina, capaz de se ligar ao domínio modificado de calmodulina e promover a ligação de cálcio em proteínas BiP.

A análise de domínios da proteína BiP1 mostrou a ocorrência de mutação no sítio de ligação de ATP que impede a ligação desse nucleotídeo (dados não apresentados). A hidrólise de ATP é essencial para que as proteínas HSP exerçam suas funções de chaperone molecular. Esse dado reforça a idéia de que o gene BiP seqüenciado constitui um pseudogene. Entretanto, esse pseudogene de BiP da soja está sendo transcrito significando que a região reguladora (promotor) permanece conservada e, portanto, o processo de transcrição é ativado. Contudo, as mutações na seqüência traduzida do gene BiP1 (domínio de ligação de calmodulina e sítio de ligação de ATP) constituem fortes evidências sobre a perda da função da proteína BiP1 sintetizada. Considerando que, em soja, ocorrem várias cópias do gene BiP, a transcrição e a tradução de uma cópia normal

seriam suficientes para suprir as células com as funções desempenhadas pela proteína.

Referências

- ANDERSON, J.V.; LI, Q.B.; HASKELL, D.W.; GUY, C.L. Structural organization of the spinach endoplasmic reticulum-luminal 70 kilodalton heat-shock cognate gene and expression of 70-kilodalton heat-shock genes during cold acclimation. **Plant Physiology**, Bethesda, v.104, p.1359 – 1370, 1994.
- CAROLINO, S.M.B.; VAEZ, J.R.; IRSIGLER, A.S.T.; VALENTE, M.A.S.; RODRIGUES, L.A.Z.; FONTES, E.P.B. Plant BiP gene family: differential expression, stress induction and protective role against physiological stresses, review article. **Brazilian Journal of Plant Physiology**, Campinas, v.15, n.2, p. 59 – 66, 2003.
- CHOU, P.Y.; FASMAN, G.D. amino acid scale: Conformational parameter for alpha helix. **Advances in Enzymology**, New York, v.47, p. 45 –148, 1978.
- DENECKE, J.; GOLDMAN, M.H.S.; DEMOLDER, J.; SEURINCK, J.; BOTTERMAN, J. The tobacco luminal binding protein is encoded by a multigene family. **Plant Cell**, Rockville, v. 3, p.1025 –1035, 1991.
- FIGUEIREDO, J.E.F.; CASCARDO, J.C.M.; CAROLINO, S.M.B.; ALVIN, F.; FONTES, E.P.B. Water-stress regulation and molecular analysis of the soybean BiP gene family. **Brazilian Journal of Plant Physiology**, Campinas, v.9, n. 2, p.103 –110, 1997.
- GAUT, B. S.; DOEBLEY, J. F. dna sequence evidence for the segmental allotetraploid origin of maize. **Proceedings of the National Academy of Sciences**, Washington, v. 94, p. 6809 – 6814, 1997.
- GEOURJON, C.; DELEAGE, G. SOPM: a self optimized prediction method for protein secondary structure prediction. **Protein Engineering**, Washington, v.7, p.157-164, 1994.
- GEOURJON, C.; DELEAGE, G. SOPMA: Significant improvements in protein secondary structure prediction by prediction from multiple alignments. **Computer Applications in the Bioscience**, Oxford, v.11, p. 681-684, 1995.
- HANLEY, R.M.; MEANS, A.R.; KEMP, B.E.; SHENOLIKAR, S. Mapping of calmodulin-binding domain of Ca^{2+} /calmodulin dependent protein kinase II from rat brain. **Biochemical and Biophysical Research Communications**, New York, v.152, p.122 –128, 1988.
- KALINSKI, A.; ROWLEY, D.L.; LOER, D.S.; FOLEY, C.; BUTA, G.; HERMAN, E.M. Binding-Protein expression is subject to temporal, developmental and stress-induced regulation in terminally differentiated soybean organs. **Planta**, Berlin, v.195, p.611- 621, 1995.
- LING, V.; ASSMANN, S.M. Cellular distribution of calmodulin and calmodulin-binding proteins in *Vicia faba* L. **Plant Physiology**, Bethesda, v.100, p970-978, 1992.
- LU, Y.; HARRINGTON, H.M. Isolation of tobacco cDNA clones encoding calmodulin-binding proteins and characterization of a known calmodulin-binding domain. **Plant Physiology and Biochemistry**, Paris, v.32, p.413 – 422, 1994.
- NORMINGTON, K.; KOHNO, K.; KOZUTSUMI, Y.; GETHING, M.J.; SAMBROOK, J. *S. cerevisiae* encodes an essential protein homologous in sequence and function to mammalian BiP. **Cell**, Cambridge, v.57, p.1223 –1236, 1989.
- O'NEIL, K.T.; DEGRADO, W.F. How calmodulin binds its targets: sequence independent recognition of amphiphilic α -helices. **Trends Biochemical Sciences**, Amsterdam, v.15, p.59 – 64, 1990.
- ROBERTS, D.M.; LUKAS, T.J.; WATTERSON, D.M. Structure, function, and mechanism of action of calmodulin. **CRC - Critical Reviews in Plant Science**, Boca Raton, v.4, p.311 – 339, 1985.
- SANDER, C. The journal Bioinformatics, key medium for computational biology. **Bioinformatics**, Oxford, v.18, n.1, p.1 – 2, 2002.
- SMITH, T.F. The history of the genetic sequence databases. **Genomics**, San Diego, v. 6, n. 4, p.701 – 707, 1990.
- STEVENSON, M. A.; CALDERWOOD, S. K. Members of the 70-kilodalton heat shock protein contain a highly conserved calmodulin-binding domain. **Molecular Cellular Biology**, Washington, v.10, p.1234 –1238, 1990.
- TING, J.; LEE, A.S. Human gene encoding the 78,000-Dalton glucose-regulated protein and its pseudogene: structure, conservation, and regulation. **DNA**, New York, v. 7, n.4, p. 275 – 286, 1988.
- WATERMAN, M.S. Genomic sequence databases. **Genomics**, San Diego, v.6, n.4, p.700 – 770, 1990.

Comunicado Técnico, 63

Exemplares desta edição podem ser adquiridos na:
Embrapa Milho e Sorgo
Caixa Postal 151 CEP 35701-970 Sete Lagoas, MG
Fone: 0xx31 3779 1000
Fax: 0xx31 3779 1088
E-mail: sac@cnpmis.embrapa.br

Ministério da Agricultura,
Pecuária e Abastecimento

1ª edição
1ª impressão (2003) Tiragem: 200

Comitê de Publicações

Presidente: Ivan Cruz
Secretário-Executivo: Frederico Ozanan Machado Durães
Membros: Antônio Carlos de Oliveira, Arnaldo Ferreira da Silva, Carlos Roberto Casela, Fernando Tavares Fernandes e Paulo Afonso Viana

Expediente

Supervisor editorial: José Heitor Vasconcellos
Revisão de texto: Dilermando Lúcio de Oliveira
Editoração eletrônica: Tânia Mara Assunção Barbosa