



República Federativa do Brasil
Ministério da Economia
Instituto Nacional da Propriedade Industrial

(11) PI 0701826-6 B1



(22) Data do Depósito: 16/03/2007

(45) Data de Concessão: 17/02/2021

(54) Título: PROTEÍNAS DA TEIA DE ARANHA NEPHILENGYS CRUENTATA, AVICULARIA JURUENSIS E PARAWIXIA BISTRIATA ISOLADAS DA BIODIVERSIDADE BRASILEIRA

(51) Int.Cl.: C12N 15/12; C12N 5/10; C12N 15/29; C12N 15/82; A01H 5/00; (...).

(73) Titular(es): EMBRAPA - EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA.

(72) Inventor(es): ELÍBIO LEOPOLDO RECH FILHO; NATALIA CRISTINA VERZA FERREIRA; GIOVANNI RODRIGUES VIANNA; FELIPE RODRIGUES DA SILVA; FRANCISCO JOSÉ LIMA ARAGÃO; LUIZ ALBERTO COLNAGO; ALAN CARVALHO ANDRADE; DANIELA MATIAS DE CARVALHO BITTENCOURT; PEDRO ISMAEL DA SILVA JUNIOR.

(57) Resumo: PROTEÍNAS DA TEIA DE ARANHA NEPHILENGYS CRUENTATA, AVICULARIA JURUENSIS E PARAWIXIA BISTRIATA ISOLADAS DA BIODIVERSIDADE BRASILEIRA A presente invenção está relacionada com moléculas isoladas de ácido nucleico que codificam proteínas da teia de aranha ou um fragmentos das mesmas ou outras derivações das mesmas A invenção ainda discorre sobre um gene quimérico e um vetor de expressão contendo a molécula isolada de ácido nucleico que codifica para a proteína da teia de aranha de Nephilengv cruentata, Avicularia juruensis e Parawixia bistriata. Uma outra concretização da presente invenção são células transformadas contendo o gene quimérico ou o vetor de expressão da presente invenção. Ainda uma outra concretização da presente invenção diz respeito a um método de obtenção de organismos geneticamente modificados contendo genes quiméricos ou vetores de expressão inventivos e a um método de obtenção de proteínas recombinantes da seda das aranhas Nephilengy cruentata, Avicularia juruensis e Parawixia bistriata. Finalmente, a invenção descreve produtos, tais como biofilamentos e composições, utilizando-se da proteína recombinante da presente invenção. A descoberta de novas proteínas da seda de aranhas, bem como sua caracterização e expressão em diferentes sistemas heterólogos será de grande utilidade para inúmeras áreas, como a médica e industrial.

PROTEÍNAS DA TEIA DE ARANHA NEPHILENGYS CRUENTATA, AVICULARIA JURUENSIS E PARAWIXIA BISTRIATA ISOLADAS DA BIODIVERSIDADE BRASILEIRA

CAMPO DA INVENÇÃO

[1] A presente invenção está relacionada com moléculas isoladas de ácido nucleico que codificam proteínas relacionadas à teia de aranha, fragmentos delas ou outras derivações das mesmas. A invenção ainda discorre sobre um gene quimérico e um vetor de expressão contendo moléculas isoladas de ácido nucleico que codificam para proteínas da teia de aranha de *Nephilengys cruentata*, *Avicularia juruensis* e *Parawixia bistriata*. Uma outra concretização da presente invenção são células transformadas contendo uma construção gênica ou um vetor de expressão da presente invenção.

[2] Ainda uma outra concretização da presente invenção diz respeito a um método de obtenção de organismos geneticamente modificados contendo construções gênicas ou vetores de expressão inventivos e a um método de obtenção de proteínas recombinantes da seda das aranhas *Nephilengys cruentata*, *Avicularia juruensis* e *Parawixia bistriata*. Finalmente, a invenção descreve produtos, tais como biofilamentos e composições, utilizando-se de proteínas recombinantes da presente invenção.

FUNDAMENTOS DA INVENÇÃO

[3] A indústria tem demonstrado um grande interesse na obtenção de fibras sintéticas ou naturais que apresentem, simultaneamente, alta resistência, leveza e versatilidade. A maior parte das fibras sintéticas utilizadas atualmente, como o Nylon ou o Kevlar, além do custo elevado de produção, apresentam outras características indesejáveis como densidade alta e campos restritos de aplicação.

[4] Entre as fibras naturais, a seda extraída do bicho da seda (*Bombyx mori*) tem sido utilizada por mais de 5.000 anos na indústria têxtil (Hyde, N. 1984. The queen of textiles. *Natl. Geogr.* 165, 3-49). As fibras desses casulos são compostas por dois filamentos contínuos de seda, fibroína pesada (≈ 350 kDa) e leve (≈ 25 kDa), ligados por proteínas

adesivas denominadas sericinas (Jin H. J., Kaplan D. L. (2003). Mechanism of silk processing in insects and spiders. *Nature* 424:1057-1061). Comercialmente, a sericina é extraída dos casulos pela imersão em água quente e sabão, liberando entre 300 e 1200 m de fibra (fibroína) utilizável por casulo.

[5] Diferente do bicho da seda, as aranhas ainda não foram domesticadas para aplicações têxteis. Esta diferença básica deve-se ao fato da dificuldade de obter grandes populações de aranhas devido a sua natureza solitária e predatória; em adição, sedas de aranhas são produzidas em pequena quantidade e não são possíveis de serem enoveladas como uma fibra simples, de forma semelhante ao que ocorre com o casulo do bicho da seda. Entretanto, as características físicas apresentadas pelas sedas produzidas por aranhas são muito superiores às encontradas na seda de *B.mori* (Dickinson M. H. (1999). Bionics: Biological insight into mechanical design. *Proc. Natl. Acad. Sci. USA* 96:14208-14209). Devido a sua grande elasticidade e resistência, as sedas de teias de aranhas têm despertado grande interesse não só da indústria têxtil como também de indústrias dos mais variados segmentos, como é o caso da indústria de cosméticos (US20050019297).

[6] As aranhas são alguns dos organismos que apresentam maior diversidade e abundância sobre o planeta terra. A ordem Araneae é o segundo maior grupo dentre os aracnídeos e o sétimo dentre os artrópodes, com mais de 39.000 espécies incluídas em 110 famílias (Selden P.A. 1989. Orb-weaving spiders in the early Cretaceous. *Nature*, 340: 711-712; Shear W.A., Palmer J.A., Coddington J.A., Bonamo P.M. 1989. a Devonian Spinnert: early evidence of spiders and silk use. *Science* 246: 479-481; Platnick, N. I. 2006. The world spider catalog, version 6.5. American Museum of Natural History). Só no Brasil, estima-se que hajam de 4.000 a 10.000 espécies de aranhas [Brescovit, A.D. 1999. Araneae. *In*: Biodiversidade do Estado de São Paulo, Brasil. Joly, C.A. & C.E.M. Bicudo (orgs.). Fundação de Amparo à Pesquisa do Estado de São Paulo, São Paulo, SP].

[7] As fibras obtidas de proteínas da seda de aranhas são até cinco vezes mais resistentes que o aço e 30% mais flexíveis que o Nylon, podendo ser utilizadas na fabricação de cabos, fios para pesca, coletes à prova de balas, confecção de pára-quedas, entre outros. Além disso, por serem compostas por material biodegradável, as sedas de

aranhas podem ter aplicações médicas como a fabricação de suturas e curativos cirúrgicos, bandagens, tendões e ligamentos artificiais, matriz para veículo de fármacos, etc. (WO2004016651; Gosline, J.M.; P.A. Guarette; C.S. Orllepp & K.N. Savage, 1999. The mechanical design of spider silks: from fibroin sequence to mechanical function. *The Journal of Experimental Biology*, 202: 3295-3303; Heslot, 1998. Artificial fibrous proteins: a review. *Biochimie*, 80: 19-31).

[8] Sedas produzidas por aranhas são sintetizadas em glândulas localizadas na região do abdômen, e polimerizadas através de uma série de fiandeiras que transformam as proteínas da seda solúveis em água e com alto peso molecular, em fibras insolúveis em água (Benito B., 2002. Synthesizing spider silk. *Trends Biotechnol.* 20:189). Os tipos e a natureza das fibras são diversos e dependem da espécie da aranha [Denny, M.W., 1980. Silks – their properties and functions. *In: Mechanical properties of Biological Materials*. Vincent, J.F.V., Currey, J.D. (Eds), Cambridge University Press, Cambridge, pp. 247-272]. As aranhas possuem sete glândulas produtoras de seda: a glândula aciniforme responsável pela produção de sedas utilizadas no encapsulamento de insetos, a glândula tubiliforme que produz a seda formadora do casulo para depósito dos ovos, e as glândulas flageliforme, “major ampullate”, “minor ampullate”, piriforme e glândula coronata, que são responsáveis pela produção das sedas que constituem a teia em orbital. Entretanto, não são conhecidas famílias de aranhas que possuam todas as sete glândulas.

[9] Dentre as diferentes sedas produzidas pelas aranhas a linha de segurança, sintetizada pela glândula “major ampullate”, é extremamente rígida, com uma força tensora similar à do Kevlar ($4 \times 10^9 \text{ N/m}^2$) associada a uma razoável visco-elasticidade (linha de segurança 35%, Kevlar 5%) (Oroudjev E., Soares J., Arcidiacono S., Thompson J.B., Fossey A.S., Hansma H.G. 2002. Segment nanofibers of spider dragline silk: Atomic force microscopy and single-molecule force spectroscopy. *Proc. Natl. Acad. Sci USA* 99:6460-6465). A seda do tipo linha de segurança é utilizada pelas aranhas na fuga de predadores e como moldura para a construção de sedas. A seda produzida pela glândula “minor ampullate”, usada como reforço na construção da teia, possui uma força tensora semelhante à linha de segurança, mas com menor elasticidade (Colgin M.A., Lewis R.V. 1998. Spider minor ampullate silk proteins contain new repetitive sequences

and highly conserved non-silk-like ‘spacer regions’. Protein Sci. 7:667-672; Hayashi C.Y., Blackledge T.A., Lewis R.V. 2004. Molecular e mechanical characterization of aciniform silk: uniformity of iterated sequence modules in a novel member of the spider silk fibroin gene family. Mol. Biol. Evol. 21:1950-1959). Sedas de aranhas são biopolímeros que apresentam extraordinárias propriedades físicas (Cunniff P.M., Fossey S.A., Auerbach M. A., 1994a. Mechanical and thermal properties of dragline silk from the spider *Nephila clavipes*. Poly. Adv. Technol. 5:401-410, Cunniff P.M., Fossey S.A., Auerbach M. A., 1994b. Mechanical properties of major ampullate gland silk fibers extracted from *Nephila clavipes* spiders. In: Kaplan,D.L., Adams, W.W., Farmer, B., Viney, C. (Eds.). Silk Polymers: Materials science and Biotechnology, American Chemical Society Symposium Series, 544,pp. 234-251; Ko F.K., Jovicic J., 2004. Modelling of mechanical properties and structural design of spider web. Biomacromolecules 5:780-785), mas existe somente uma limitada informação sobre a composição das diferentes sedas produzidas por uma determinada espécie de aranha. As diferentes proteínas da seda contêm aminoácidos repetitivos que variam de acordo com a função da seda, conferindo assim diferentes propriedades mecânicas entre os biopolímeros (Gosline J.M., Guerette P.A., Ortlepp C.S., Savage K.N. (1999). The mechanical design of spider silks: from fibroin sequence to mechanical function. The J. Exp. Biol. 202:3295-3303). Dependendo das condições ambientais e necessidades, a composição dos aminoácidos da seda pode variar consideravelmente, não somente entre diferentes aranhas, mas também na mesma aranha em diferentes dias (Work R.W., Young C.T., 1987. The amino-acid compositions of major and minor ampullate silks of certain orb-web-building spiders (Araneae, Araneidae). J. Arachnol. 15:65-80; Volltrah F. 1999. Biology of spider silk. Int. J. Biol. Macromol. 24:81-88; Craig C.L., Riekel C., Herberstein M.E., Webwe R.S., aplan D., Pierce N.E., 2000. Evidence for diet effects on the composition of silk proteins produced by spiders. Mol. Biol. Evol. 17:1904-1913). Este fato levanta questões sobre as seqüências genômicas e organização dos genes que codificam estas proteínas.

a. Os primeiros estudos para utilização industrial da seda se concentraram principalmente na análise de proteínas de duas espécies: *Nephila clavipes* e *Araneus*

diadematus. A seda linha de segurança, isolada dessas duas espécies é a mais estudada entre todas as fibras sintetizadas por aranhas. A linha de segurança é formada por dois tipos de proteínas produzidas pela glândula “major ampullate”, denominadas de MaSp1 e MaSp 2 (Major Ampullate Spidroin) na *N. clavipes*, e ADF-3 e ADF-4 na *A. diadematus* (Araneus Diadematus Fibroin) (Hinman M.B., Lewis, R.V. 1992. Isolation of a clone encoding a second dragline silk fibroin, *Nephila clavipes* dragline silk is a two protein fiber. J. Biol. Chem. 267:19320-19324; Guerette P., Ginzinger D., Weber B., Gosline J. 1996. Silk properties determined by gland-specific expression of a spider fibroin gene family. Science 272:112-115; Beckwitt R., Arcidiacono S. 1994. Sequence conservation in the C- terminal region of spider silk proteins (Spidroin) from *Nephila clavipes* (Tetragnathidae) and *Araneus bicentarius* (Araneidae). J. Biol. Chem. 269:6661-6663, Beckwit R., Arcidiacono S., Stote R. 1998. Evolution of repetitive proteins: spider silks from *Nephila clavipes* (Tetragnathidae) and *Araneus bicentarius* (Araneidae). Insect. Biochem. Molec. 28:121-130). As proteínas da linha de segurança possuem massa molecular entre 180 kDa e 720 kDa, dependendo das condições de análise (Mello C.M., Senecal K. Yeung B., Vouros P., Kaplan D.I. 1994. Initial characterization of *Nephila clavipes* dragline protein. In: Kaplan D. L. Adams W.W Farmer B. Viney C. (Eds.). Silk Polymers: Materials Science and Biotechnology. American Chemical Society Symposium Series. 544:67-79). De acordo com a composição de aminoácidos destas proteínas, acredita-se que a proporção molecular entre MaSp1 e MaSp2, e entre ADF-4 e ADF-3 seja de aproximadamente 3:1 na fibra linha de segurança (Hinman M.B., Lewis, R.V. 1992. Isolation of a clone encoding a second dragline silk fibroin, *Nephila clavipes* dragline silk is a two protein fiber. J. Biol. Chem. 267:19320-19324; Lombardi S.J., Kaplan D.L. 1990. The amino acid composition of major ampullate gland silk (dragline) of *Nephila clavipes* (Araneae, Tetragnathidae). J. Arachnol. 18:297-306; Guerette P., Ginzinger D., Weber B., Gosline J. 1996. Silk properties determined by gland-specific expression of a spider fibroin gene family. Science 272:112-115). Embora produzidas em espécies diferentes, as proteínas da glândula “major ampullate” possuem em sua composição um alto número de repetições dos mesmos aminoácidos. Em MaSp2 e ADF-3, por exemplo, glicina, alanina, prolina,

serina e tirosina são responsáveis por até 99% dos aminoácidos presentes na sua estrutura (Hayashi C.Y., Lewis R.V. 1998. Evidence from flagelliform silk. cDNA for the structural basis of elasticity and modular nature of spider silks. J. Mol. Biol. 275:773-784).

[10] Diversos trabalhos foram desenvolvidos a respeito das características e aplicações da seda linha de segurança de *Nephila clavipes* [US6268169; US6412261; WO9116351; Beckwitt R. & Arcidiacono S., 1994. Sequence conservation in the C- terminal region of spider silk proteins (Spidroin) from *Nephila clavipes* (Tetragnathidae) and *Araneus bicentenarius* (Araneidae). J. Biol. Chem. 269,6661-6663; Arcidiacono S., Mello, C., Kaplan D.L., Cheley, S., Bayley, H., 1998. Purification and characterization of recombinant spider silk expressed in *Escherichia coli*. Appl. Microbiol. Biotechnol. 49, 31-38].

[11] Além das proteínas da seda produzida pela glândula “Major Ampullate”, outra seda bastante estudada de *N. clavipes* é a produzida pela glândula “Minor Ampullate”. Tal como as anteriores, esta também é formada por dois peptídeos (MiSP1 e MiSP2) compostos de repetições imperfeitas de seqüências de aminoácidos (US5733771).

[12] Hayashi & Lewis, em 1998 (Hayashi C.Y., Lewis R.V., 1998. Evidence from flagelliform silk. cDNA for the structural basis of elasticity and modular nature of spider silks. J. Mol. Bio, 275: 773-784), seqüenciaram a proteína da seda produzida pela glândula Flageliforme de *N. clavipes* (Gosline, J.M.; P.A Guarette; C.S. Ortlepp & K.N. Savage, 1999. The mechanical design of spider silks: from fibroin sequence to mechanical function. The Journal of Experimental Biology, 202: 3295-3303). Resultados semelhantes têm sido publicados para a espécie *Araneus diadematus*, envolvendo a glândula “Major Ampullate” que produz as proteínas ADMAG1 e ADMAG2 (Guerette P., Ginzinger D., Weber B ., Gosline J., 1996. Silk properties determined by gland-specific expression of a spider fibroin gene family. Science, 272:112-115).

[13] Baseado na análise de DNA pode-se afirmar que todas as proteínas constituintes das sedas de aranhas são formadas por unidades peptídicas repetitivas. Estas podem ser agrupadas em quatro grandes grupos: GPGXX (onde X freqüentemente representa Q), seqüências ricas em alanina (An ou (GA)n), GGX (onde X = A, Y, L ou Q) e os

espaçadores. Uma quinta categoria é representada por regiões não repetitivas nas extremidades N e C-terminais das proteínas, normalmente cadeias constituídas por 100 ou mais aminoácidos (Xu M., Lewis R.V., 1990. Structure of a protein superfiber: Spider dragline silk. Proc. Natl. Acad. Sci. USA 87:7120-7124; Hinman M.B., Lewis, R.V. 1992. Isolation of a clone encoding a second dragline silk fibroin, *Nephila clavipes* dragline silk is a two protein fiber. J. Biol. Chem. 267:19320-19324; Colgin M.A., Lewis R.V. 1998. Spider minor ampullate silk proteins contain new repetitive sequences and highly conserved non-silk-like 'spacer regions'. Protein Sci. 7:667-672; Hayashi C.C.Y., Shipley N.H., Lewis R.V. 1999. Hypothesis that correlate the sequence, structure and mechanical properties of spider silk proteins. Int. Biol. Macromol. 24:271-275; Oroudjev E., Soares J., Arcidiacono S., Thompson J.B., Fossey A.S., Hansma H.G. 2002. Segment nanofibers of spider dragline silk: Atomic force microscopy and single-molecule force spectroscopy. Proc. Natl. Acad. Sci USA 99:6460-6465; Tai P.L., Hwang G.Y., Tso I.M., 2004. Inter-specific sequence conservation and intra-individual sequence variation in a spider silk gene. Inter. J. Biol. Macromolecules 34:295-301).

[14] De acordo com diferentes estudos, a maioria das unidades repetitivas presentes nas sedas de aranhas, apresentam propriedades estruturais específicas (Xu M., Lewis R.V., 1990. Structure of a protein superfiber: Spider dragline silk. Proc. Natl. Acad. Sci. USA 87:7120-7124; Hayashi C.Y., Lewis R.V. 1998. Evidence from flagelliform silk. cDNA for the structural basis of elasticity and modular nature of spider silks. J. Mol. Biol 275:773-784; Van Beek J.D., Hess S., Vollrath F., Meier B.H., 2002. The molecular structure of spider dragline silk: Folding and orientation of the protein backbone. Proc. Natl. Acad. Sci USA 99:10266-10271; Bini E., Knight D.P., Kaplan D.L. 2004. Mapping domain structures in silks from insects and spiders related to protein assembly. J. Mol. Biol. 335:27-40; Scheibel T. 2004. Spider silks: recombinant synthesis, assembly, spinning, and engineering of synthetic proteins. Microb. Cell Fact. 3:14; Stantcheva N.N.P., Mason S.J.M. 2004. Molecular studies of a novel dragline silk from nursery web spider, *Euprosthenops* sp (Psauridae). Comp. Biochem. Physiol. 138:371-376). O módulo GPGXX é responsável pela formação de estruturas β -espiral, e provavelmente confere elasticidade à seda. A seda flageliforme, produzida pela glândula

flageliforme, apresenta elasticidade superior a 200%, sendo esta constituída por pelo menos 43 módulos GPGXX em cada unidade repetitiva (Hayashi C.Y., Lewis R.V. 2000. Molecular architecture and evolution of a modular of spider silk protein gene. Science 287:1477-1479). Em concordância com a baixa elasticidade da seda linha de segurança, esta apresenta apenas nove repetições deste motivo antes de ser interrompido por outro módulo. Módulos ricos em alanina normalmente são constituídos por 6-9 resíduos deste aminoácido, estes são responsáveis pela formação de folhas- β que conferem rigidez à fibra. Sedas produzidas pelas glândulas “major” e “minor ampullate” são ambas muito fortes e apresentam motivos An ou (GA)_n, entretanto tal motivo não é encontrado na seda flageliforme (Gatesy J., Hayashi C., Motriuk D., Woods J., Lewis R. 2001. Extreme diversity, conservation, and convergence of spider silk fibroin sequences. Science 291:2603-2605). GGX, uma hélice 310, por sua vez, forma uma matriz amorfa que conecta regiões cristalinas conferindo elasticidade à fibra provavelmente em conjunto com GPGXX. Tal motivo pode ser encontrado tanto em sedas produzidas pelas glândulas flageliforme, “major” e “minor ampullate”. As regiões espaçadoras são constituídas por grupos carregados que separam as regiões ricas em glicina (Colgin M.A., Lewis R.V. 1998. Spider minor ampullate silk proteins contain new repetitive sequences and highly conserved non-silk-like ‘spacer regions’. Protein Sci. 7:667-672; Hayashi C.C.Y., Shipley N.H., Lewis R.V. 1999. Hypothesis that correlate the sequence, structure and mechanical properties of spider silk proteins. Int. Biol. Macromol. 24:271-275), entretanto não se tem conhecimento da sua função estrutural. As terminações não repetitivas são comuns em todas as fibras produzidas por aranhas da família Araneidae, com seqüências C-terminais altamente conservadas entre as espécies (Bini E., Knight D.P., Kaplan D.L. 2004. Mapping domain structures in silks from insects and spiders related to protein assembly. J. Mol. Biol. 335:27-40; Hayashi C.Y., Blackledge T.A., Lewis R.V. 2004. Molecular e mechanical characterization of aciniform silk: uniformity of iterated sequence modules in a novel member of the spider silk fibroin gene family. Mol. Biol. Evol. 21:1950-1959; Stantcheva N.N.P., Mason S.J.M. 2004. Molecular studies of a novel dragline silk from nursery web spider, *Euprosthenops sp* (Psauridae). Comp. Biochem. Physiol. 138:371-376; Tian M., Liu C., Lewis R., 2004. Analysis of

major ampullate silk cDNA from two non-orb-weaving spiders. *Biomacromolecules* 5:657-660). Estudos recentes com ADF-3 e 4 revelaram uma estrutura em α -hélice formada pela região C-terminal, levantando a hipótese desta região possuir importante papel na polimerização da fibra (Huemmerich D., Scheibel T., Vollrath F., Cohen S., Gat U., Ittah I. 2004. Novel assembly properties of recombinant spider dragline silk proteins. *Curr. Biol.* 14:472-476).

[15] O mecanismo de fiação, ou seja, a polimerização das proteínas solúveis em água para fibras insolúveis, é um processo que se inicia com um aumento na concentração da proteína no lúmen glandular formando a “solução de fiação”. Na glândula major ampullate, por exemplo, as proteínas da seda linha de segurança estão presentes numa concentração superior a 50% (m/v) (Artkins E. 2003. Silk’s secrets. *Nature* 424:1010; Scheibel T. 2004. Spider silks: recombinant synthesis, assembly, spinning, and engineering of synthetic proteins. *Microb. Cell Fact.* 3:14). O aumento da concentração da MaSp faz com que ocorra uma modificação estrutural nestas proteínas, passando de uma estrutura “coil” para β -hélice e aumentando a estabilidade das mesmas (Dicko C., Knight D., Kenney J.M., Vollrath F. 2004. Structural conformation of spidroin in solution: A synchrotron radiation circular dichroism study. *Biomacromolecules* 5:758-767). Deste modo, a aranha mantém uma concentração relativamente alta de proteína em solução aquosa, sem resultar na formação de folhas- β insolúveis. A polimerização das proteínas ocorre durante a passagem da solução de fiação pelo ducto glandular, concomitantemente com a extração de água, sódio e cloreto. Íons de hidrogênio e potássio são secretados causando uma diminuição no pH de 6.9 para 6.3 (Chen X., Knight D.P., Shao Z., Vollrath F. 2002. Conformation transition in silk protein films monitored by time-resolved fourier transform infrared spectroscopy: Effect of potassium ions on *Nephila* spidroin films. *Biochemistry* 41:14944-14950; Dicko C., Vollrath F., Kenney J.M. 2004. Spider silk protein refolding is controlled by changing pH. *Biomacromolecules* 5:704-710). Tais alterações desencadeiam o alinhamento das proteínas na parte distal do ducto, enquanto seus seguimentos hidrofóbicos de poli-A se alinham e se aproximam, estas são expostas a um ambiente cada vez mais hidrofóbico, o que muito provavelmente instiga a conversão estrutural destas proteínas para folhas- β

(Scheibel T. 2004. Spider silks: recombinant synthesis, assembly, spinning, and engineering of synthetic proteins. *Microb. Cell Fact.* 3:14), e conseqüente polimerização da fibra.

[16] A alta organização das estruturas das fibras, extensivas pontes de hidrogênio e interações de Van der Waals, induzem a exclusão da água das regiões entre as folhas- β . Sedas de aranhas são insolúveis em água, ácidos e bases diluídas, agentes caotrópicos como uréia e hidrocloreto de guanidina e a maioria dos solventes orgânicos (Lombardi S.J., Kaplan D.L. 1990. The amino acid composition of major ampullate gland silk (dragline) of *Nephila clavipes* (Araneae, Tetragnathidae). *J. Arachnol.* 18:297-306). As sedas são também resistentes à grande parte das enzimas proteolíticas. Uma pequena solubilização das sedas foi obtida com a utilização de soluções salinas de brometo de lítio, tiocianato de lítio, cloreto de cálcio e outros sais de cálcio. Altas concentrações da mistura de ácido propiônico/hidroclorídrico bem como ácido fórmico, também pode ser utilizado (Mello C.M., Senecal K. Yeung B., Vouros P., Kaplan D.I. 1994. Initial characterization of *Nephila clavipes* dragline protein. In: Kaplan D. L. Adams W.W Farmer B. Viney C. (Eds.). *Silk Polymers: Materials Science and Biotechnology*. American Chemical Society Symposium Series. 544:67-79 Lewis R.V., Hinman M., Kothakota S., Fournier M.J. 1996. Expression and purification of a spider silk proteins: A new strategy for producing repetitive proteins. *Express. Prif* 4:400-406).

[17] Diferentes grupos têm tentado processar a seda artificialmente utilizando diferentes tipos de diluentes. A maioria dos esforços tem envolvido o processamento líquido utilizado para *B. mori*. Proteínas recombinantes da seda têm sido processadas utilizando solventes como hexafluoroisopropanol (WO 9429450) como diluente ou soluções de proteínas diluídas em soluções concentradas de ácido fórmico [Lewis R.V., Hinman M., Kothakota S., Fournier M.J., 1996. Expression and purification of a spider silk protein: A new strategy for producing repetitive proteins. *Express. Prif* (4): 400-406]. Entretanto, em ambos casos, as propriedades mecânicas da seda natural não tem sido reproduzidas com eficiência (Fahnestock S.R., Bedzyk L.A, 1997. Production of synthetic spider dragline silk protein in *Pichia pastoris*. *Appl. Microbiol. Biotechnol.* 47: 23-32).

[18] Outros ensaios conseguiram solubilizar as sedas de aranhas através da imersão das fibras em concentrações salinas, como brometo de lítio, tiocianato de lítio, cloreto de cálcio e outros sais de cálcio. Altas concentrações da mistura de ácido propiônico/hidroclorídrico bem como ácido fórmico também podem ser utilizados [Mello C.M., Senecal K., Yeung B., Vouros P., Kaplan D.I., 1994. Initial characterization of *Nephila clavipes* dragline protein. In: Silk Polymers: Materials Science and Biotechnology. American Chemical Society Symposium Series. Kaplan D. L., Adams W.W, Farmer B., Viney C. (Eds.), 544 pp].

[19] Inicialmente, o processo que conduz à alta hidrofobia das proteínas da seda de aranhas direciona a formação de seqüências cristalinas repetitivas. No bicho da seda, mudanças nas condições fisiológicas como pH e concentração de sais na glândula acompanham o processamento e presumivelmente auxiliam na manutenção da solubilidade. A quebra física gerada durante o processamento (spinning) da seda solúvel parece ser, em grande parte, responsável pela conversão da proteína solúvel em fibra insolúvel no processo natural de processamento [Ilzuka, E., 1985. Silk: an overview. J. Appl. Polymer. Sci. Jpn. 41: 163-171; Ilzuka, E., 1985. Silk thread: Mechanism of spinning and its mechanical properties. J. Appl. Polymer Sci Jpn. 41: 173-185; Magoshi J., Magoshi Y., Nakamura S., 1985. Crystallization, liquid crystal, and fiber formation of silk fibroin. J. Appl. Polymer Sci. 41: 187-204; Magoshi J., Magoshi Y., Nakamura S., 1994. Mechanism of fiber formation of silkworm. In: Silk Polymers: materials Science and Biotechnology, American Chemical Society Symposium Series, Kaplan D.L., Adams W. W., Farmer B., Viney C. (Eds.), 544 pp].

[20] A produção em larga escala das fibras da seda de aranhas tornaria possível a produção de uma nova geração de biomateriais com alto grau de biodegradabilidade, que envolvem aplicações práticas em diferentes segmentos do setor industrial. A inabilidade de domesticar as aranhas para produzir quantidade suficiente de proteínas para o seu estudo e utilização comercial induziu o desenvolvimento de estudos para viabilizar a produção das proteínas da seda em sistemas de expressão heterólogos em larga escala. O recente sucesso da clonagem de cDNAs e genes sintéticos e a expressão de proteínas recombinantes da seda de aranhas em diferentes sistemas foram fundamentais para o

desenvolvimento do entendimento da estrutura, processamento e função destas proteínas, e de suas importantes propriedades mecânicas (Kaplan D.L., Adams W.W., Farmer B., Viney C. 1994. Silk Polymers: Materials Science and Biotechnology. American Chemical Society Symposium Series Volume 544; Kaplan D.L., Mello C. M., Arcidiacono S., Fossey S., Senecal K., Muller W., 1998. Silk. In: McGrath, K., Kaplan. D. L. (Eds.), Protein Based Materials. Birkhauser, Boston).

[21] Estudos estão sendo conduzidos para obter conhecimentos sobre estes processos. Entretanto, a natureza altamente repetitiva dos genes, os específicos códons utilizados pelas aranhas e a estrutura secundária incomum adotada pelo mRNA tem resultado em uma tradução ineficiente das proteínas, limitando o tamanho da fibra produzida (Fahnestock S.R., Bedzyk L.A. 1997. Production of synthetic spider dragline silk protein in *Pichia pastoris*. Appl. Microbiol. Biotechnol. 47:23-32; Scheibel T. 2004. Spider silks: recombinant synthesis, assembly, spinning, and engineering of synthetic proteins. Microb. Cell Fact. 3:14). Em função da natureza repetitiva das seqüências, as primeiras pesquisas feitas com mRNAs coletados da glândula “major ampullate” de *N. clavipes* não obtiveram sucesso na tradução in vitro (Candelas G.C., Cintron J.J. 1981. A spider fibroin and its synthesis. J.Exp. Zool. 216:1-6; Candelas G.C., Lopez, F. 1983. Synthesis of fibroin in the cultured glands of *Nephila clavipes*. Comp. Biochem. Physiol. 74:637-641; Candelas G.C., Candelas T., Ortiz A., Rodriguez O. 1983. Translation pauses during a spider fibroin synthesis. Biochem. Biophys. Res. Commun. 116:1033-1038).

[22] Diferentes sistemas heterólogos de expressão estão sendo utilizados na tentativa de produzir fibras de aranhas. Estudos recentes, utilizando construções feitas a partir de cDNAs parciais dos genes da linha de segurança, produziram proteínas recombinantes em *E. coli* (Arcidiacono S., Mello, C., Kaplan D.L., Cheley, S., Bayley, H. 1998. Purification and characterization of recombinant spider silk expressed in *Escherichia coli*. Appl. Microbiol. Biotechnol. 49:31-38), em cultura de células MAC-T (bovino) e BHK (hamster) (Lazaris A., Arcidiacono S., Huang Y., Zhou J.F, Duguay F., Chretien N., Welsh E. A., Soares J.W., Karatzas C.N. 2002. Spider silk fibers spun from soluble recombinant silk produced in mammalian cells. Science 295:472-476), e em linhagens celulares do inseto *Spodoptera frugiperda* usando o sistema de expressão de baculovirus

(Huemmerich D., Scheibel T., Vollrath F., Cohen S., Gat U., Ittah I. 2004. Novel assembly properties of recombinant spider dragline silk proteins. *Curr. Biol.* 14:472-476). Vários estudos mostram a utilização de construções contendo cDNA de genes que codificam proteínas das glândulas “minor ampullate” e flageliforme de aranhas, como é o caso dos documentos de patente US576677 e US5994099. O documento US5728810 descreve a expressão de seqüências de Spidroina 1 e 2 de *N. clavipes* em microorganismos. Os documentos US6608242, US20050010035 e WO0194393 relatam um método para produzir proteínas sintéticas da seda de aranha em plantas e construções expressando proteínas sintéticas da seda derivadas de *Nephila clavipes* e outras espécies de aranhas. O documento de patente CN1380418 relata a construção sintética da proteína da teia de aranha “Spidroin” para expressar em plantas de algodão. Estudos mostram a expressão da proteína da seda de aranhas em animais, como é o caso dos documentos de patente WO9947661 e US2001042255, que descrevem métodos para produção recombinante de biofilamentos no leite e/ou urina de animais transgênicos.

[23] Genes sintéticos baseados na seqüência de MaSp de *N. clavipes* e *Araneus gemmoides* também foram utilizados para a expressão de proteínas heterólogas em *E. coli* (Fahnestock S.R., Bedzyk L.A. 1997. Production of synthetic spider dragline silk protein *Pichia pastoris*. *Appl. Microbiol. Biotechnol.* 47:23-32), *Pichia Pastoris* (Fahnestock S.R., Bedzyk L.A. 1997. Production of synthetic spider dragline silk prote in *Pichia pastoris*. *Appl. Microbiol. Biotechnol.* 47:23-32) e plantas (US20040210956; Scheller J., Gurhuns K.H., Grosse F., Conrad U. 2001. Production of spider silk proteins in tobacco and potato. *Nat. Biotechnol.* 19:573-577; Piruzian E.S., Bogush, V.G., Sidoruk K.V., Goldenkova I.V., Mysiyshuk K.A., Debabov V.G. 2003. Construction of synthetic genes for analogues of spider silk spidroin 1 and their expression in Tabacco plants. *Mol. Biol.* 27:554-560; Scheller J., Henggeler D., Viviani A., Conrad U. 2004. Purification of spider-elastin from transgenic plants and application for human chondrocyte proliferation. *Transg. Res.* 13:51-57). Infelizmente nenhum gene MaSp foi clonado em sua totalidade, e os dados disponíveis dizem respeito a clones de cDNA parciais, iniciando pela terminação 3’ dos genes da seda linha de segurança de *N. clavipes*, *A. diadematus* e outras espécies (Xu M., Lewis R.V., 1990. Structure of a protein superfiber:

Spider dragline silk. Proc. Natl. Acad. Sci. USA 87:7120-7124; Hinman M.B., Lewis, R.V. 1992. Isolation of a clone encoding a second dragline silk fibroin, *Nephila clavipes* dragline silk is a two protein fiber. J. Biol. Chem. 267:19320-19324; Beckwitt R., Arcidiacono S. 1994. Sequence conservation in the C- terminal region of spider silk proteins (Spidroin) from *Nephila clavipes* (Tetragnathidae) and *Araneus bicentenarius* (Araneidae). J. Biol. Chem. 269:6661-6663 Guerette P., Ginzinger D., Weber B., Gosline J. 1996. Silk properties determined by gland-specific expression of a spider fibroin gene family. Science 272:112-115 Hayashi C.Y., Lewis R.V. 1998. Evidence from flagelliform silk. cDNA for the structural basis of elasticity and modular nature of spider silks. J. Mol. Biol 275:773-784). Uma explicação para tais resultados seria a possível degradação do mRNA durante a sua extração das glândulas produtoras de seda para a construção das bibliotecas de cDNA, uma vez que mRNAs de grande comprimento são mais susceptíveis a degradação enzimática (Stantcheva N.N.P., Mason S.J.M. 2004. Molecular studies of a novel dragline silk from nursery web spider, *Euprosthenops sp* (Psauridae). Comp. Biochem. Phisiol. 138:371-376). A produção de proteínas das sedas de aranhas em sistemas heterólogos, e também a manipulação da estrutura primária destas proteínas, fazendo uso da engenharia da estrutura modular tem sido baseada em conhecimentos sobre as sedas de aranhas naturais (Cappello J., Crissman J., Dorman M. 1990. Genetic engineering of structural protein polymers. Biotech.Prog. 6:198 – 202; Scheibel T. 2004. Spider silks: recombinant synthesis, assembly, spinning, and engineering of synthetic proteins. Microb. Cell Fact. 3:14; Kang W.J., Cho S.S., Huh H., Chung D.T 1997. Identification of dynamic behavior of sheet metals for an auto-body with tension split Hopkinson bar. Trans. KSME 21:2209-2219, Kang W.J., Cho S.S., Huh H., Chung D.T. 1999. Modified Johnson-Cook model for dynamic behaviour of sheet metals for auto-body Crash-worthiness. Int. J. Vehicle Design, 21:424-435).

[24] A possibilidade de produzir proteínas das sedas de aranhas em sistemas heterólogos em larga escala com a cinética e função desejada deverá permitir sua aplicação em vários produtos médicos como curativos e microfilamentos de suturas para neurocirurgias. Fibras de alta performance que poderão ser utilizadas em diferentes aplicações técnicas e industriais. As sedas poderão ser usadas em cordas e redes de pesca

especiais, pára-quedas, em aplicações balísticas (coletes à prova de balas), produtos esportivos, na indústria têxtil, na indústria de cosméticos e como matéria prima de baixo peso para a construção de aviões. Um adicional benefício será o uso das proteínas das sedas de aranhas para a produção de microbiocidas e defensinas que poderão ser utilizadas contra doenças e pragas da agricultura, pecuária e saúde humana.

[25] Há portanto a necessidade de se identificar novas proteínas da seda de aranhas, expressá-las em diferentes sistemas e desenvolver outros métodos que permitam solucionar os problemas ainda encontrados na área e acima descritos.

[26] A presente invenção descreve novas proteínas da seda extraídas de *Nephilengys cruentata*, *Avicularia juruensis* e *Parawixia bistrata* e a expressão dessas proteínas em sistemas recombinantes. A presente invenção descreve ainda a expressão de proteínas da seda em células vegetais, animais e microrganismos produtores de fibra a fim de desenvolver novos biomateriais fibrosos com características aprimoradas.

SUMÁRIO DA INVENÇÃO

[27] A descoberta de novas proteínas da seda de aranhas, bem como sua caracterização e expressão em diferentes sistemas heterólogos será de grande utilidade para inúmeras áreas, como a médica e a industrial.

[28] As proteínas da seda de aranha podem ser obtidas através de polipeptídeos sintéticos tendo seqüências de aminoácidos substancialmente similares a uma seqüência consenso de proteína da seda ou através de polipeptídeos expressos de seqüências de ácido nucleico codificando uma proteína da seda natural ou engenheirada, ou derivados destes. Dependendo da aplicação para a qual a proteína da seda é requerida, pode ser desejável formar fibras de uma simples proteína da teia de aranha ou de combinações de diferentes proteínas da teia de aranha.

[29] De acordo com um aspecto da invenção são providas moléculas isoladas de ácido nucléico de aranha caracterizadas por compreender:

- a. seqüências substancialmente similares a qualquer uma das seqüências selecionadas do grupo identificado como SEQ ID NO 1-19;
- b. complementos das seqüências descritas em SEQ ID NO1-19;

- c. complementos reversos das seqüências descritas em SEQ ID NO1-19;
- d. seqüências reversas das seqüências descritas em SEQ ID NO1-19.

[30] De acordo com um segundo aspecto da invenção é provido um gene quimérico caracterizado por compreender:

um promotor opcionalmente ligado a uma seqüência líder e operacionalmente ligado a;
uma seqüência codificadora substancialmente similar à qualquer uma das seqüências identificadas como SEQ ID NO1-19.

[31] De acordo com um outro aspecto da presente invenção, é provido um vetor de expressão caracterizado por compreender:

- a. um promotor opcionalmente ligado a uma seqüência líder e operacionalmente ligado a;
- b. uma seqüência codificadora substancialmente similar à qualquer uma das seqüências identificadas como SEQ ID NO1-19 operacionalmente ligada a;
- c. um sinal de terminação;
- d. uma origem de replicação;
- e. um marcador seletivo; e
- f. um sítio de clonagem.

[32] Uma quarta concretização da presente invenção diz respeito a moléculas isoladas de proteína da seda de aranha caracterizadas por compreender seqüências substancialmente similares a qualquer uma das seqüências selecionadas do grupo identificado como SEQ ID NO 20-38.

[33] Em um outro aspecto da invenção, células hospedeiras compreendendo pelo menos uma das proteínas da seda de aranha codificadas por ácidos nucleicos são providas. Tais células hospedeiras incluem mas não estão limitadas a células bacterianas, células de fungos, células de insetos, células de mamíferos e células vegetais. Células hospedeiras superexpressando uma ou mais proteínas da seda de aranha codificadas por ácidos nucleicos da invenção provém reagentes úteis para inúmeras aplicações, incluindo, mas não limitado a, produção de fibras de seda compreendendo pelo menos uma proteína

da seda que pode ser incorporada dentro de um material para modular as propriedades estruturais do material.

[34] A invenção discorre também sobre um método para produzir um organismo geneticamente modificado caracterizado pelo fato de compreender as seguintes etapas:

- a. transformar uma célula, tecido, órgão ou embrião com um gene quimérico de acordo com qualquer uma das reivindicações 3 a 11 ou um vetor de expressão de acordo com qualquer uma das reivindicações 12 a 23;
- b. selecionar células transformadas, calos de células, embriões ou sementes;
- c. regenerar plantas maduras, embriões maduros ou microorganismos de células transformadas, calos de células, embriões ou sementes selecionados na etapa (b);
- d. selecionar plantas maduras, embriões maduros ou células de microorganismos da etapa (c) contendo o gene quimérico ou vetor de expressão com as seqüências de nucleotídeos que codificam a proteína da seda da aranha.

[35] A presente invenção discorre ainda sobre um método para produzir proteínas recombinantes da seda de aranhas em células procariotas e eucariotas caracterizado pelo fato de compreender as seguintes etapas:

- a. transformar uma célula, tecido, órgão ou embrião com um vetor de expressão de acordo com qualquer uma das reivindicações 12 a 23;
- b. selecionar células transformadas, calos de células, embriões ou sementes;
- c. regenerar plantas maduras, embriões maduros ou microorganismos de células transformadas, calos de células, embriões ou sementes selecionados na etapa (b);
- d. selecionar plantas maduras, embriões maduros ou células de microorganismos da etapa (c) contendo vetor de expressão com as seqüências de nucleotídeos que codificam a proteína da seda da aranha;
- e. Fazer a extração da proteína recombinante da seda de aranha produzida nos organismos selecionados na etapa (d).

[36] A invenção engloba também as proteínas recombinantes com atividade microbicida, defensina e dermatológica, bem como composições dermatológicas para uso farmacêutico e composições microbicidas para uso agrícola.

[37] A invenção também relata uma composição dermatológica caracterizada pelo fato de compreender:

- a. Proteína recombinante de acordo com a reivindicação 42;
- b. Veículo farmaceuticamente aceitável.

[38] A invenção discorre também sobre uma composição microbicida caracterizada pelo de compreender:

- a. Proteína recombinante de acordo com a reivindicação 43;
- b. Veículo agriculturalmente aceitável e, opcionalmente,
- c. Aditivos.

[39] Por fim a invenção descreve biopolímeros produzidos a partir das proteínas recombinantes da presente invenção.

[40] BREVE DESCRIÇÃO DAS FIGURAS

[41] Figura 1: Alinhamento entre SEQ ID NO 1 e gi|7106229. As seqüências inseridas nos retângulos com fundo escuro destacam aminoácidos idênticos entre as duas seqüências. Os números sobre a seqüência identificam a posição no alinhamento, não correspondendo à posição em qualquer uma das seqüências. A diferença se deve à inserção de espaços (gaps) para manter o alinhamento.

[42] Figura 2 – Vetores contendo peptídeo sinal e promotor da beta-conglicina utilizados para bombardeamento em sistema de transformação e de co-transformação de plantas de soja e algodão.

[43] Figura 3 - Vetor pAG1 utilizado para bombardeamento em sistema de transformação e de co-transformação de plantas de soja e algodão.

[44] Figura 4 - Vetor pET19b:SEQ 1-19 utilizado transformação de *E. coli* para expressão das proteínas de seda.

[45] Figura 5 - Vetor pCMV-Script contendo as seqs 1 a 19 para expressão em células de mamíferos. MCS: multiple cloning site; sítio de clonagem contendo 15 enzimas diferentes com sítios únicos.

[46] Figura 6 - Vetor pBC1 contendo as seqs 1 a 19 para expressão das proteínas de teia no leite de animais transgênicos.

[47] Figura 7: Modelo de estrutura protéica obtido para a SEQ ID NO 1 (Produto do Gene: *Nephilengys cruentata* – NCFlag). A. representa visão horizontal da estrutura e B. representa a visão do topo a partir de uma das extremidades.

DESCRIÇÃO DETALHADA DA INVENÇÃO

[48] As seguintes definições são providas para facilitar o entendimento da presente invenção:

[49] O termo “molécula isolada de ácido nucleico” é utilizado para se referenciar aos ácidos nucleicos da presente invenção. Este termo, quando aplicado para DNA, refere-se à molécula de DNA que é separada das seqüências imediatamente contíguas (nas direções 5’ e 3’) que ocorrem naturalmente no genoma do organismo do qual foram derivadas. Por exemplo, a “molécula isolada de ácido nucleico” pode estar inserida em um vetor, tal como um plasmídeo ou um vetor de vírus, ou integrado dentro de um DNA genômico de um procarioto ou eucarioto. Uma “molécula isolada de ácido nucleico” pode compreender também uma molécula de cDNA. Uma molécula isolada de ácido nucleico inserida em um vetor é também referida às vezes aqui como molécula de ácido nucleico recombinante. O termo “molécula isolada de ácido nucleico” também poderá ser aplicado à moléculas de RNA transcritas de uma molécula isolada de DNA conforme descrita acima. Alternativamente o termo pode se referir a uma molécula de RNA que tem sido suficientemente separada das moléculas de RNA às quais estava associada em seu estado natural (isto é, em células ou tecidos).

[50] A definição dos termos “complemento”, “complemento reverso” e “seqüência reversa” como usados aqui é ilustrada pelo seguinte exemplo: para a seqüência 5’AGTGAAGT3’, o complemento é 3’TCACTTCA5’, o complemento reverso é 3’ACTTCACT5’ e a seqüência reversa é 5’TGAAGTGA3’.

[51] “Seqüência codificadora” refere-se à seqüência de DNA que codifica uma proteína específica e exclui a seqüência não codificadora. Uma “seqüência codificadora interrompida” significa a seqüência que atua como separadora (por exemplo, um ou mais

íntrons ligados através de junções). Um “íntron” é uma sequência de nucleotídeo que é transcrita e está presente no pré mRNA, mas é removida através de clivagem e a re-ligação do mRNA dentro da célula gerando um mRNA maduro que pode ser traduzido em uma proteína. Exemplos de íntrons incluem, mas não são limitados a, íntron pdk2, íntron catalase da mamona, íntron Delta 12 desnaturase de algodão, Delta 12 desnaturase de Arabidopsis, íntron ubiquitina de milho, íntron de SV40, íntrons do gene da malato sintase.

[52] Uma “construção gênica” é um gene compreendendo um promotor e uma região codificadora de diferentes origens. No caso da presente invenção, a construção gênica compreende os polinucleotídeos da presente invenção ligados de forma associada ou isoladamente à regiões reguladoras da expressão, como promotores e sinais de terminação.

[53] A obtenção de construções gênicas compreendendo promotores ligados a ácidos nucleicos é conhecida no estado da técnica e pode ser encontrada em Sambrook, et al. (Molecular Cloning, A Laboratory Manual, 2nd ed. (1989), Cold Spring Harbor Laboratory Press).

[54] O termo “vetor” diz respeito a um replicon, tal como plasmídeo, cosmídeo, BAC, fago ou vírus, no qual outras seqüências genéticas ou elementos (seja DNA ou RNA) possam ser ligadas para serem replicadas juntas com o vetor. Preferencialmente o vetor derivado de vírus é selecionado entre os bacteriófagos, vaccínias, retrovírus ou vírus do papiloma bovino. Um “vetor de expressão” é um vetor especializado que contém um gene com regiões regulatórias necessárias para a expressão em uma célula hospedeira. Tais vetores podem ser obtidos comercialmente, incluindo Clontech Laboratories, Inc (Palo Alto, Calif.), Stratagene (La Jolla, Calif.), Invitrogen (Carlsbad, Calif.), New England Biolabs (Beverly, Mass.) e Promega (Madison, Wis.). Alguns exemplos de vetores utilizados na presente invenção são, mas não estão limitados a pMAC/PS, pCMV-Gal, pGFP/NEO. O termo “operacionalmente ligado” significa que as seqüências regulatórias necessárias para expressão da seqüência codificante são colocadas na molécula de DNA em posições apropriadas relativas à seqüência codificante para efeito de sua expressão. Essa mesma definição é às vezes aplicada para o arranjo de seqüências

codificantes e elementos controladores da transcrição (por exemplo, promotores, acentuadores ou “enhancers” e elementos de terminação) no vetor de expressão. Uma região codificante exógena é tipicamente flanqueada por regiões regulatórias operacionalmente ligadas que regulam a expressão da região codificante exógena em uma célula transformada (podendo ser microorganismo, vegetal ou animal). Uma região regulatória típica operacionalmente ligada a uma região codificante exógena inclui um promotor, isto é, um fragmento de ácido nucleico que pode causar transcrição de regiões codificantes exógenas, posicionado na região 5’ da região codificante exógena. A presente invenção não está limitada para o uso de qualquer promotor em particular e uma ampla variedade de promotores são conhecidos no estado da técnica. Os promotores podem ser, mas não estão limitados a, induzíveis, constitutivos e tecido-específicos. Preferencialmente, o promotor da presente invenção é escolhido do grupo dos promotores dos genes da fibra de algodão, podendo ser, mas não estando limitado a E6, H6S, Rac13, LTP, ACP, Expansina, CAP, Anexina, FbL2A e actina 2.

[55] Em um dos aspectos da invenção, o promotor é um promotor constitutivo. Em outro aspecto da invenção, a atividade do promotor é estimulada por fatores externos ou internos tais como, mas não limitado a, hormônios, compostos químicos, impulsos mecânicos, e condições de estresse biótico ou abiótico. A atividade do promotor também pode ser regulada de maneira temporal e espacial (como por exemplo, promotores tecido-específicos e promotores regulados durante o desenvolvimento).

[56] O promotor pode conter elementos “enhancers”. Um “enhancer” é uma sequência de DNA que pode estimular a atividade do promotor. Ela pode ser um elemento inato do promotor ou um elemento heterólogo inserido para aumentar o nível e/ou a tecido-especificidade de um promotor. “Promotores constitutivos” referem-se àqueles que dirigem a expressão gênica em todos os tecidos e durante todo tempo. Promotores “tecido-específicos” ou “desenvolvimento-específicos” são aqueles que dirigem a expressão gênica quase que exclusivamente em tecidos específicos, tais como folhas, raízes, caules, flores, frutos ou sementes, ou em estágios do desenvolvimento específicos em um tecido, como no início ou final da embriogênese.

[57] Em um dos aspectos da invenção, o promotor é um promotor expresso em plantas. Como usado aqui, o termo “promotor expresso em plantas” significa uma sequência de DNA que é capaz de iniciar e/ou controlar a transcrição em uma célula de planta. Isso inclui qualquer promotor de origem vegetal; qualquer promotor de origem não vegetal que seja capaz de direcionar a expressão em uma célula vegetal, por exemplo promotores de origem viral ou bacteriana tais como o 19S e 35S do CaMV (como mencionado no pedido de patente US20030175783, Hapster et al, 1988 Mol. Gen. Genet. 212, 182-190), promotor do bacteriófago T7 e promotores do gene presentes no T-DNA de *Agrobacterium*; promotores tecido-específicos ou órgão-específicos, incluindo mas não limitados a promotores semente-específicos (WO8903887), promotores específicos de órgãos primordiais (como mencionado no pedido de patente US20030175783, An et al., 1996 The Plant Cell 8, 15-30), promotores específicos de caule (como mencionado no pedido de patente US20030175783, Keller et al., 1988 EMBO J. 7: 3625-3633), promotores específicos de folhas (como mencionado no pedido de patente US20030175783, Hudspeth et al., 1989 Plant Mol Biol 12:579-589), promotores específicos de mesófilo, promotores específicos de raiz (como mencionado no pedido de patente US20030175783, Keller et al., 1989 Genes Devel. 3:1639-1646), promotores específicos de tubérculos (como mencionado no pedido de patente US20030175783, Keil et al., 1989 EMBO J. 8: 1323:1330), promotores específicos de tecidos vasculares (como mencionado no pedido de patente US20030175783, Peleman et al., 1989 Gene 84: 359-369), promotores específicos de estames (WO8910396, WO9213956), promotores específicos da zona de deiscência (WO9713865); e semelhantes. Além dos promotores específicos, existem outros promotores de planta endógenos que incluem, mas não estão limitados ao promotor da subunidade pequena da ribulose 1,6 bifosfato (RUBP), promotor da beta-conglicinina, promotor da beta-faseolina, promotor da γ -kafirina, beta-amilase, álcool desidrogenase de milho, cruciferina (semente-específico), rubisco, gene RD2 de tabaco, gene SAG de Arabidopsis (folhas), poligalacturonase (frutos), patatina (tubérculos), hordeína de cevada, napina, actina de arroz, promotor da ubiquitina de milho, promotor do ADH, promotor do GPAL2, promotor do GPAL3, promotores das proteínas de choque térmico entre outros. Para a expressão das proteínas da seda em

plantas que produzem fibras tais como o algodão, o sisal, o esparto, a palma, a juta, a cana, o bambu, o agave, o cânhamo, entre outras, podem ser utilizados os promotores dos genes beta-tubulina, celulose sintetase A1, A2, A4, MYB (MYB- like transcription factor), entre outros (US6608242). A invenção preferencialmente inclui promotores de genes da fibra de algodão, que incluem mas não estão limitados aos promotores dos genes E6, H6S, Rac13, LTP, ACP, expansina, CAP, anexina, FbL2A e actina2.

[58] Em um dos aspectos da invenção, o promotor é um promotor expresso em animais. Como usado aqui, o termo “promotor expresso em animais” significa uma sequência de DNA que é capaz de iniciar e/ou controlar a transcrição em uma célula animal. Isso inclui qualquer promotor de origem animal; qualquer promotor de origem não animal que seja capaz de direcionar a expressão em uma célula animal, por exemplo o promotor da beta-caseína do leite (Invitrogen). Os promotores preferidos da invenção dirigem a transcrição de uma proteína em células produtoras de leite, tais como, mas não limitado a, promotores dos seguintes genes: proteína ácida do soro de leite (WAP), alfa caseínaS1, alfa caseína S2, beta caseína, kappa caseína, beta lactoglobina, alfa lactalbumina, entre outros. Outros promotores preferidos da invenção dirigem a transcrição de uma proteína nas células produtoras de urina (por exemplo uma célula uroepitelial ou uma célula da mesma natureza); tais promotores incluem mas não são limitados ao promotor do gene uroplakina. Ainda outros promotores preferidos da invenção dirigem a transcrição de uma proteína em uma célula embrionária.

[59] Além dos promotores descritos, uma das formas de concretização da presente invenção diz respeito aos promotores expressos em bactérias, fungos e insetos. Como usado aqui, o termo “promotor expresso em bactérias” significa uma sequência de DNA que é capaz de iniciar e/ou controlar a transcrição em uma célula bacteriana. Como usado aqui, o termo “promotor expresso em fungos” significa uma sequência de DNA que é capaz de iniciar e/ou controlar a transcrição em uma célula de fungo. Como usado aqui, o termo “promotor expresso em insetos” significa uma sequência de DNA que é capaz de iniciar e/ou controlar a transcrição em uma célula de inseto.

[60] Uma “sequência líder” ou “sequência sinal” na presente invenção significa uma sequência de ácido nucleico que, quando operacionalmente ligada a uma molécula de

ácido nucleico, permite a secreção do produto da molécula de ácido nucleico. A sequência líder está preferencialmente localizada na região 5' da molécula de ácido nucleico. Preferencialmente, a sequência líder é obtida do mesmo gene que o promotor utilizado para dirigir a transcrição da molécula de ácido nucleico, ou é obtida do gene onde a molécula de ácido nucleico é derivada. Preferencialmente a presente invenção utiliza a sequência sinal da α -coixina.

[61] O sinal de terminação da transcrição e a região de poliadenilação da presente invenção inclui, mas não está limitado a, sinal de terminação de SV40, sinal de adenilação de HSV TK, sinal de terminação do gene da nopalina sintetase de *Agrobacterium tumefaciens* (NOS), sinal de terminação do gene da octopina sintetase, sinal de terminação do gene 19S e 35S do CaMV, sinal de terminação do gene da álcool desidrogenase de milho, sinal de terminação do gene da manopina sintetase, sinal de terminação do gene da beta-faseolina, sinal de terminação do gene da ssRUBISCO, sinal de terminação do gene da sucrose sintetase, sinal de terminação do vírus que ataca o *Trifolium subterranean* (SCSV), sinal de terminação do gene trpC de *Aspergillus nidulans*, e outros semelhantes.

[62] Conforme descrito anteriormente, a expressão “vetores de expressão” pode compreender um promotor induzível operacionalmente ligado a uma sequência de ácido nucleico codificando uma proteína da teia de aranha. Promotores “induzíveis” podem dirigir a expressão de um polinucleotídeo com o qual eles estejam operacionalmente ligados, em um tecido ou estágio específico do desenvolvimento ou respondendo a condições ambientais. Em um dos aspectos da invenção, vetores de expressão compreendem um promotor induzível firmemente regulado operacionalmente ligado à uma molécula de ácido nucleico codificando uma proteína de teia de aranha. Tal vetor de expressão pode adicionalmente compreender um gene marcador de seleção (por exemplo, um gene codificando uma proteína que confere resistência a antibiótico) operacionalmente ligado a um promotor constitutivo ou a um promotor induzível firmemente regulado. Dependendo da aplicação, ele pode beneficiar a expressão da sequência de ácido nucleico codificando uma proteína da teia de aranha através de um promotor induzível de patógeno. Tais promotores incluem aqueles promotores derivados

de proteínas relacionadas à patogênese (proteínas PR) as quais são induzidas através de infecções por um patógeno, pex, proteínas PR, proteínas SAR, glucanase beta 1,3, quitinase, etc. Em um aspecto da presente invenção pode ser vantajoso utilizar promotores que são expressos localmente ou próximo do sítio de infecção do patógeno. Adicionalmente, porque muitos patógenos entram nas plantas através dos ferimentos, muitas vezes resultantes de danos de insetos, um promotor induzível por ferimentos pode ser usado em vetores de expressão da invenção. Promotores induzíveis de ferimentos incluem, mas não estão limitados a, promotor do gene inibidor da proteinase de batata (pinII), promotores dos genes win 1 e win 2, promotor do gene sistemina, promotor do gene MPI.

[63] A atividade transcricional dos promotores induzíveis pode também ser regulada por vários fatores ambientais, incluindo, mas não limitado a, temperatura, estresse anaeróbico, luz. Exemplos de promotores induzíveis incluem o promotor Adh1 (induzido por hipoxia ou estresse a frio), o promotor Hsp70 (induzido pelo estresse de calor) e o promotor PPKK (induzido pela luz).

[64] Como usado aqui, o termo “variante” ou “substancialmente similar” compreende seqüências de aminoácidos ou nucleotídeos diferentes de seqüências especificamente identificadas, em que um ou mais nucleotídeos ou resíduos de aminoácidos é deletado, substituído ou adicionado. As variantes podem ser variantes alélicas, de ocorrência natural, ou variantes de ocorrência não natural. As seqüências variantes ou substancialmente similares dizem respeito a fragmentos de ácidos nucléicos ou peptídeos que podem ser caracterizados pela porcentagem de identidade de suas seqüências de nucleotídeos ou aminoácidos com as seqüências de nucleotídeo (SEQ ID Nos 1-19) ou de aminoácidos (SEQ ID Nos 20-38) descritas aqui, como determinada por algoritmos comuns empregados no estado da técnica. Os fragmentos de ácidos nucléicos ou peptídeos preferidos são aqueles cujas seqüências de nucleotídeos ou aminoácidos têm pelo menos cerca de 40 ou 45% de identidade de seqüência, preferencialmente cerca de 50% ou 55% de identidade de seqüência, mais preferencialmente cerca de 60% ou 65% de identidade de seqüência, mais preferencialmente cerca de 70% ou 75% de identidade de seqüência, mais preferencialmente cerca de 80% ou 85% de identidade de seqüência,

mais preferencialmente ainda com cerca de 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% ou 99% de identidade de sequência quando comparada com a sequência de referência. A identidade percentual é determinada pelo alinhamento de duas seqüências a serem comparadas, determinando o número de resíduos idênticos na porção alinhada, dividindo este número pelo número total de resíduos na seqüência pesquisada e multiplicando o resultado por 100. Esse alinhamento pode ser feito através de ferramentas de domínio público, uma delas é o BLASTN, que está disponível na página do National Center for Biotechnology Information/NCBI (www.ncbi.nlm.nih.gov). O alinhamento de seqüência e o cálculo de porcentagem de identidade da presente invenção foram realizados e as seqüências depositadas no Gene Bank, através da integração do Web browser.

[65] O termo “hibridizando especificamente” diz respeito a associação entre duas moléculas de ácidos nucleicos de cadeia simples que possuam seqüências suficientemente complementares para permitir tal hibridização sob condições pré-determinadas geralmente descritas no estado da técnica (às vezes referida na presente invenção como “substancialmente complementar”). Em particular, o termo se refere à hibridização de um oligonucleotídeo com uma seqüência substancialmente complementar contendo uma molécula de DNA ou RNA de cadeia simples da presente invenção. Condições apropriadas necessárias para realização da hibridização específica entre moléculas de ácidos nucleicos de cadeia simples de complementariedade variada estão bem descritas no estado da técnica. Uma fórmula comum para se calcular as condições de estringência requeridas para se ter uma hibridização entre moléculas de ácido nucleico segue abaixo (Sambrook et al., Molecular Cloning, A Laboratory Manual, 2nd ed. (1989), Cold Spring Harbor Laboratory Press):

- a. $T_m = 81,5^{\circ}\text{C} + 16,6 \log [\text{Na}^+] + 0,41 (\% \text{G+C}) - 0,63 (\% \text{formamida}) - 600/\text{pb no duplex (sonda)}$
- b. Como ilustração da fórmula acima, usando $[\text{Na}^+] = [0,368]$ e 50% de formamida, com conteúdo de GC de 42% e um tamanho de sonda médio de 200 bases, a T_m será de 57°C .

[66] O termo “oligonucleotídeo” é referido aqui como ‘primers’ e ‘sondas’ da presente invenção, e é definido como uma molécula de ácido nucleico compreendendo dois ou mais ribo ou deoxiribonucleotídeos, preferencialmente mais do que três. O tamanho exato dos oligonucleotídeos dependerá de vários fatores e da aplicação particular e uso dos oligonucleotídeos. Oligonucleotídeos preferidos compreendem 15-50 bases consecutivas.

[67] O termo “sonda” utilizado na presente invenção refere-se a um oligonucleotídeo, polinucleotídeo ou ácido nucleico, sendo RNA ou DNA, se ocorrendo naturalmente como em uma digestão de enzima de restrição purificada ou produzida sinteticamente, que seja capaz de se anelar com ou especificamente hibridizando com um ácido nucleico contendo seqüências complementares à sonda. Uma sonda pode ser ainda de cadeia simples ou cadeia dupla. O comprimento exato da sonda dependerá de muitos fatores, incluindo temperatura, origem da sonda e uso do método. Por exemplo, dependendo da complexidade da seqüência alvo, a sonda de oligonucleotídeo tipicamente contém 15-25 ou mais nucleotídeos, embora ela possa conter menos nucleotídeos. As sonda aqui são selecionadas para serem complementares para diferenciar cadeias de uma seqüência de um ácido nucleico particular. Isto significa que a sonda pode ser suficientemente complementar para ser capaz de “hibridizar especificamente” ou anelar com suas respectivas cadeias-alvo sob uma série de condições pré-determinadas. Conseqüentemente, a seqüência da sonda não necessita refletir exatamente a seqüência complementar do alvo. Por exemplo, um fragmento de nucleotídeo não complementar pode ser ligado ao final 5’ ou 3’ da sonda, com o restante da seqüência da sonda sendo complementar à cadeia alvo. Alternativamente, bases não complementares ou seqüências longas podem estar intercaladas dentro da sonda se esta tiver complementariedade suficiente com a seqüência do ácido nucleico alvo para anelar especificamente com ele.

[68] O termo “primer” como utilizado aqui refere-se a um oligonucleotídeo, sendo RNA ou DNA, cadeia simples ou cadeia dupla, derivado de um sistema biológico, gerado através de digestão com enzimas de restrição, ou produzido sinteticamente que, quando colocado em um ambiente próprio, é capaz de agir funcionalmente como um iniciador de uma síntese de ácido nucleico dependente de molde. Quando apresentado com um molde de ácido nucleico apropriado, trifosfatos nucleosídeos adequados precursores de ácidos

nucleicos, uma enzima polimerase, cofatores adequados e condições tais como temperatura e pH adequados, o primer pode ser estendido em seu terminal 3' pela adição de nucleotídeos pela ação de uma polimerase ou com atividade similar para produzir uma primeira extensão do produto. O 'primer' pode variar em comprimento dependendo de condições particulares e requerimentos para aplicação. Por exemplo, em aplicações de diagnósticos, o 'primer' oligonucleotídeo possui tipicamente de 15-25 ou mais nucleotídeos em comprimento. O 'primer' deve ter complementariedade suficiente com o molde desejado para iniciar a síntese da extensão do produto desejado. Isso não significa que a sequência do 'primer' deva representar um complemento exato do molde desejado. Por exemplo, uma sequência de nucleotídeo não complementar pode ser ligada ao final 5' de um 'primer' complementar. Alternativamente, bases não complementares podem estar intercaladas dentro da sequência de oligonucleotídeo do 'primer', desde que o 'primer' tenha complementariedade suficiente com a sequência da cadeia molde desejada para prover funcionalmente um complexo molde-primer para a síntese da extensão do produto. Descrever os primers utilizados na presente invenção na parte dos exemplos onde tiverem utilizado os mesmos (Reações de PCR, por exemplo).

[69] O termo "proteína isolada" ou "proteína isolada e purificada" é às vezes utilizado na presente invenção. Este termo refere-se a uma proteína produzida pela expressão de uma molécula de ácido nucleico isolada da presente invenção. Alternativamente, este termo pode referir-se a uma proteína que tem sido suficientemente separada de outras proteínas que ela poderia estar naturalmente associada, tal como ela existe na sua forma "substancialmente pura". O termo "isolado" não exclui misturas sintéticas ou artificiais com outros compostos ou materiais, ou a presença de impurezas que não interferem com a atividade fundamental da proteína, e que pode estar presente, por exemplo, em uma purificação incompleta, adição de estabilizadores, ou combinados dentro, por exemplo, em preparações imunogênicas ou preparações farmacêuticas aceitáveis. Preparações farmacêuticas aceitáveis podem ser utilizadas na produção de fibras e polímeros sintéticos, por exemplo, que possam ser incorporadas dentro de uma variedade de implementos medicinais, incluindo, mas não limitado a, suturas, cobertura de ferimentos, e implantes.

[70] O termo “veículo farmacêuticamente aceitável” refere-se a soluções nas quais uma proteína da teia de aranha ou uma sequência de ácido nucleico codificadora de uma proteína da teia de aranha possa ser mantida sem a alteração das propriedades funcionais da molécula da teia de aranha aqui descrita para uso farmacêutico. Para a administração em mamíferos, por exemplo, uma proteína da teia de aranha ou uma sequência de ácido nucleico codificando uma proteína da teia de aranha pode ser suspensa em qualquer veículo farmacêutico aceitável, por exemplo o tampão salino “HEPES” com um pH de aproximadamente 7,8. Outros veículos farmacêuticamente aceitáveis que são úteis incluem, mas não estão limitados a, glicerol, água, solução salina, etano e outra solução salina farmacêuticamente aceitável tais como fosfatos e sais de ácidos orgânicos. Exemplos desses e outros veículos farmacêuticamente aceitáveis estão descritos em Remington’s Pharmaceutical Sciences (1991, Mack Publication Co., New Jersey).

[71] O termo “veículo agriculturalmente aceitável” refere-se a soluções nas quais uma proteína da teia de aranha ou uma sequência de ácido nucleico codificadora de uma proteína da teia de aranha possa ser mantida sem a alteração das propriedades funcionais da molécula da teia de aranha aqui descrita para uso agrícola. Os veículos utilizados na presente invenção podem ser líquidos ou sólidos. Veículos líquidos que podem ser utilizados para formar composições utilizando proteína recombinante da presente invenção podem ser, mas não estão limitados a água ou solventes orgânicos, tais como polióis, ésteres, cloreto de metileno, álcool, ou óleo vegetal. Outros componentes que podem ser incluídos na formulação incluem umectantes, preservativos, espessantes, agentes antimicrobianos, antioxidantes, emulsificadores, polímeros formadores de filme e misturas destes. Umectantes podem incluir polióis, açúcares (tais como melaço), glicóis e sais higroscópicos. Membranas vítreas, polímeros formadores de filmes incluem goma rosina, látex, pirrolidona polivinil, álcool polivinil, cloreto polivinil, polietileno, acetato de polivinil, e misturas destes. Aditivos opcionais adicionais incluem metil, metacrilato, e misturas destes.

[72] Os termos “proteína madura” ou “polipeptídeo maduro” significam um polipeptídeo possuindo uma sequência de aminoácidos após qualquer evento de processamento que normalmente ocorre com o polipeptídeo durante a sua geração, tal

como o processamento proteolítico de um precursor poliproteico. Na designação da seqüência ou limites da proteína madura, o primeiro aminoácido da seqüência da proteína madura é designado como resíduo de aminoácido 1. Para a presente invenção, qualquer resíduo de aminoácido associado com uma proteína madura não naturalmente encontrado associado com aquela proteína que precede o aminoácido 1 são designados aminoácido -1,-2,-3 e assim por diante. Para sistemas de expressão recombinantes, o codon iniciador da metionina é freqüentemente utilizado para propósitos de uma tradução eficiente. Este resíduo de metionina no polipeptídeo resultante, como usado aqui, deve estar na posição -1 na seqüência da proteína madura.

[73] O termo “análogo peptídico” significa um análogo natural ou mutante de uma proteína, compreendendo uma série de fragmentos lineares ou descontínuos daquela proteína e a qual pode ter um ou mais aminoácidos recolocado(s) com outro(s) aminoácido(s) e que pode ter sua atividade biológica alterada, auxiliada ou diminuída quando comparada com a proteína parental ou não-mutada.

[74] O termo “atividade biológica” diz respeito a uma função ou um grupo de funções executado por uma molécula em um contexto biológico (isto é, em um organismo ou substituto in vitro ou algum outro modelo similar). Para as proteínas da teia de aranha, atividade biológica é caracterizada pelas propriedades físicas (por exemplo, força tensora e elasticidade) conforme descritas aqui.

[75] O termo “substancialmente pura” refere-se a preparações compreendendo pelo menos 50-60% de peso do componente de interesse (por exemplo, ácido nucleico, oligonucleotídeo, polipeptídeo, proteína, etc). Mais preferencialmente, a preparação compreende pelo menos 75% de peso, e mais preferencialmente 90-99% de peso do componente de interesse. Pureza é medida através de métodos apropriados para o componente de interesse (por exemplo, métodos cromatográficos, gel de eletroforese de poliacrilamida ou agarose, análise de HPLC, espectrometria de massa e similares).

[76] O termo “vetor” diz respeito a um replicon, tal como plasmídeo, cosmídeo, bacmídeo, fago ou vírus, no qual outras seqüências genéticas ou elementos (seja DNA ou RNA) possam ser ligadas para serem replicadas juntas com o vetor. Preferencialmente o vetor derivado de vírus é selecionado entre os bacteriófagos, vaccínias, retrovírus ou

vírus do papiloma bovino. Um “vetor de expressão” é um vetor especializado que contém um gene com regiões regulatórias necessárias para a expressão em uma célula hospedeira. Tais vetores podem ser obtidos comercialmente, incluindo Clontech Laboratories, Inc (Palo Alto, Calif.), Stratagene (La Jolla, Calif), Invitrogen (Carlsbad, Calif.), New England Biolabs (Beverly, Mass.) e Promega (Madison, Wis.). Alguns exemplos de vetores utilizados na presente invenção, mas não limitados a pMAC/PS, pCMV-Gal, pGFP/NEO. O termo “operacionalmente ligado” significa que as seqüências regulatórias necessárias para expressão da seqüência codante são colocadas na molécula de DNA em posições apropriadas relativa à seqüência codante para efeito de expressão da seqüência codante. Essa mesma definição é às vezes aplicada para o arranjo de seqüências codantes e elementos controladores da transcrição (por exemplo, promotores, acentuadores ou “enhancers” e elementos de terminação) no vetor de expressão. Uma região codante exógena é tipicamente flanqueada por regiões regulatórias operacionalmente ligadas que regulam a expressão da região codante exógena em uma célula transformada (podendo ser microorganismo, vegetal ou animal). Uma região regulatória típica operacionalmente ligada a uma região codante exógena inclui um promotor, isto é, um fragmento de ácido nucleico que pode causar transcrição de regiões codantes exógenas, posicionado na região 5’ da região codante exógena. A presente invenção não está limitada para o uso de qualquer promotor particular e uma ampla variedade de promotores são conhecidos no estado da técnica. Os promotores podem ser, mas não estão limitados a, induzíveis, constitutivos e tecido-específicos.

[77] Em um dos aspectos da invenção, o promotor é um promotor constitutivo. Em outro aspecto da invenção, a atividade do promotor é estimulada por fatores externos ou internos tais como, mas não limitado a, hormônios, compostos químicos, impulsos mecânicos, e condições de estresse biótico ou abiótico. A atividade do promotor também pode ser regulada de maneira temporal e espacial (como por exemplo, promotores tecido-específicos e promotores regulados durante o desenvolvimento).

[78] O promotor pode conter elementos “enhancers”. Um “enhancer” é uma seqüência de DNA que pode estimular a atividade do promotor. Ela pode ser um elemento inato do promotor ou um elemento heterólogo inserido para aumentar o nível e/ou a tecido-

especificidade de um promotor. “Promotores constitutivos” referem-se àqueles que dirigem a expressão gênica em todos os tecidos e durante todo tempo. Promotores “tecido-específicos” ou “desenvolvimento-específicos” são aqueles que dirigem a expressão gênica quase que exclusivamente em tecidos específicos, tais como folhas, raízes, caules, flores, frutos ou sementes, ou em estágios do desenvolvimento específicos em um tecido, como no início ou final da embriogênese

[79] Em um dos aspectos da invenção, o promotor é um promotor expresso em plantas. Como usado aqui, o termo “promotor expresso em plantas” significa uma sequência de DNA que é capaz de iniciar e/ou controlar a transcrição em uma célula de planta. Isso inclui qualquer promotor de origem vegetal; qualquer promotor de origem não vegetal o qual seja capaz de direcionar a expressão em uma célula vegetal, por exemplo promotores de origem viral ou bacteriana tais como o 19S e 35S do CaMV (como mencionado no pedido de patente US20030175783, Hapster et al, 1988 Mol. Gen. Genet. 212, 182-190), promotor do bacteriófago T7 e promotores do gene presentes no T-DNA de *Agrobacterium*; promotores tecido-específicos ou órgão-específicos, incluindo mas não limitados a promotores semente-específicos (WO8903887), promotores específicos de órgãos primordiais (como mencionado no pedido de patente US20030175783, An et al., 1996 The Plant Cell 8, 15-30), promotores específicos de caule (como mencionado no pedido de patente US20030175783, Keller et al., 1988 EMBO J. 7: 3625-3633), promotores específicos de folhas (como mencionado no pedido de patente US20030175783, Hudspeth et al., 1989 Plant Mol Biol 12:579-589), promotores específicos de mesófilo, promotores específicos de raiz (como mencionado no pedido de patente US20030175783, Keller et al., 1989 Genes Devel. 3:1639-1646), promotores específicos de tubérculos (como mencionado no pedido de patente US20030175783, Keil et al., 1989 EMBO J. 8: 1323:1330), promotores específicos de tecidos vasculares (como mencionado no pedido de patente US20030175783, Peleman et al., 1989 Gene 84: 359-369), promotores específicos de estames (WO8910396, WO9213956), promotores específicos da zona de deiscência (WO9713865); e semelhantes. Além dos promotores específicos, existem outros promotores de planta endógenos que incluem, mas não estão limitados ao promotor da subunidade pequena da ribulose 1,6 bifosfato (RUBP),

promotor da beta-conglicinina, promotor da beta-faseolina, promotor da γ -kafirina, beta-amilase, álcool desidrogenase de milho, cruciferina (semente-específico), rubisco, gene RD2 de tabaco, gene SAG de Arabidopsis (folhas), poligalacturonase (frutos), patatina (tubérculos), hordeína de cevada, napina, actina de arroz, promotor da ubiquitina de milho, promotor do ADH, promotor do GPAL2, promotor do GPAL3, promotores das proteínas de choque térmico entre outros. Para a expressão das proteínas da seda em plantas que produzem fibras tais como o algodão, o sisal, o esparto, a palma, a juta, a cana, o bambu, o agave, o cânhamo, entre outras, podem ser utilizados os promotores dos genes beta-tubulina, celulose sintetase A1, A2, A4, MYB (MYB-like transcription factor), entre outros (US6608242). A invenção preferencialmente inclui promotores de genes da fibra de algodão, que incluem mas não estão limitados aos promotores dos genes E6, H6S, Rac13, LTP, ACP, expansina, CAP, anexina, FbL2A e actina2.

[80] Em um dos aspectos da invenção, o promotor é um promotor expresso em animais. Como usado aqui, o termo “promotor expresso em animais” significa uma sequência de DNA que é capaz de iniciar e/ou controlar a transcrição em uma célula de animal. Isso inclui qualquer promotor de origem animal; qualquer promotor de origem não animal o qual seja capaz de direcionar a expressão em uma célula animal, por exemplo o promotor da beta-caseína do leite (Invitrogen). Os promotores preferidos da invenção dirigem a transcrição de uma proteína em células produtoras de leite, tais como, mas não limitado a, promotores dos seguintes genes: proteína ácida do soro de leite (WAP), alfa caseína S1, alfa caseína S2, beta caseína, kappa caseína, beta lactoglobina, alfa lactalbumina, entre outros. Outros promotores preferidos da invenção dirigem a transcrição de uma proteína nas células produtoras de urina (por exemplo, uma célula uroepitelial ou uma célula da mesma natureza); tais promotores incluem mas não são limitados ao promotor do gene uroplakina. Ainda outros promotores preferidos da invenção dirigem a transcrição de uma proteína em uma célula embrionária.

[81] Além dos promotores descritos, uma das formas de concretização da presente invenção diz respeito aos promotores expressos em bactérias, fungos e insetos. Como usado aqui, o termo “promotor expresso em bactérias” significa uma sequência de DNA que é capaz de iniciar e/ou controlar a transcrição em uma célula de bactéria. Como

usado aqui, o termo “promotor expresso em fungos” significa uma sequência de DNA que é capaz de iniciar e/ou controlar a transcrição em uma célula de fungo. Como usado aqui, o termo “promotor expresso em insetos” significa uma sequência de DNA que é capaz de iniciar e/ou controlar a transcrição em uma célula de inseto. Uma “sequência líder” ou “sequência sinal” na presente invenção significa uma sequência de ácido nucleico que, quando operacionalmente ligada a uma molécula de ácido nucleico, permite a secreção do produto da molécula de ácido nucleico. A sequência líder está preferencialmente localizada na região 5’ da molécula de ácido nucleico. Preferencialmente, a sequência líder é obtida do mesmo gene que o promotor utilizado para dirigir a transcrição da molécula de ácido nucleico, ou é obtida do gene onde a molécula de ácido nucleico é derivada. Preferencialmente a presente invenção utiliza a sequência sinal da α -coixina.

[82] O sinal de terminação da transcrição e a região de poliadenilação da presente invenção inclui, mas não está limitado a, sinal de terminação de SV40, sinal de adenilação de HSV TK, sinal de terminação do gene da nopalina sintetase de *Agrobacterium tumefaciens* (NOS), sinal de terminação do gene da octopina sintetase, sinal de terminação do gene 19S e 35S do CaMV, sinal de terminação do gene da álcool desidrogenase de milho, sinal de terminação do gene da manopina sintetase, sinal de terminação do gene da beta-faseolina, sinal de terminação do gene da ssRUBISCO, sinal de terminação do gene da sucrose sintetase, sinal de terminação do vírus que ataca o *Trifolium subterranean* (SCSV), sinal de terminação do gene trpC de *Aspergillus nidulans*, e outros semelhantes.

[83] Conforme descrito anteriormente, a expressão “vetores de expressão” pode compreender um promotor induzível operacionalmente ligado a uma sequência de ácido nucleico codificando uma proteína da teia de aranha. Promotores “induzíveis” podem dirigir a expressão de um polinucleotídeo para o qual eles estejam operacionalmente ligados, em um tecido ou estágio específico do desenvolvimento ou respondendo a condições ambientais. Em um dos aspectos da invenção, vetores de expressão compreendem um promotor induzível firmemente regulado operacionalmente ligado a uma molécula de ácido nucleico codificando uma proteína de teia de aranha. Tal vetor de

expressão pode adicionalmente compreender um gene marcador de seleção (por exemplo, um gene codificando uma proteína que confere resistência a antibiótico) operacionalmente ligado a um promotor constitutivo ou a um promotor induzível firmemente regulado. Dependendo da aplicação, ele pode beneficiar a expressão da seqüência de ácido nucleico codificando uma proteína da teia de aranha através de um promotor induzível de patógeno. Tais promotores incluem aqueles promotores derivados de proteínas relacionadas à patogênese (proteínas PR) as quais são induzidas através de infecções por um patógeno, pex, proteínas PR, proteínas SAR, glucanase beta 1,3, quitinase, etc. Em um aspecto da presente invenção pode ser vantajoso utilizar promotores que são expressos localmente ou próximo do sítio de infecção do patógeno. Adicionalmente porque muitos patógenos entram nas plantas através dos ferimentos, muitas vezes resultantes de danos de insetos, um promotor induzível por ferimentos pode ser usado em vetores de expressão da invenção. Promotores induzíveis de ferimentos incluem, mas não estão limitados a, promotor do gene inibidor da proteinase de batata (pinII), promotores dos genes win 1 e win 2, promotor do gene sistemina, promotor do gene MPI.

[84] A atividade transcricional dos promotores induzíveis pode também ser regulada por vários fatores ambientais, incluindo, mas não limitado a, temperatura, estresse anaeróbico, luz. Exemplos de promotores induzíveis incluem o promotor Adh1 (induzido por hipoxia ou estresse a frio), o promotor Hsp70 (induzido pelo estresse de calor) e o promotor PPDK (induzido pela luz).

[85] A construção de vetores compreendendo promotores ligados à ácido nucleicos é conhecida no estado da técnica e pode ser encontradas em Sambrook, et al. (Molecular Cloning, A Laboratory Manual, 2nd ed. (1989), Cold Spring Harbor Laboratory Press).

[86] Vetores de expressão que compreendem seqüências de ácidos nucléicos codificando proteínas da teia de aranha estão dentro do escopo da presente invenção. Também estão incluídos na presente invenção as células vegetais, sementes recombinantes, embriões de plantas recombinantes, plantas recombinantes, células animais, embriões de animais recombinantes, animais recombinantes, células de insetos,

insetos recombinantes e microorganismos recombinantes compreendendo vetores de expressão codificando as proteínas da teia de aranha descritas aqui.

[87] Uma “célula transfectada” ou uma “célula transformada” significa uma célula dentro da qual tem sido introduzida, através de técnicas do DNA recombinante, uma molécula de ácido nucleico codificando um polipeptídeo da presente invenção. As células podem ser de um organismo hospedeiro que incluem, mas não estão limitados a, células bacterianas, células de fungos, células de insetos, célula de plantas e células de animais. Preferencialmente, a célula é uma célula eucariótica de um organismo pluricelular (pex plantas e animais).

[88] Os vetores de expressão podem ser introduzidos dentro do genoma da planta hospedeira desejada através de uma variedade de técnicas convencionais. Por exemplo, podem ser introduzidos diretamente dentro do DNA genômico da célula vegetal utilizando técnicas tais como eletroporação e microinjeção de protoplastos de células de plantas, ou o vetor de expressão pode ser introduzido diretamente no tecido vegetal utilizando-se métodos balísticos, tais como bombardeamento de partículas recobertas com DNA.

[89] Técnicas de microinjeção são conhecidas no estado da técnica e bem descritas em literatura científica e patentária. A introdução de vetores de expressão utilizando-se precipitações de polietileno glicol é descrita em Paszkowski et al. *Embo J.* 3:2717-2722, 1984 (como mencionado no pedido de patente US20020152501). Técnicas de eletroporação são descritas em From et al. *Proc. Natl. Acad. Sci. USA* 82:5824, 1985 (como mencionado no pedido de patente US20020152501). Técnicas de transformações balísticas são descritas em Klein et al. *Nature* 327:70-73, 1987 (como mencionado no pedido de patente US20020152501).

[90] Alternativamente, os vetores de expressão contendo a molécula de ácido nucleico recombinante podem ser combinados com regiões flangeadoras de T-DNA apropriadas e introduzidos dentro de vetor convencional hospedeiro *Agrobacterium tumefaciens*. A função de virulência do hospedeiro *Agrobacterium tumefaciens* direcionará a inserção das moléculas de ácido nucleico recombinante e marcador adjacente dentro do DNA da célula vegetal quando a célula é infectada pela bactéria. Técnicas de transformação

mediadas por *Agrobacterium tumefaciens*, incluindo desarmamento e o uso de vetores binários, são bem descritas na literatura científica (como mencionado no pedido de patente US 20020152501, Horsch et al. Science 233:496-498, 1984; e Fraley et al. Proc. Natl. Acad. Sci. USA 80:4803, 1983).

[91] Células de plantas transformadas que são derivadas de qualquer uma das técnicas de transformação descritas acima podem ser cultivadas para regenerar uma planta inteira que possua o genótipo transformado e então o fenótipo desejado tal como produção de proteínas da teia de aranha. Tais técnicas de regeneração contam com a manipulação de certos fitohormônios em meio de crescimento de cultura de tecidos, tipicamente contendo um marcador biocida e/ou herbicida, que deve ser introduzido junto com a sequência de nucleotídeos desejada. Preferencialmente a presente invenção utiliza os marcadores de seleção selecionados dos genes de resistência a antibióticos e herbicidas tais como canamicina, neomicina, ampicilina, cloranfenicol, estreptomicina, higromicina, geneticina, fosfinotricina, glifosato, glufosinato de amônio, entre outros. A presente invenção também se utiliza de genes repórteres para verificar o potencial de transformação das células, tais como AHAS, BAR, GUS. Regeneração de plantas a partir de cultura de protoplastos é descrita em Evans et al., Protoplasts Isolation and Culture, Handbook of Plant Cell Culture, pp. 124-176, MacMillan Publishing Company, New York, 1983; e Binding, Regeneration of Plants, Plant Protoplasts, pp. 21-73, CRC Press, Boca Raton, 1985 (como mencionado no pedido de patente US20020152501). A regeneração pode ser também obtida através de calos de planta, explantes, órgãos, ou parte da mesma. Tais técnicas de regeneração são descritas geralmente em Klee et al., Ann. Ver. Of Plant Phys. 38:467-486, 1987 1985 (como mencionado no pedido de patente US20020152501) e também podem ser encontradas em Clark, 1997 (Clark, M. S. eds., 1997. Plant Molecular Biology A laboratory Manual. Springer-Verlag, Berlin, Heidelberg); Maliga et al., 1995 (Maliga, P.; D. F. Flessing, A. R. Cashmore, W. Cruissem, J. E. Varner, eds., 1995. Methods in Plant Molecular Biology, A Laboratory Course Manual. Cold Spring Harbor Laboratory Press) e Martinez-Zapater & Salinas, 1998 (Martinez-Zapater, J.M. & J. Salinas, eds., 1998. Methods in Molecular Biology, v. 82: Arabidopsis Protocols. Humana Press, Totowa, N.J.).

[92] A metodologia para manutenção e crescimento de culturas de microorganismos (bactérias, fungos, leveduras) é bem conhecida por técnicos no assunto. A descrição de tais técnicas também pode ser encontrada em manuais da área como Gerhardt et al., 1994 (Gerhardt, P.; R.G.E. Murray; R.N. Costilow; E. W. Nester; W. A. Wood; N.R. Krieg & G.B. Phillips eds. *Manual of Methods for General Bacteriology*. American Society for Microbiology, Washington, D.C) ou Brock, 1989 (Brock, T.D. 1989. *Biotechnology: A Textbook of Industrial Microbiology*. Second edition, Sinauer Associates, Inc., Sunderland, Mass).

[93] Uma “célula embrionária” significa uma célula que é capaz de ser progenitora de todas as células de linhagem somática e germinativa de um organismo. Exemplos de células embrionárias são as células tronco (células ES) e oócitos fertilizados. Preferencialmente, as células embrionárias da invenção são células embrionárias de mamíferos.

[94] “Célula de linhagem germinativa” significa uma célula eucariota, progenitora, ou semelhante à células progenitoras, que é produto de uma divisão celular meiótica.

[95] Um “clone” ou uma “população celular clonal” é uma população de células derivadas de uma simples célula ou ancestral comum através da mitose.

[96] Uma “linhagem celular” é um clone de uma célula primária ou população celular que é capaz de crescer estavelmente in vitro por muitas gerações.

[97] “Plantas” referem-se a organismos fotossintéticos, ambos eucariotos e procariotos, onde o termo “plantas desenvolvidas” refere-se a plantas eucariotas. Os ácidos nucleicos da invenção podem ser utilizados para conferir tratos desejados em essencialmente qualquer planta. Então, a invenção possui uso sobre várias espécies de plantas, incluindo espécies dos gêneros *Anacardium*, *Anona*, *Arachis*, *Artocarpus*, *Asparagus*, *Atropa*, *Avena*, *Brassica*, *Carica*, *Citrus*, *Citrullus*, *Capsicum*, *Carthamus*, *Cocos*, *Coffea*, *Cucumis*, *Cucurbita*, *Daucus*, *Elaeis*, *Fragaria*, *Glycine*, *Gossypium*, *Helianthus*, *Heterocallis*, *Hordeum*, *Hyoseyamus*, *Lactuca*, *Linum*, *Lolium*, *Lupinus*, *Lycopersicon*, *Malus*, *Manihot*, *Majorana*, *Medicago*, *Nicotiana*, *Olea*, *Oryza*, *Panicum*, *Pannesetum*, *Passiflora*, *Persea*, *Phaseolus*, *Pistachia*, *Pisum*, *Pyrus*, *Prunus*, *Psidium*,

Raphanus, Ricinus, Secale, Senecio, Sinapis, Solanum, Sorghum, Theobromus, Trigonella, Triticum, Vicia, Vitis, Vigna, e Zea.

[98] “Animais” referem-se a organismos eucariotos que podem pertencer ao filo dos vertebrados ou invertebrados, onde o termo “animais superiores” refere-se aos animais do filo vertebrado. Os ácidos nucleicos da invenção podem ser utilizados para conferir tratos desejados em essencialmente qualquer animal. Então, a invenção possui uso sobre várias espécies de animais vertebrados, incluindo espécies de mamíferos, que incluem mas não estão limitados aos primatas, cetáceos, insetívoros, dermópteros, quirópteros, roedores, lagomorfos, carnívoros, perssodáctilos, hiracóides, prosboscídeos, artiodáctilos, desdentados, folídotas, tubilidentados, sirênios, marsupiais e os monotremados. Preferencialmente a presente invenção diz respeito a utilização dos ácidos nucleicos da presente invenção em espécies de mamíferos que incluem, mas não estão limitadas aos grupos de Afrotheria, Euarchontoglires, Laurasiatheria e Xenarthra. Preferencialmente a presente invenção diz respeito aos mamíferos selecionados entre camundongos, bovinos, ovinos, caprinos e eqüinos.

[99] O termo “animais inferiores” refere-se aos animais do filo dos invertebrados. A invenção também possui uso sobre várias espécies de animais invertebrados, incluindo espécies de aracnídeos, que incluem mas não estão limitados a Acari, Amblypygi, Araneae, Opiliones, Palpigradi, Pseudoscorpiones, Ricinulei, Scorpiones, Solifugae, Uropygi. Preferencialmente a presente invenção diz respeito a utilização dos ácidos nucleicos da presente invenção em espécies de aranhas que incluem, mas não estão limitadas aos grupos de Araneomorphae, Mesothelae e Mygalomorphae.

a. “Microorganismos” referem-se a organismos microscópicos tais como bactérias, vírus, fungos e protozoários. Preferencialmente a presente invenção tem como microorganismos os organismos escolhidos do grupo das bactérias e fungos. “Bactérias” referem-se a organismos procariotos, com exceção das cianofíceas. Os ácidos nucleicos da invenção podem ser utilizados para conferir tratos desejados em essencialmente qualquer bactéria. Então, a invenção possui uso sobre várias espécies de bactérias, que incluem mas não estão limitadas ao grupo das Actinobacteria, Aquificae, Bacteroidetes/grupo Chlorobi, Chlamydiae/ grupo Verrucomicrobia, Chloroflexi,

Chrysiogenetes, Cyanobacteria, Gloeobacteria, Nostocales, Oscillatoriales, Pleurocapsales, Prochlorales, Stigonematales, Deferribacteres, Deinococcus-Thermus, Dictyoglomi, Fibrobacteres/grupo Acidobacteria, Firmicutes, Fusobacteria, Gemmatimonadetes, Nitrospirae, Planctomycetes, Proteobacteria, Spirochaetes, Thermodesulfobacteria, Thermotogae. “Fungos” referem-se a organismos do reino Fungi que podem ser unicelulares ou pluricelulares. Os ácidos nucleicos da invenção podem ser utilizados para conferir tratos desejados em essencialmente qualquer fungo. Então, a invenção possui uso sobre várias espécies de fungos, que incluem mas não estão limitadas ao grupo dos Ascomycota e Basidiomycota. Preferencialmente a presente invenção utiliza espécies de microorganismos selecionadas dos gêneros *Aspergillus*, *Bacillus*, *Escherichia*, *Pichia*, *Saccharomyces* ou *Streptomyces*.

[100] Uma “resposta imune” significa qualquer reação produzida na resposta a um antígeno, tal como um antígeno viral, em um hospedeiro tendo um sistema imune funcionando. Respostas imunes podem ser humoral “in nature”, isto é, que envolve a produção de imunoglobulinas ou anticorpos; ou celular “in nature”, que envolve vários tipos de linfócito “B” e “T”, células dendríticas, macrófagos, células apresentadoras de antígenos e semelhantes, ou ambas as respostas. A resposta imune também pode envolver a produção ou elaboração de várias moléculas efetoras tais como citocininas, limfocinas, e semelhantes. Respostas imunes podem ser mensuradas in vitro, células ou sistemas animais. Tais respostas imunes podem ser importantes na proteção do hospedeiro de doenças e podem ser usadas profilaticamente e terapêuticamente.

[101] Um “derivado” de uma proteína da teia de aranha ou um fragmento dele significa um polipeptídeo modificado pela variação na sequência de aminoácido da proteína., por exemplo pela manipulação de ácidos nucleicos codificando a proteína ou pela alteração da proteína por ela mesma. Tais derivações da sequência de aminoácido natural podem envolver inserção, adição, deleção ou substituição de um ou mais aminoácidos, podendo ou não alterar a atividade essencial da proteína da teia de aranha.

[102] O termo “proteína nativa ou natural da teia de aranha” refere-se àquelas proteínas que estão presentes nas teias produzidas pelas aranhas. Essas proteínas podem ser derivadas da teia por ela mesma através de dissolução ou da glândula específica da teia

localizada no abdomen da aranha antes da seda ser tecida. O termo também pode ser aplicado a proteínas da teia de aranha produzidas usando uma variedade de sistemas de expressão, os quais compreendem substancialmente a mesma seqüência de aminoácido conforme produzida pela aranha.

[103] O termo “proteína sintética da teia de aranha” refere-se a uma proteína que tem sido produzida por um sistema de expressão e cuja seqüência pode ser baseada em uma seqüência de proteína da teia de aranha natural ou uma seqüência de ácido nucleico produzida artificialmente que codifica motifs de aminoácidos das proteínas da teia de aranha.

[104] O termo “biofilamento” significa uma proteína fibrosa que é normalmente produzida e secretada por qualquer uma das variedades de insetos e aracnídeos. Biofilamentos são compostos de alternância de regiões cristalinas e amorfas. Exemplos de biofilamentos incluem teia de aranha, uma secreção proteica fibrosa que foi tecida externamente encontrada em uma variedade de aracnídeos (por exemplo, *Nephilengys cruentata*, *Avicularia juruensis* e *Parawixia bistriata*), e fibroína, uma secreção proteica fibrosa fiada externamente e encontrada em uma variedade de insetos (por exemplo *Bombyx mori*). Preferivelmente o biofilamento, quando secretado na forma de secreção sujeita à força do tear e extensão mecânica, tem um segmento de polialanina, formando um domínio de cristal, que passa por uma transição de hélice para folha beta, formando um cristal (beta) que estabiliza essa estrutura. Preferencialmente, o domínio amorfo de um biofilamento forma uma folha do tipo beta onde os espaços entre as folhas são entre 3 ângstrons e 8 ângstrons, preferencialmente entre 3,5 e 7,5 ângstrons.

[105] Preferencialmente, o biofilamento tem uma porção C-terminal com um motivo de aminoácidos repetidos que possui entre 20-40 aminoácidos em comprimento, mais preferencialmente 34 aminoácidos em comprimentos, e uma seqüência consenso que está entre 35 e 55 aminoácidos em comprimento, mais preferencialmente 47 aminoácidos em comprimento. Preferencialmente, o biofilamento tem um motivo de aminoácidos repetido (criando ambos os domínios: amorfo e cristalino) tendo uma seqüência que é pelo menos 50% idêntica as seqüências selecionadas do grupo: SEQ ID NO 19-34), mais preferencialmente pelo menos 70% idêntica, e mais preferencialmente pelo menos 90%

idêntica às seqüências identificadas como SEQ ID NO 19-34. “Meio de cultura” significa o meio que circunda a célula e que seja responsável pela sobrevivência da mesma. Se a célula está secretando uma proteína (pex, um biofilamento), o meio de cultura da célula irá conter a proteína secretada por esta célula.

[106] A descoberta de novas proteínas da seda de aranhas, bem como sua caracterização e expressão em diferentes sistemas heterólogos será de grande utilidade para inúmeras áreas, como a médica e industrial.

[107] As proteínas da seda de aranha podem ser obtidas através de polipeptídeos sintéticos tendo seqüências de aminoácidos substancialmente similar a uma unidade consenso de proteína da seda ou através de polipeptídeos expressos de seqüências de ácido nucleico codificando uma proteína da seda natural ou engenheirada, ou derivados destes. Dependendo da aplicação para a qual a proteína da seda é requerida, pode ser desejável formar fibras de uma simples proteína da teia de aranha ou de combinações de diferentes proteínas da teia de aranha.

[108] De acordo com um aspecto da invenção são providas seqüências de ácido nucleicos codificando novas proteínas de teia de aranha. As seqüências de ácido nucleico da presente invenção compreendem SEQ ID NOS 1-19.

[109] Em um aspecto particular da invenção, são providas seqüências de ácidos nucleicos que codificam proteínas da seda relacionadas principalmente, mas não limitando-se à glândula Major Ampullate. Exemplos de seqüências de ácidos nucleicos que estão relacionadas com essa glândula compreendem SEQ ID NOS 3, 17 e 18.

a. Em um aspecto particular da invenção, são providas seqüências de ácidos nucleicos que codificam proteínas da seda relacionadas principalmente, mas não limitando-se à glândula Minor Ampullate. Exemplos de seqüências de ácidos nucleicos que estão relacionadas com essa glândula compreendem SEQ ID NOS 4, 5, 6 e 16.

[110] Em um aspecto particular da invenção, são providas seqüências de ácidos nucleicos que codificam proteínas da seda relacionadas principalmente, mas não limitando-se à glândula Flageliforme. Exemplos de seqüências de ácidos nucleicos que estão relacionadas com essa glândula compreendem SEQ ID NOS 1 e 15.

[111] Em um aspecto particular da invenção, são providas seqüências de ácidos nucleicos que codificam proteínas da seda relacionadas principalmente, mas não limitando-se à glândula Tubuliforme. Exemplo de seqüência de ácido nucleico que está relacionada com essa glândula compreende a SEQ ID NO 2.

[112] Em um aspecto particular da invenção, são providas seqüências de ácidos nucleicos que codificam proteínas da seda relacionadas principalmente, mas não limitando-se à glândula Aciniforme. Exemplo de seqüência de ácido nucleico que está relacionada com essa glândula compreende a SEQ ID NO 14.

[113] As moléculas de ácido nucleico codificando polipeptídeos da presente invenção podem ser preparadas através de dois métodos gerais: síntese artificial de nucleotídeos que codificam para a proteína da teia de aranha ou através do isolamento de nucleotídeos oriundos das próprias aranhas. Ambos os métodos utilizam protocolos bem descritos no estado da técnica. A informação da seqüência de nucleotídeo, tais como seqüências de DNA codificando uma proteína da teia de aranha sintética ou natural, pode ser preparada através de uma molécula isolada de ácido nucleico da invenção pela síntese de oligonucleotídeo. Síntese de oligonucleotídeos pode ser preparada pelo método fosforamidite empregado pelo Sintetizador de DNA da Applied Biosystems 38A ou aparelhos similares. O constructo resultante pode ser utilizado diretamente ou purificado de acordo com métodos utilizados no estado da técnica, tais como a cromatografia líquida (HPLC).

[114] De acordo com a presente invenção, ácidos nucleicos tendo um nível apropriado de homologia de seqüência com seqüências codificando uma proteína da teia de aranha podem ser identificados através da utilização de condições de hibridização e lavagem de estringência apropriada. Tais métodos são úteis para uma variedade de propósitos, incluindo o triagem de bibliotecas compreendendo seqüências mutadas de ácido nucleico codificando uma proteína da teia de aranha. Hibridizações podem ser feitas através da utilização de metodologias descritas em Sambrook, et al. (Molecular Cloning, A Laboratory Manual, 2nd ed. (1989), Cold Spring Harbor Laboratory Press).

[115] A invenção também relata moléculas de novas proteínas da seda de aranhas. Em um aspecto particular da invenção, as proteínas da seda de aranha compreendem as seqüências SEQ ID NOs 20-38

[116] Em um aspecto particular da invenção, são providas seqüências de aminoácidos da seda relacionadas principalmente, mas não limitando-se à glândula Major Ampullate. Exemplos de seqüências de aminoácidos que estão relacionadas com essa glândula compreendem as seqüências SEQ ID NOs 22, 35 e 36.

[117] Em um aspecto particular da invenção, são providas seqüências de aminoácidos relacionadas principalmente, mas não limitando-se à glândula Minor Ampullate. Exemplos de seqüências de aminoácidos que estão relacionadas com essa glândula compreendem as seqüências SEQ ID NOs 23, 24, 25 e 37.

[118] Em um aspecto particular da invenção, são providas seqüências de aminoácidos relacionadas principalmente, mas não limitando-se à glândula Flagelliforme. Exemplos de seqüências de aminoácidos que estão relacionadas com essa glândula compreendem as seqüências SEQ ID NOs 20 e 34.

[119] Em um aspecto particular da invenção, são providas seqüências de aminoácidos relacionadas principalmente, mas não limitando-se à glândula Tubuliforme. Exemplo de seqüência de aminoácido que está relacionada com essa glândula compreende a SEQ ID NO 21.

[120] Em um aspecto particular da invenção, são providas seqüências de aminoácidos relacionadas principalmente, mas não limitando-se à glândula Aciniforme. Exemplo de seqüência de aminoácido que está relacionada com essa glândula compreende a SEQ ID NO 33.

[121] De acordo com um outro aspecto da presente invenção, uma molécula de ácido nucleico isolada é provida, a qual tem uma seqüência selecionada do grupo de SEQ ID NOs: 1-19 e cuja expressão é controlada através de promotores específicos ou constitutivos e terminadores possuindo a região de poliadenilação.

[122] A presente invenção discorre sobre um método para produzir biofilamentos a partir das proteínas da seda de aranhas produzidas em células procariotas e eucariotas.

[123] Em um outro aspecto da invenção, células hospedeiras compreendendo pelo menos uma das proteínas da seda de aranha codificadas por ácidos nucleicos são providas. Tais células hospedeiras incluem mas não estão limitadas a células bacterianas, células de fungos, células de insetos, células de mamíferos e células vegetais. Células hospedeiras superexpressando uma ou mais proteínas da seda de aranha codificadas por ácidos nucleicos da invenção provém reagentes úteis para inúmeras aplicações, incluindo, mas não limitado a, produção de fibras de seda compreendendo pelo menos uma proteína da seda que pode ser incorporada dentro de um material para modular as propriedades estruturais do material.

- a. A invenção também discorre sobre composições dermatológicas caracterizadas por apresentarem:
- b. Proteína recombinante de seda de aranha;
- c. Veículo farmacêuticamente aceitável.

[124] E ainda discorre sobre composições microbidas caracterizadas por apresentarem:

- a. Proteína recombinante de seda de aranha;
- b. Veículo agriculturalmente aceitável e, opcionalmente,
- c. Aditivos

[125] Um outro objeto da presente invenção é prover células procariotas, e organismos procariotos contendo moléculas de DNA da presente invenção que podem ser qualquer uma das seqüências identificadas do grupo SEQ ID NO 1-19, ou células contendo as construções gênicas capazes de produzir as proteínas da presente invenção (SEQ ID NO 20-38), ou variantes das mesmas. As construções gênicas podem estar estavelmente integradas no genoma das células de organismos procariotos.

[126] Um outro objeto da presente invenção é prover células eucariotas, e organismos eucariotos contendo moléculas de DNA da presente invenção que podem ser qualquer uma das seqüências identificadas do grupo SEQ ID NO 1-19, ou células contendo as construções gênicas capazes de produzir as proteínas da presente invenção (SEQ ID NO

20-38), ou variantes das mesmas. As construções gênicas podem estar estavelmente integradas no genoma das células de organismos eucariotos.

[127] Em outro aspecto da invenção, as construções gênicas podem estar providas em uma molécula de DNA capaz de se replicar de forma autônoma nas células de organismos eucariotos, tais como vetores virais. A construção gênica pode também estar disposta de forma transitória nas células de organismos eucariotos.

[128] A presente invenção também diz respeito a um método para produzir um organismo geneticamente modificado caracterizado pelo fato de compreender as seguintes etapas:

- a. transformar uma célula, tecido, órgão ou embrião com uma construção gênica de acordo com qualquer uma das reivindicações 3 a 11 ou um vetor de expressão de acordo com qualquer uma das reivindicações 12 a 23;
- b. selecionar células transformadas, calos de células, embriões ou sementes;
- c. regenerar plantas maduras, embriões maduros ou microorganismos de células transformadas, calos de células, embriões ou sementes selecionados na etapa (b);
- d. selecionar plantas maduras, embriões maduros ou células de microorganismos da etapa (c) contendo a construção gênica ou vetor de expressão com as seqüências de nucleotídeos que codificam a proteína da seda da aranha.

[129] A invenção também relata um método para produção de proteína recombinante caracterizado pelo fato de compreender as seguintes etapas:

- a. transformar uma célula, tecido, órgão ou embrião com um vetor de expressão de acordo com qualquer uma das reivindicações 12 a 23;
- b. selecionar células transformadas, calos de células, embriões ou sementes;
- c. regenerar plantas maduras, embriões maduros ou microorganismos de células transformadas, calos de células, embriões ou sementes selecionados na etapa (b);
- d. selecionar plantas maduras, embriões maduros ou células de microorganismos da etapa (c) contendo vetor de expressão com as seqüências de nucleotídeos que codificam a proteína da seda da aranha;

- e. Fazer a extração da proteína recombinante da seda de aranha produzida nos organismos selecionados na etapa (d).

[130] A possibilidade de moléculas de ácido nucleico codificarem proteínas da teia de aranha torna viável a produção de grandes quantidades da proteína da teia de aranha em um sistema procariota ou eucariota viável. Por exemplo, parte ou o todo de pelo menos uma molécula de DNA codificando uma proteína da teia de aranha natural ou sintética, tais como uma seqüência de ácido nucleico selecionada do grupo de SEQ ID NO 1-19, pode ser inserida dentro de um vetor plasmidial adaptado para expressão em células de bactérias, tais como *E. coli*. Tais vetores compreendem elementos regulatórios necessários para a expressão do DNA na célula hospedeira posicionados de tal maneira a permitir a expressão do DNA na célula hospedeira. Tais elementos regulatórios requeridos para a expressão incluem seqüências de promotor, seqüências de iniciação da transcrição e, opcionalmente, seqüências acentuadoras (*enhancers*). Tais métodos podem ser usados para avaliar constructos para expressão das proteínas da teia de aranha, por exemplo, em um sistema bacteriano que proporciona uma rápida e real técnica de triagem.

[131] As proteínas da teia de aranha produzidas pela expressão gênica em um sistema recombinante procariota ou eucariota pode ser purificada de acordo com métodos conhecidos no estado da técnica. Preferencialmente, um sistema de secreção/expressão viável comercialmente pode ser usado, por meio do qual uma proteína recombinante é expressa e em seguida secretada da célula hospedeira, para ser facilmente purificada do meio de cultura. Se vetores de expressão/secreção não são usados, uma técnica alternativa envolve purificar a proteína recombinante das células lisadas derivadas de células procariotas ou eucariotas onde a proteína foi expressa. Métodos para gerar tais lisados de células são bem conhecidos no estado da técnica. Proteínas recombinantes podem ser purificadas através de separação de afinidade, tais como através de interação imunológica com anticorpos que ligam especificamente a proteínas recombinantes ou colunas de níquel para isolamento das proteínas recombinantes marcadas com 6-8 resíduos de histidina no N-terminal ou C-terminal. Marcações alternativas compreendem epitopo

FLAG ou epítopo hemaglutinina. Tais métodos são bem descritos no estado da técnica e são comumente usados por especialistas.

[132] Alternativamente, estratégias de purificação padrão desenhadas para isolar diferencialmente proteínas da seda de homogenados de plantas podem ser utilizadas como vantagem. Purificação de uma proteína da teia de aranha expressada em plantas pode ser facilitada por sua extrema estabilidade sob condições que denaturam proteínas típicas, tais como, por exemplo, alta temperatura e baixo pH. Estratégias de purificação de proteínas em geral podem ser adaptadas para otimizar a purificação da proteína da teia de aranha a partir de de folhas. Porções acima do solo de plantas transgênicas podem ser colhidas e secadas através de práticas normais. Essas partes das plantas desidratadas podem ser homogeneizadas em tampão apropriado seguido por vários tratamentos designados para eliminar contaminantes diferencialmente. As proteínas da seda recuperadas podem ser otimizadas seguindo tratamentos nos quais os extratos vegetais estão sujeito a uma ou mais combinações dos seguintes passos: 1) ferver na presença ou ausência de detergente; 2) centrifugação diferencial; 3) decréscimo progressivo de pH; e 4) precipitação com concentrações variadas de ureia ou sulfato de amônia. Esses passos podem variar de acordo com a otimização para a produção e eficiência de purificação das proteínas da teia de aranha em plantas.

[133] O nível da proteína da teia de aranha pode ser determinado por *imunoblotting* e a pureza e concentração através da análise dos aminoácidos. Proteína da teia de aranha purificada pode ser analisada através das suas propriedades mecânicas para assegurar que a proteína recombinante possui as características desejadas. As proteínas da teia de aranha, preparadas conforme descrito anteriormente, podem ser analisadas de acordo com procedimentos padrão. Por exemplo, tais proteínas podem ser sujeitas a análises de seqüência de aminoácidos de acordo com métodos conhecidos.

[134] As proteínas da teia de aranha da presente invenção podem ser utilizadas como microbiocidas contra replicação viral; como defensinas contra insetos e pragas; como cosméticos ou composição dermatológica; em combinação com outras matérias. Podem ainda ser introduzidas em plantas de algodão para serem expressas juntamente com a fibra de algodão para aumentar a resistência e flexibilidade da fibra. As proteínas da

presente invenção também estão relacionadas com a geração de novas variações em relação as sedas naturalmente produzidas por aranhas e produção de novas proteínas, peptídeos e polipeptídeos com diferentes características físico-químicas.

[135] A presente invenção é ainda definida nos seguintes exemplos. Deve ser entendido que esses exemplos, enquanto indicam parte da invenção, são colocados como forma de ilustração somente, não tendo, portanto, qualquer cunho limitante do escopo das presentes invenções.

[136] Técnicas usuais de biologia molecular tais como transformação de bactérias e eletroforese em gel de agarose de ácidos nucléicos são referidas através de termos comuns para descrevê-las. Detalhes da prática dessas técnicas, bem conhecidos no estado da técnica, são descritos em Sambrook, et al. (Molecular Cloning, A Laboratory Manual, 2nd ed. 1989, Cold Spring Harbor Laboratory Press). Várias soluções utilizadas nas manipulações experimentais são referidas por seus nomes comuns tais como “solução de lise”, “SSC”, “SDS”, etc. A composição dessas soluções pode ser encontrada na referência Sambrook, et al. (supracitada).

Exemplo 1: Coleta e classificação de aranhas e de suas teias

[137] Sedas das espécies de aranhas *Argiope* sp., *Ephebopus* sp., *Nephila clavipes*, *Nephilengys cruentata*, *Avicularia juruensis* e *Parawixia bistriata* foram coletadas na biodiversidade brasileira, principalmente na região amazônica, da mata atlântica e dos cerrados. As sedas foram secas em temperatura ambiente e analisadas através da microscopia de infravermelho (FTIR). Foram verificadas as significativas diferenças nos resultados das folhas alfa e beta, associadas a flexibilidade e resistência das sedas em relação a *Nephila clavipes*. O FTIR vem sendo usado como um método de determinação de estrutura secundária de proteínas em estado sólido e tem se mostrado bastante viável, principalmente para proteínas insolúveis como as das teias de aranha. Para a quantificação de estruturas secundárias de proteínas a partir de espectros de infravermelho foi utilizado o método de reconhecimento de padrões desenvolvido por

Forato et al., 1998. Os espectros da teia de aranha e produtos de expressão transgênicos foram obtidos com amostras preparadas em pastilhas de KBr, e foi usada para a quantificação banda de amida I, entre 1600 e 1800 cm⁻¹.

Tabela 01 – Porcentagem de estruturas secundárias encontradas na análise de infravermelho para as diferentes espécies de aranhas coletadas.

[138] Espécie	[139] α hélice (%)	[140] β - folha (%)	[141] Volta s (%)	[142] Outr os (%)
[143] <i>Argiope</i>	[144] 7	[145] 56	[146] 27	[147] 13
[148] <i>Nephila clavipes</i>	[149] 16	[150] 43	[151] 24	[152] 13
[153] <i>Epebopus</i>	[154] 2	[155] 57	[156] 37	[157] 13
[158] <i>Nephylengys cruentata</i>	[159] 6	[160] 53	[161] 32	[162] 11
[163] <i>Parawixia bistriata</i>	[164] 17	[165] 47	[166] 32	[167] 12
[168] <i>Avicularia juruensis</i>	[169] 5	[170] 58	[171] 32	[172] 12

Exemplo 2: Obtenção de Seqüências Polinucleotídicas: Construção das Bibliotecas de cDNA, Seqüenciamento

[173] Após a coleta, as glândulas produtoras de seda das aranhas foram isoladas em laboratório, imediatamente congeladas em nitrogênio líquido e mantidas a -70oC. Após pulverização, a extração de RNA Total foi realizada, utilizando-se o reagente TRIZOL (Invitrogen), de acordo com as recomendações do fabricante. O kit Oligotex (Qiagen) foi utilizado para a purificação do mRNA utilizado para síntese do cDNA, preferencialmente

através da utilização do Kit “SUPERScript II Plasmid System with GATEWAY Technology for cDNA Synthesis and Cloning” (Invitrogen), segundo as orientações fornecidas pelo fabricante. Após a síntese e o fracionamento por tamanho em colunas de cromatografia, utilizando a resina Sepharose CL-2B (Pharmacia). Fragmentos de cDNA grandes (1-5Kb) e pequenos (0,5-2Kb) foram inseridos em vetores apropriados tais como o pSPORT-1 (Life), o pCMV-SPORT 6 ou o pTrueBlue (Genetix). As bibliotecas assim obtidas foram introduzidas em células hospedeiras, preferencialmente por eletroporação (25mF, 200W, 1,8KV) em bactérias DH5 α (Invitrogen).

[174] As bactérias transformadas foram cultivadas em placas de petri de 7,5 cm contendo meio LB (Luria-Bertani) com agente seletivo apropriado (Sambrook et al., 1989). Aquelas apresentando inserto (brancas) foram transferidas para placas de 96 poços. Uma cópia de cada placa de 96 poços foi feita com auxílio de um replicador e as 2 réplicas mantidas a -70°C .

[175] O DNA para seqüenciamento foi preparado a partir de inóculos provenientes de uma das placas-réplica através do método de lise alcalina (Sambrook et al., 1989) modificado para funcionar com placas de 96 poços. As reações de seqüenciamento foram realizadas utilizando-se o BigDye Terminator Cycle Sequencing Kit (Applied Biosystems). Os primers empregados nas reações de seqüenciamento foram escolhidos de acordo com o vetor no qual a biblioteca foi construída. Todos os clones foram seqüenciados a partir da extremidade 5' original da molécula de mRNA do inserto, e uma parte deles também foi seqüenciada a partir da extremidade 3'.

[176] As reações de seqüenciamento foram lidas em seqüenciadores automáticos 3700 (Applied Biosystems). Os eletroferogramas resultantes foram transferidos para uma Base de Dados centralizada, localizada no Laboratório de Bioinformática da Embrapa Recursos Genéticos e Biotecnologia, para processamento e análises.

[177] As seqüências produzidas foram depositadas no GenBank (Benson et al. 1999) e transferidas para o BCCC (<http://www.bcccenter.fcav.unesp.br>) onde estarão à disposição da comunidade científica internacional. Tais seqüências também estão descritas na Listagem de Seqüências, SEQ. ID. No. 1 a SEQ. ID. No. 19.

[178] As técnicas de engenharia genética aqui descritas e conhecidas por técnicos no assunto, estão também descritas em Sambrook et al., 1989 (Sambrook, J., Fritsch, E. F. and Maniatis, T. 1989. *Molecular Cloning: A Laboratory Manual* – volumes 1, 2 e 3. Cold Spring Harbor Laboratory, Cold Spring Harbor, Nova Iorque), Silhavy et al., 1984 (Silhavy, T. J.; M. L. Bannan & L. W. Enquist. 1984. *Experiments with Gene Fusions*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.) e Ausubel, et al., 1987. (*Current Protocols in Molecular Biology*, pub. by Greene Publishing Assoc. and Wiley-Interscience).

Exemplo 3: Análise mecânica de sedas de diferentes espécies de aranhas.

[179] *Coleta da seda* – Cinco amostras de cada espécie com aproximadamente 5 cm de comprimento foram posicionadas em um cartão de teste mecânico, como descrito anteriormente por Stauffer et al. (Stauffer SL, Coguil SL, Lewis RV. Comparison of physical properties of three silks from *Nephila clavipes* and *Araneus gemmoides*. The journal of Arachnology. 1994; 22:5-11).

[180] *Medição do diâmetro das fibras* – As fibras foram analisadas utilizando um microscópio Nikon Eclipse E200 equipado com uma câmera. Cada seda foi observada com uma magnitude de 800 vezes, e as imagens foram visualizadas no programa SPOT Basic. Os diâmetros foram determinados utilizando o programa ImageJ versão 1.32 (<http://rsb.info.nih.gov/ij/>), e o valor final foi a média de 5 medidas ao longo das fibras.

[181] *Teste mecânico* – Cada fibra foi testada em um sistema de teste mecânico (MTS - Mechanical Testing System) Synergie 100, using a custom 10-g load cell. As fibras foram esticadas em uma taxa de 2mm/min, e os dados foram coletados em uma taxa de 35 Hz. A coleta dos dados foi feita utilizando o programa Testworks 4 software (MTS Systems Corporation, Cary, NC). Gráficos de stress vs. strain foram construídos utilizando o Microsoft Office Excel 2003.

[182] As equações abaixo foram usadas para calcular os valores de *stress*, *strain*, and *stiffness*.

σ (stress)= F/A , onde F é a força aplicada e A área da secção transversal.

ε (strain)= $\Delta L/L_0$, onde ΔL é a mudança no comprimento da fibra e L_0 é o comprimento inicial.

Y (stiffness or Young's modulus)= σ/ε

ESPÉCIE	DIÂMETRO (μm)	STRAIN (%)	STRESS (GPa)	STIFFNESS (GPa)
<i>Avicularia juruensis</i>	9.62 $\pm$ 3.92	7.54 $\pm$ 5.71	0.07 $\pm$ 0.03	0.017 $\pm$ 0.018
<i>Nephilengys cruentata</i>	4.82 $\pm$ 0.61	11.88 $\pm$ 2.97	0.81 $\pm$ 0.23	0.253 $\pm$ 0.306
<i>Parawixia bistrata</i>	7.56 $\pm$ 1.21	21.54 $\pm$ 5.01	0.76 $\pm$ 0.14	0.069 $\pm$ 0.024
<i>Nephila clavipes</i> *	?	22.89 $\pm$ 3.58	9.53 $\pm$ 0.06	0.062 $\pm$ 0.019
<i>Araneus diadematus</i> *	?	28.00 $\pm$ 4.00	1.08 $\pm$ 0.16	6.90 $\pm$ 1.22
<i>Argiope aurantia</i> ***	?	24.14 $\pm$ 4.73	1.36 $\pm$ 0.58	0.202 $\pm$ 0.205
<i>Latrodectus geometricus</i> **	2.78 $\pm$ 0.24	14.00 $\pm$ 6.00	0.83 $\pm$ 0.19	12.91 $\pm$ 7.38

i. Brooks AE, Steinkraus HB, Nelson SR, Lewis RV. An investigation of the divergence of major ampullate silk fibers from *Nephila clavipes* and *Argiope aurantia*. *Biomacromolecules*. 2005 Nov-Dec; 6(6):3095-9.

[183] ** Motriuk-Smith D, Lewis RV. Brown widow (*Latrodectus geometricus*) major ampullate silk protein and its material properties. *Biomed Sci Instrum*. 2004; 40:64-9.

[184] *** Madsen B, Shao ZZ, Vollrath F. Variability in the mechanical properties of spider silks on three levels: interspecific, intraspecific and intraindividual. *Int J Biol Macromol*. 1999 Mar-Apr; 24(2-3):301-6.

Exemplo 4: Comparação das seqüências da presente invenção com as seqüências já existentes no estado da técnica

[185] Com relação à SEQ ID NO 1 (Produto do Gene: *Nephilengys cruentata* – NCFlag) uma busca no GenBank utilizando-se o programa BLASTP revelou as seguintes 10 seqüências protéicas como as mais parecidas já descritas:

1. gi|7106228|gb|AAF36091.1

2. gi|7106224|gb|AAF36090.1
3. gi|2833649|gb|AAC38847.1
4. gi|13561982|gb|AAK30594.1
5. gi|2833647|gb|AAC38846.1
6. gi|70913024|gb|AAZ15322.1
7. gi|13562004|gb|AAK30605.1
8. gi|93138993|gb|ABE99838.1
9. gi|7106229|gb|AAF36092.1
10. gi|89276819|gb|ABD66603.1

[186] Conforme pode ser verificado na Tabela 1, a seqüência número 9 (gi|7106229, destacada com asterisco na tabela) é a que possui menor porcentagem de discrepância de aminoácidos quando alinhada com a SEQ ID NO 1. Do total de aminoácidos alinhados entre estas duas seqüências protéicas, 23% são divergentes. É importante notar que aminoácidos alinhados com espaços (*gaps*) não entram neste cálculo, e que o alinhamento considerado na análise é o alinhamento múltiplo entre todas as onze proteínas. Observando a Error: Reference source not found. é possível notar que a discrepância seria ainda maior se os *gaps* fosse incluídos no cálculo. Na Tabela é possível ainda constatar que várias outras seqüências já descritas (valores sublinhados) apresentam nível de similaridade maior entre si do que a observada entre SEQ ID NO 1 e as seqüências previamente descritas.

Tabela 1: Porcentagem de discrepância de aminoácidos entre os alinhamentos par a-par de 11 seqüências relacionadas à SEQ ID NO 1. “SI 1” indica a SEQ ID NO 1. Os números que identificam linhas e colunas correspondem aos números da lista das seqüências encontradas no estado da técnica descritos anteriormente. O número encontrado no cruzamento de uma linha com uma coluna representa o percentual de discrepância observado no alinhamento das duas seqüências. Destacado com asterisco encontra-se a maior semelhança observada entre SEQ ID NO 1 e seqüências já descritas.

Valores sublinhados representam semelhanças entre seqüências já descritas maiores que a encontrada para SEQ ID NO 1. Apenas metade da tabela está preenchida porque os valores são recíprocos.

	SI 1	1	2	3	4	5	6	7	8	9
1	44									
2	39	<u>20</u>								
3	25	46	40							
4	42	47	43	43						
5	42	<u>10</u>	27	45	47					
6	52	54	52	50	53	53				
7	52	54	52	50	53	53	<u>0</u>			
8	53	57	55	52	55	55	<u>16</u>	<u>16</u>		
9	23*	40	41	<u>8</u>	43	37	51	51	54	
10	59	65	62	60	59	62	58	58	61	60

[187] A Figura 1 indica um alinhamento entre a SEQ ID NO1 e a seqüência de número 9 (gi|7106229) destaca aminoácidos idênticos entre as duas seqüências. Os números sobre a seqüência identificam a posição no alinhamento, não correspondendo à posição em qualquer uma das seqüências. A diferença se deve à inserção de espaços (*gaps*) para manter o alinhamento.

[188] Para calcular a porcentagem de discrepância, a seqüência da presente invenção foi empregada em uma busca contra o banco de seqüências protéicas não redundante CDS do GenBank (Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Wheeler, D.L. (2006). GenBank. Nucleic Acids Res. 34 (Database issue): D16-20). A busca foi realizada sem uso do filtro de baixa complexidade, usando palavras de tamanho 3, empregando a matriz BLOSUM62, utilizando-se 11 pontos de penalidade para abertura de um *gap* e 1 ponto por aminoácido estendido. As seqüências com os 10 maiores *scores* foram utilizadas em um alinhamento múltiplo empregando o clustalx (Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. (1997). The ClustalX

windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, 24: 4876-4882). Para o alinhamento múltiplo foi empregada a matriz Gonnet 250 penalizando em 10 pontos a abertura de um *gap* e 0.20 pontos por aminoácido estendido No alinhamento par-a-par foi empregada a estratégia “lenta – precisa”, baixando-se a penalidade por extensão de *gap* para 0.10 pontos. O resultado do alinhamento múltiplo final foi empregado em uma comparação par-a-par de distância, com auxílio do programa MEGA3 (Kumar, S., Tamura, K. e Nei, M. (2004). MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Briefings in Bioinformatics* 5: 150-163). A porcentagem de discrepância foi calculada pela divisão do número de aminoácidos diferentes pelo número total de aminoácidos comparados. Os *gaps* de alinhamento foram mantidos na análise e eliminados par-a-par. A figura de alinhamento foi gerada a partir de um alinhamento entre as 2 seqüências consideradas, com os mesmos parâmetros empregados para o alinhamento único. O resultado de alinhamento foi formatado com auxílio do ESPript (Gouet, P., Courcelle, E., Stuart, D.I. and Metoz, F. (1999). ESPript: multiple sequence alignments in PostScript. *Bioinformatics*. 15: 305-8. <http://esprict.ibcp.fr/ESPript/ESPript/>).

[189] Com relação à SEQ ID NO2 (Produto do Gene: *Nephilengys cruentata* – NCTuSp), as 10 seqüências protéicas mais parecidas já descritas são:

1. gi|83758427
2. gi|68342501
3. gi|89365776
4. gi|89365774
5. gi|63054329
6. gi|61387244
7. gi|70927654
8. gi|63054327
9. gi|61387237
10. gi|61387234

Tabela 2: Porcentagem de discrepância de aminoácidos entre os alinhamentos par-a-par de 11 seqüências relacionadas à SEQ ID NO 2. “SI 2” indica a SEQ ID NO 2. Os números que identificam linhas e colunas correspondem aos números da lista acima. O número encontrado no cruzamento de uma linha com uma coluna representa o percentual de discrepância observado no alinhamento das duas seqüências. Destacado com asterisco encontra-se a maior semelhança observada entre SEQ ID NO 2 e seqüências já descritas. Os valores sublinhados representam as semelhanças entre seqüências já descritas maiores que a encontrada para SEQ ID NO 2. Apenas metade da tabela está preenchida porque os valores são recíprocos.

	SI 2	1	2	3	4	5	6	7	8	9
1	33									
2	33	<u>7</u>								
3	36	45	45							
4	37	48	45	<u>11</u>						
5	38	48	48	<u>11</u>	<u>13</u>					
6	37	22	20	47	47	49				
7	47	51	51	40	40	41	53			
8	46	51	50	39	39	41	53	<u>1</u>		
9	27*	47	44	<u>23</u>	<u>24</u>	22	47	33	34	
10	39	46	45	<u>10</u>	06	<u>14</u>	47	47	47	<u>21</u>

[190] Com relação a SEQ ID NO 3 (*Nephilengys cruentata* - NCMaSp), as 10 seqüências protéicas mais parecidas já descritas são:

1. gi|83758427
2. gi|68342501
3. gi|89365776
4. gi|89365774
5. gi|63054329
6. gi|61387244

7. gi|70927654
8. gi|63054327
9. gi|61387237
10. gi|61387234

Tabela 3: Porcentagem de discrepância de aminoácidos entre os alinhamentos par-a-par de 11 seqüências relacionadas à SEQ ID NO 3. “SI 3” indica a SEQ ID NO 3.

Os números que identificam linhas e colunas correspondem aos números da lista acima. O número encontrado no cruzamento de uma linha com uma coluna representa o percentual de discrepância observado no alinhamento das duas seqüências. Destacado com asterisco encontra-se a maior semelhança observada entre SEQ ID NO 3 e seqüências já descritas. Os valores sublinhados representam as semelhanças entre seqüências já descritas maiores que a encontrada para SEQ ID NO 3. Apenas metade da tabela está preenchida porque os valores são recíprocos.

	SI 3	1	2	3	4	5	6	7	8	9
1	7*									
2	15	7								
3	18	9	11							
4	24	9	14	<u>2</u>						
5	22	10	13	<u>2</u>	<u>4</u>					
6	17	10	10	<u>1</u>	<u>1</u>	<u>1</u>				
7	17	9	14	<u>2</u>	<u>2</u>	<u>2</u>	<u>2</u>			
8	16	7	14	9	9	7	8	9		
9	22	10	13	<u>2</u>	<u>3</u>	<u>2</u>	<u>2</u>	<u>2</u>	10	
10	11	8	5	<u>1</u>	<u>1</u>	<u>0</u>	<u>0</u>	<u>1</u>	<u>6</u>	<u>1</u>

11. Com relação a SEQ ID NO 4 (*Nephilengys cruentata* - NCMaSp), as 10 seqüências protéicas mais parecidas já descritas são:

1. gi|2605800|
2. gi|85680899
3. gi|2605798|
4. gi|50363143
5. gi|50363141
6. gi|50363145
7. gi|2911274|
8. gi|50363137
9. gi|13561992
10. gi|50363139

Tabela 4: Porcentagem de discrepância de aminoácidos entre os alinhamentos par-a-par de 11 seqüências relacionadas à SEQ ID NO 4. “SI 4” indica a SEQ ID NO 4. Os números que identificam linhas e colunas correspondem aos números da lista acima. O número encontrado no cruzamento de uma linha com uma coluna representa o percentual de discrepância observado no alinhamento das duas seqüências. Destacado com asterisco encontra-se a maior semelhança observada entre SEQ ID NO 4 e seqüências já descritas. Os valores sublinhados representam as semelhanças entre seqüências já descritas maiores que a encontrada para SEQ ID NO 4. Apenas metade da tabela está preenchida porque os valores são recíprocos.

	SI 4	1	2	3	4	5	6	7	8	9
1	25*									
2	41	34								
3	50	34	<u>8</u>							
4	45	50	54	54						
5	50	51	59	58	<u>2</u>					
6	50	52	55	55	<u>19</u>	<u>21</u>				

7	51	53	59	57	<u>4</u>	<u>7</u>	19			
8	47	52	57	55	<u>1</u>	<u>1</u>	18	4		
9	50	52	58	58	29	34	37	38	30	
10	49	54	51	50	<u>2</u>	<u>2</u>	17	2	2	33

[191] Com relação a SEQ ID NO 5 (*Nephilengys cruentata* - NCMiSp 06A01), as 10 seqüências protéicas mais parecidas já descritas são:

1. gi|2605798
2. gi|8572061
3. gi|765323
4. gi|27228959
5. gi|47007923
6. gi|47007963
7. gi|13561992
8. gi|108707764
9. gi|89276817
10. gi|2914731

Tabela 5: Porcentagem de discrepância de aminoácidos entre os alinhamentos par-a-par de 11 seqüências relacionadas à SEQ ID NO 5. “SI 5” indica a SEQ ID NO 5. Os números que identificam linhas e colunas correspondem aos números da lista acima. O número encontrado no cruzamento de uma linha com uma coluna representa o percentual de discrepância observado no alinhamento das duas seqüências. Destacado com asterisco encontra-se a maior semelhança observada entre SEQ ID NO 5 e seqüências já descritas. Os valores sublinhados representam as semelhanças entre seqüências já descritas maiores que a encontrada para SEQ ID NO 5. Apenas metade da tabela está preenchida porque os valores são recíprocos.

	SI 5	1	2	3	4	5	6	7	8	9
1	13*									
2	39	52								
3	39	47	2							
4	46	47	49	49						
5	46	56	49	49	57					
6	45	53	36	39	54	53				
7	53	58	60	60	58	64	59			
8	47	51	48	48	56	53	52	59		
9	53	57	62	61	56	66	61	37	58	
10	54	53	60	60	54	67	57	36	58	22

[192] Com relação a SEQ ID NO 6 (*Nephilengys cruentata* - NCMiSp 11F12), as 10 seqüências protéicas mais parecidas já descritas são:

1. gi|85680899
2. gi|2605798|
3. gi|89113992
4. gi|13562018
5. gi|63054333
6. gi|63054353
7. gi|63054329
8. gi|61387237
9. gi|13561994
10. gi|89365776

Tabela 6: Porcentagem de discrepância de aminoácidos entre os alinhamentos par-a-par de 11 seqüências relacionadas à SEQ ID NO 6. “SI 6” indica a SEQ ID NO 6. Os números que identificam linhas e colunas correspondem aos números da lista acima.

O número encontrado no cruzamento de uma linha com uma coluna representa o percentual de discrepância observado no alinhamento das duas seqüências. Destacado com asterisco encontra-se a maior semelhança observada entre SEQ ID NO 6 e seqüências já descritas. Os valores sublinhados representam as semelhanças entre seqüências já descritas maiores que a encontrada para SEQ ID NO 6. Apenas metade da tabela está preenchida porque os valores são recíprocos.

	SI 6	1	2	3	4	5	6	7	8	9
1	67									
2	64*	<u>8</u>								
3	74	<u>50</u>	<u>54</u>							
4	89	83	84	87						
5	82	77	77	78	78					
6	87	76	80	76	79	<u>60</u>				
7	83	78	81	80	77	<u>54</u>	<u>43</u>			
8	85	77	81	80	75	<u>53</u>	<u>47</u>	<u>22</u>		
9	77	74	76	81	75	<u>63</u>	74	69	69	
10	83	79	81	80	80	<u>54</u>	<u>42</u>	<u>11</u>	<u>22</u>	69

[193] Com relação a SEQ ID NO 7 (*Nephilengys cruentata* - NCfibroin) as 10 seqüências protéicas mais parecidas já descritas são:

1. gi|63054331
2. gi|6984160|
3. gi|70905642
4. gi|89365776
5. gi|70905641
6. gi|63054329
7. gi|70905643
8. gi|14973269

9. gi|13562018

10. gi|8885520|

Tabela 7: Porcentagem de discrepância de aminoácidos entre os alinhamentos par-a-par de 11 seqüências relacionadas à SEQ ID NO 7. “SI 7” indica a SEQ ID NO 7. Os números que identificam linhas e colunas correspondem aos números da lista acima. O número encontrado no cruzamento de uma linha com uma coluna representa o percentual de discrepância observado no alinhamento das duas seqüências. Destacado com asterisco encontra-se a maior semelhança observada entre SEQ ID NO 7 e seqüências já descritas. Os valores sublinhados representam as semelhanças entre seqüências já descritas maiores que a encontrada para SEQ ID NO 7. Apenas metade da tabela está preenchida porque os valores são recíprocos.

	SI 7	1	2	3	4	5	6	7	8	9
1	80									
2	77	<u>72</u>								
3	78	<u>73</u>	<u>68</u>							
4	81	<u>58</u>	77	<u>75</u>						
5	78	<u>73</u>	<u>68</u>	<u>3</u>	<u>75</u>					
6	81	<u>61</u>	<u>73</u>	<u>74</u>	<u>11</u>	<u>74</u>				
7	78	<u>72</u>	<u>72</u>	<u>25</u>	78	25	<u>74</u>			
8	76*	<u>70</u>	<u>45</u>	<u>61</u>	<u>72</u>	<u>61</u>	<u>72</u>	<u>68</u>		
9	84	77	78	<u>72</u>	79	72	79	<u>74</u>	<u>72</u>	
10	79	<u>72</u>	<u>51</u>	<u>65</u>	<u>75</u>	<u>65</u>	<u>74</u>	<u>66</u>	<u>33</u>	76

[194] Com relação a SEQ ID NO 8 (*Nephilengys cruentata* - NCdefensina) as 10 seqüências protéicas mais parecidas já descritas são:

1. gi|89512121

2. gi|41019463

3. gi|41019465
4. gi|77158011
5. gi|90192368
6. gi|33348850
7. gi|62275780
8. gi|57792507
9. gi|49458046
10. gi|49458052

Tabela 8: Porcentagem de discrepância de aminoácidos entre os alinhamentos par-a-par de 11 seqüências relacionadas à SEQ ID NO 8. “SI 8” indica a SEQ ID NO 8.

Os números que identificam linhas e colunas correspondem aos números da lista acima. O número encontrado no cruzamento de uma linha com uma coluna representa o percentual de discrepância observado no alinhamento das duas seqüências. Destacado com asterisco encontra-se a maior semelhança observada entre SEQ ID NO 8 e seqüências já descritas. Os valores sublinhados representam as semelhanças entre seqüências já descritas maiores que a encontrada para SEQ ID NO 8. Apenas metade da tabela está preenchida porque os valores são recíprocos.

	SI 8	1	2	3	4	5	6	7	8	9
1	63*									
2	66	<u>62</u>								
3	67	63	<u>11</u>							
4	67	<u>62</u>	<u>1</u>	<u>11</u>						
5	63	<u>62</u>	<u>49</u>	<u>51</u>	<u>50</u>					
6	65	<u>61</u>	<u>2</u>	<u>12</u>	<u>2</u>	<u>50</u>				
7	64	<u>60</u>	<u>0</u>	<u>10</u>	<u>0</u>	<u>48</u>	<u>0</u>			
8	67	63	<u>17</u>	<u>15</u>	<u>17</u>	<u>51</u>	<u>18</u>	<u>15</u>		

9	66	<u>61</u>	<u>18</u>	<u>18</u>	<u>18</u>	<u>51</u>	<u>18</u>	<u>18</u>	<u>13</u>	
10	65	<u>61</u>	<u>15</u>	<u>13</u>	<u>15</u>	<u>50</u>	<u>15</u>	<u>15</u>	<u>1</u>	<u>13</u>

[195] Com relação a SEQ ID NO 9 (*Avicularia juruensis* AJFibroin 1A) as 10 seqüências protéicas mais parecidas já descritas são:

1. gi|17536963
2. gi|50548483
3. gi|76580883
4. gi|52428273
5. gi|52209673
6. gi|71416355
7. gi|89276819
8. gi|13562018
9. gi|71654760
10. gi|109511662

Tabela 9: Porcentagem de discrepância de aminoácidos entre os alinhamentos par-a-par de 11 seqüências relacionadas à SEQ ID NO 9. “SI 9” indica a SEQ ID NO 9. Os números que identificam linhas e colunas correspondem aos números da lista acima. O número encontrado no cruzamento de uma linha com uma coluna representa o percentual de discrepância observado no alinhamento das duas seqüências. Destacado com asterisco encontra-se a maior semelhança observada entre SEQ ID NO 9 e seqüências já descritas. Os valores sublinhados representam as semelhanças entre seqüências já descritas maiores que a encontrada para SEQ ID NO 9. Apenas metade da tabela está preenchida porque os valores são recíprocos.

	SI 9	1	2	3	4	5	6	7	8	9
--	-------------	----------	----------	----------	----------	----------	----------	----------	----------	----------

1	84									
2	86	87								
3	82	83	86							
4	82	83	86	<u>1</u>						
5	82	83	86	<u>1</u>	<u>1</u>					
6	80	84	87	77	77	77				
7	76*	81	83	87	87	87	80			
8	76*	85	87	86	86	86	84	84		
9	79	84	88	76	76	76	<u>6</u>	79	83	
10	82	86	83	81	81	81	79	84	88	78

[196] Com relação a SEQ ID NO 10 (*Avicularia juruensis* AJFibroin 1B) as 10 seqüências protéicas mais parecidas já descritas são:

1. gi|68171564
2. gi|17536963
3. gi|89276819
4. gi|38197745
5. gi|13562020
6. gi|38197743
7. gi|83758429
8. gi|71416355
9. gi|70913022
10. gi|1263289

Tabela 10: Porcentagem de discrepância de aminoácidos entre os alinhamentos par-a-par de 11 seqüências relacionadas à SEQ ID NO 10. “SI 10” indica a SEQ ID NO 10. Os números que identificam linhas e colunas correspondem aos números da lista acima. O número encontrado no cruzamento de uma linha com uma coluna representa o

percentual de discrepância observado no alinhamento das duas seqüências. Destacado com asterisco encontra-se a maior semelhança observada entre SEQ ID NO 10 e seqüências já descritas. Os valores sublinhados representam as semelhanças entre seqüências já descritas maiores que a encontrada para SEQ ID NO 10. Apenas metade da tabela está preenchida porque os valores são recíprocos.

	SI 10	1	2	3	4	5	6	7	8	9
1	88									
2	83	91								
3	75	91	81							
4	75	86	85	<u>45</u>						
5	76	90	84	76	<u>70</u>					
6	74	88	85	<u>44</u>	<u>0</u>	<u>68</u>				
7	77	88	84	85	87	84	85			
8	82	89	84	78	81	85	80	84		
9	76	88	83	<u>41</u>	<u>50</u>	<u>72</u>	<u>50</u>	82	78	
10	73*	88	83	<u>30</u>	<u>45</u>	<u>70</u>	<u>45</u>	85	78	<u>49</u>

[197] Com relação a SEQ ID NO 11 (Produto do Gene: *Avicularia juruensis* AJFibroin 2) as 10 seqüências protéicas mais parecidas já descritas são:

1. gi|1405387
2. gi|38197749
3. gi|38197751
4. gi|38197745
5. gi|50309199
6. gi|38197755
7. gi|38197759
8. gi|38197747

9. gi|50423563

10. gi|38197757

Tabela 11: Porcentagem de discrepância de aminoácidos entre os alinhamentos par-a-par de 11 seqüências relacionadas à SEQ ID NO 11. “SI 11” indica a SEQ ID NO 11. Os números que identificam linhas e colunas correspondem aos números da lista acima. O número encontrado no cruzamento de uma linha com uma coluna representa o percentual de discrepância observado no alinhamento das duas seqüências. Destacado com asterisco encontra-se a maior semelhança observada entre SEQ ID NO 11 e seqüências já descritas. Os valores sublinhados representam as semelhanças entre seqüências já descritas maiores que a encontrada para SEQ ID NO 11. Apenas metade da tabela está preenchida porque os valores são recíprocos.

	SI 11	1	2	3	4	5	6	7	8	9
1	66									
2	41	76								
3	42	76	<u>1</u>							
4	55	77	<u>1</u>	<u>0</u>						
5	68	83	83	83	86					
6	33	75	<u>8</u>	<u>10</u>	<u>10</u>	81				
7	42	76	<u>2</u>	<u>2</u>	<u>2</u>	81	<u>9</u>			
8	42	76	<u>1</u>	<u>1</u>	<u>1</u>	83	<u>9</u>	<u>2</u>		
9	76	86	87	88	89	68	86	86	87	
10	32*	77	<u>17</u>	<u>17</u>	<u>17</u>	82	<u>8</u>	<u>17</u>	<u>18</u>	84

[198] Com relação a SEQ ID NO 12 (Produto do Gene: *Avicularia juruensis* AJNegProtein 1) as 10 seqüências protéicas mais parecidas já descritas são:

[199] gi|87133239

- [200] gi|87133241
- [201] gi|17539308
- [202] gi|72011370
- [203] gi|70913024
- [204] gi|68365042
- [205] gi|49871101
- [206] gi|110756487
- [207] gi|85111705
- [208] gi|39973263

Tabela 12: Porcentagem de discrepância de aminoácidos entre os alinhamentos par-a-par de 11 seqüências relacionadas à SEQ ID NO 12. “SI 12” indica a SEQ ID NO 12. Os números que identificam linhas e colunas correspondem aos números da lista acima. O número encontrado no cruzamento de uma linha com uma coluna representa o percentual de discrepância observado no alinhamento das duas seqüências. Destacado com asterisco encontra-se a maior semelhança observada entre SEQ ID NO 12 e seqüências já descritas. Os valores sublinhados representam as semelhanças entre seqüências já descritas maiores que a encontrada para SEQ ID NO 12. Apenas metade da tabela está preenchida porque os valores são recíprocos.

	SI 12	1	2	3	4	5	6	7	8	9
1	73*									
2	80	<u>32</u>								
3	89	90	93							
4	91	92	89	93						
5	86	<u>40</u>	<u>43</u>	93	91					
6	87	92	91	89	94	90				
7	84	79	77	91	92	83	88			

8	90	93	93	92	91	93	91	92		
9	93	90	89	93	94	92	93	92	93	
10	90	88	93	92	92	92	91	91	92	93

[209] Com relação a SEQ ID NO 13 (Produto do Gene: *Avicularia juruensis* AJNegProtein 2) as 10 seqüências protéicas mais parecidas já descritas são:

1. gi|22074292
2. gi|15021422
3. gi|88713113
4. gi|55593156
5. gi|82539404
6. gi|109467082
7. gi|3236370
8. gi|6677817
9. gi|71992048
10. gi|62175305|

Tabela 13: Porcentagem de discrepância de aminoácidos entre os alinhamentos par-a-par de 11 seqüências relacionadas à SEQ ID NO 13. “SI 13” indica a SEQ ID NO 13. Os números que identificam linhas e colunas correspondem aos números da lista acima. O número encontrado no cruzamento de uma linha com uma coluna representa o percentual de discrepância observado no alinhamento das duas seqüências. Destacado com asterisco encontra-se a maior semelhança observada entre SEQ ID NO 13 e seqüências já descritas. Os valores sublinhados representam as semelhanças entre seqüências já descritas maiores que a encontrada para SEQ ID NO 13. Apenas metade da tabela está preenchida porque os valores são recíprocos.

	SI 13	1	2	3	4	5	6	7	8	9
1	72*									
2	72*	<u>0</u>								
3	74	<u>59</u>	<u>59</u>							
4	85	90	90	87						
5	90	91	91	91	94					
6	83	85	85	83	<u>64</u>	93				
7	83	84	84	83	92	92	90			
8	84	87	87	82	<u>64</u>	93	<u>24</u>	90		
9	75	<u>58</u>	<u>58</u>	<u>56</u>	88	87	84	80	84	
10	80	77	77	80	90	92	90	87	90	72

[210] Com relação a SEQ ID NO 14 (Produto do Gene: *Parawixia bistrata* - PBAciniform) as 10 seqüências protéicas mais parecidas já descritas são:

1. gi|49871101
2. gi|89114010
3. gi|44980633
4. gi|40787372
5. gi|58864899
6. gi|82936154
7. gi|16741397
8. gi|7243103
9. gi|11596144
10. gi|45439370

Tabela 14: Porcentagem de discrepância de aminoácidos entre os alinhamentos par-a-par de 11 seqüências relacionadas à SEQ ID NO 14. “SI 14” indica a SEQ ID NO

14. Os números que identificam linhas e colunas correspondem aos números da lista acima. O número encontrado no cruzamento de uma linha com uma coluna representa o percentual de discrepância observado no alinhamento das duas seqüências. Destacado com asterisco encontra-se a maior semelhança observada entre SEQ ID NO 14 e seqüências já descritas. Os valores sublinhados representam as semelhanças entre seqüências já descritas maiores que a encontrada para SEQ ID NO 14. Apenas metade da tabela está preenchida porque os valores são recíprocos.

	SI 14	1	2	3	4	5	6	7	8	9
1	28*									
2	75	74								
3	89	90	91							
4	89	89	91	90						
5	89	91	94	89	93					
6	89	91	94	89	93	<u>1</u>				
7	91	93	96	89	93	<u>0</u>	<u>0</u>			
8	89	91	94	89	93	<u>1</u>	<u>2</u>	<u>1</u>		
9	89	91	94	89	92	<u>1</u>	<u>2</u>	<u>1</u>	<u>0</u>	
10	89	91	94	89	92	<u>1</u>	<u>2</u>	<u>1</u>	<u>0</u>	<u>0</u>

[211] Com relação a SEQ ID NO 15 (Produto do Gene: *Parawixia bistrata* – PBFlag) as 10 seqüências protéicas mais parecidas já descritas são:

1. gi|62465589
2. gi|27228957
3. gi|13561980
4. gi|109500095
5. gi|47606845
6. gi|5921193

7. gi|51975245
8. gi|3236370
9. gi|109487472
10. gi|47219204

Tabela 15: Porcentagem de discrepância de aminoácidos entre os alinhamentos par-a-par de 11 seqüências relacionadas à SEQ ID NO 15. “SI 15” indica a SEQ ID NO 15. Os números que identificam linhas e colunas correspondem aos números da lista acima. O número encontrado no cruzamento de uma linha com uma coluna representa o percentual de discrepância observado no alinhamento das duas seqüências. Destacado com asterisco encontra-se a maior semelhança observada entre SEQ ID NO 15 e seqüências já descritas. Os valores sublinhados representam as semelhanças entre seqüências já descritas maiores que a encontrada para SEQ ID NO 15. Apenas metade da tabela está preenchida porque os valores são recíprocos.

	SI 15	1	2	3	4	5	6	7	8	9
1	20*									
2	20*	<u>0</u>								
3	28	<u>19</u>	20							
4	90	87	91	91						
5	38	35	35	37	91					
6	82	90	84	85	92	84				
7	70	79	70	76	91	75	84			
8	82	90	84	85	92	84	<u>17</u>	82		
9	81	90	83	84	92	84	<u>16</u>	83	<u>7</u>	
10	83	90	85	86	91	85	60	84	58	59

[212] Com relação a SEQ ID NO 16 (Produto do Gene: *Parawixia bistrata* – PBMiSp) as 10 seqüências protéicas mais parecidas já descritas são:

1. gi|1405387
2. gi|88175701
3. gi|17507879
4. gi|55619549
5. gi|71026577
6. gi|50902080
7. gi|66805291
8. gi|71408138
9. gi|50931795
10. gi|55274106

Tabela 16: Porcentagem de discrepância de aminoácidos entre os alinhamentos par-a-par de 11 seqüências relacionadas à SEQ ID NO 16. “SI 16” indica a SEQ ID NO 16. Os números que identificam linhas e colunas correspondem aos números da lista acima. O número encontrado no cruzamento de uma linha com uma coluna representa o percentual de discrepância observado no alinhamento das duas seqüências. Destacado com asterico encontra-se a maior semelhança observada entre SEQ ID NO 16 e seqüências já descritas. Os valores sublinhados representam as semelhanças entre seqüências já descritas maiores que a encontrada para SEQ ID NO 16. Apenas metade da tabela está preenchida porque os valores são recíprocos.

	SI 16	1	2	3	4	5	6	7	8	9
1	72									
2	92	93								
3	72	<u>0</u>	93							
4	81	87	92	87						
5	88	72	93	72	91					

6	83	85	92	85	86	92				
7	94	93	92	93	92	90	93			
8	82	82	93	82	91	89	88	94		
9	92	93	91	93	93	94	91	90	92	
10	46*	82	94	82	86	91	84	90	86	94

[213] Com relação a SEQ ID NO 17 (Produto do Gene: *Parawixia bistrata* – PBMaSp1) as 10 seqüências protéicas mais parecidas já descritas são:

1. gi|47569234
2. gi|32815671
3. gi|49329892
4. gi|51975246
5. gi|42782789
6. gi|109500095
7. gi|55274104
8. gi|55274084
9. gi|50363145
10. gi|1263285

Tabela 17: Porcentagem de discrepância de aminoácidos entre os alinhamentos par-a-par de 11 seqüências relacionadas à SEQ ID NO 17. “SI 17” indica a SEQ ID NO 17. Os números que identificam linhas e colunas correspondem aos números da lista acima. O número encontrado no cruzamento de uma linha com uma coluna representa o percentual de discrepância observado no alinhamento das duas seqüências. Destacado com asterisco encontra-se a maior semelhança observada entre SEQ ID NO 17 e seqüências já descritas. Os valores sublinhados representam as semelhanças entre seqüências já descritas maiores que a encontrada para SEQ ID NO 17. Apenas metade da tabela está preenchida porque os valores são recíprocos.

	SI 17	1	2	3	4	5	6	7	8	9
1	73									
2	14*	81								
3	74	14	81							
4	74	15	80	14						
5	74	<u>8</u>	81	<u>12</u>	14					
6	88	85	79	86	86	85				
7	20	83	20	83	83	84	78			
8	23	80	19	80	79	81	84	<u>2</u>		
9	33	76	21	76	77	76	87	<u>8</u>	<u>9</u>	
10	29	74	14	74	73	74	87	25	25	30

[214] Com relação a SEQ ID NO 18 (Produto do Gene: *Parawixia bistrata* – PBMaSp2) as 10 seqüências protéicas mais parecidas já descritas são:

1. gi|70913022
2. gi|32815671
3. gi|55274104
4. gi|55274080
5. gi|55274092
6. gi|55274136
7. gi|55274128
8. gi|55274086
9. gi|38197745
10. gi|55274082

Tabela 18: Porcentagem de discrepância de aminoácidos entre os alinhamentos par-a-par de 11 seqüências relacionadas à SEQ ID NO 18. “SI 18” indica a SEQ ID NO

18. Os números que identificam linhas e colunas correspondem aos números da lista acima. O número encontrado no cruzamento de uma linha com uma coluna representa o percentual de discrepância observado no alinhamento das duas seqüências. Destacado com asterico encontra-se a maior semelhança observada entre SEQ ID NO 18 e seqüências já descritas. Os valores sublinhados representam as semelhanças entre seqüências já descritas maiores que a encontrada para SEQ ID NO 18. Apenas metade da tabela está preenchida porque os valores são recíprocos.

	SI 18	1	2	3	4	5	6	7	8	9
1	55									
2	17*	45								
3	22	42	23							
4	31	46	21	<u>2</u>						
5	27	45	26	<u>4</u>	<u>6</u>					
6	39	52	23	<u>2</u>	<u>1</u>	<u>6</u>				
7	31	47	21	<u>2</u>	<u>2</u>	<u>6</u>	<u>3</u>			
8	26	44	21	<u>2</u>	<u>0</u>	<u>2</u>	<u>1</u>	<u>0</u>		
9	19	57	25	31	38	35	45	37	33	
10	34	46	23	<u>4</u>	<u>2</u>	<u>8</u>	<u>3</u>	<u>4</u>	<u>2</u>	40

[215] Com relação à SEQ ID NO19 (Produto do Gene: *Parawixia bistrata* – silk gland spidroin), as 10 seqüências protéicas mais parecidas já descritas são:

1. gi|1399945
2. gi|70913022
3. gi|17507879
4. gi|55274080
5. gi|55274086
6. gi|55274136
7. gi|55274128
8. gi|55274112

9. gi|55274138

10. gi|55274092

Tabela 19: Porcentagem de discrepância de aminoácidos entre os alinhamentos par-a-par de 11 seqüências relacionadas à SEQ ID NO 19. “SI 19” indica a SEQ ID NO 19. Os números em negrito que identificam linhas e colunas correspondem aos números da lista acima. O número encontrado no cruzamento de uma linha com uma coluna representa o percentual de discrepância observado no alinhamento das duas seqüências. Destacado com asterisco encontra-se a maior semelhança observada entre SEQ ID NO 19 e seqüências já descritas. Os valores sublinhados representam as semelhanças entre seqüências já descritas maiores que a encontrada para SEQ ID NO 19. Apenas metade da tabela está preenchida porque os valores são recíprocos.

	SI 19	1	2	3	4	5	6	7	8	9
1	78									
2	54	79								
3	78	<u>0</u>	79							
4	48	79	<u>37</u>	79						
5	47*	77	<u>37</u>	77	<u>0</u>					
6	48	77	<u>37</u>	77	<u>1</u>	<u>1</u>				
7	47*	80	<u>38</u>	80	<u>2</u>	<u>0</u>	<u>3</u>			
8	49	78	<u>37</u>	78	<u>1</u>	<u>1</u>	<u>1</u>	<u>1</u>		
9	49	78	<u>37</u>	78	<u>2</u>	<u>2</u>	<u>2</u>	<u>2</u>	<u>1</u>	
10	51	80	<u>42</u>	80	<u>8</u>	<u>8</u>	<u>8</u>	<u>8</u>	<u>6</u>	<u>7</u>

Exemplo 5: Construção de vetores de expressão contendo genes das proteínas das sedas de aranha em plantas

[216] Os vetores de expressão utilizados na presente invenção contêm pelo menos uma seqüência promotora, uma seqüência codificadora da proteína de seda de aranha selecionada do grupo SEQ ID NO 1-19 e uma seqüência de poliadenilação. O vetor de expressão pode conter ainda seqüências regulatórias responsáveis pelo processamento pós-transcricional e compartimentalização das proteínas heterólogas.

[217] Vetores de expressão foram construídos utilizando metodologias padronizadas da manipulação do DNA recombinante (Sambrook, J. Molecular Cloning: A Laboratory Manual (3-Volume Set) Cold Spring Harbor Laboratory Press; 3ª edição, 2001). Basicamente, a região codificadora referente aos fragmentos: Para soja, as seqüências codificadoras das proteínas das sedas foram clonadas sob controle do peptídeo sinal e promotor da beta-conglicina (previamente clonado no Laboratório de Transferência de Genes – EMBRAPA) (Figura 2). As alterações previstas visam adaptar o fragmento para a adição da seqüência codificadora do peptídeo sinal de planta, possibilitando seu correto processamento das proteínas recombinantes. Em algodão, as seqüências codificadoras das proteínas das sedas foram clonadas sob controle do peptídeo sinal e promotor da actina2 de Arabidopsis (Aragão F.J.L., Vianna G.R., Carvalheira S.B.R., Rech E.L. (2005) Germ line genetic transformation in cotton (*Gossypium hirsutum* L.) by selection of transgenic meristematic cells with a herbicide molecule. Plant Sci.168: 227-1233).

[218] As seqüências de genes e proteínas da presente invenção podem ser modificadas de acordo com a intenção de uso e ainda estarem presentes no escopo da invenção. Por exemplo, quando se deseja que a fibra apresente baixa elasticidade, a seqüência de repeticoes de Alanina (Ala) deve ser removida, e, em contraste, quando quando se busca uma alta taxa de elasticidade, o tamanho da porção poli-Ala pode ser aumentado. Pode-se também incluir nas construções apenas os módulos repetitivos das sequências aqui apresentadas, e, ainda, uma combinação de módulos de diferentes proteínas para se alcançar a característica de interesse.

Exemplo 6: Produção de plantas de soja e algodão contendo as seqüências dos genes das proteínas das sedas de aranha

[219] Os vetores de expressão obtidos foram utilizados em experimentos de transformação e co-transformação com o vetor pAG1 (Figura 3) contendo a seqüência codificadora do gene *ahas* sob controle do promotor do gene *ahas* e terminador NOS. Este vetor permite a seleção de plantas transgênicas de soja e algodão *in vitro*. Quando necessário, o gene marcador GUS sob controle do promotor 35SCaMV e terminador nos foram clonados em pAG1. Plantas transgênicas de soja e algodão foram desenvolvidas através do sistema de biobalística desenvolvido pela EMBRAPA de acordo com a patente PI9714887-3.

[220] As plantas transgênicas produzidas foram analisadas por PCR (Dellaporta, S.L., Wood, J. and Hicks, J.B. (1983). A plant DNA miniprep: version II. Plant Molecular Biology Reports 1:19-21; Aragão, F.J.L., Barros, L.M.G., Brasileiro, A.C.M., Ribeiro, S.G., Smith, F.D., Sanford, J.C. Rech E.L. (1996). Inheritance of foreign genes in transgenic bean (*Phaseolus vulgaris* L.) co-transformed via particle bombardment. Theor Appl Genet 93:142-150) e Southern blot (Dellaporta, S.L., Wood, J. and Hicks, J.B. (1983). A plant DNA miniprep: version II. Plant Molecular Biology Reports 1:19-21; Sambrook J., Fritsch E.F., Maniatis T. (1989). Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York) para detecção dos transgenes e por análise bioquímica e bioensaios. Foram feitas análises histoquímicas para detectar a integração de genes heterólogos.

Exemplo 7: Utilização de agroinfiltração e vetores virais para a expressão de proteínas de aranhas

[221] A expressão transitória dos genes de aranha foi avaliada pela inoculação in planta de linhagens de *Agrobacterium tumefaciens*, contendo os genes de aranha clonados no vetor binário pCambia 1300 (www.cambia.org), ou em um vetor viral baseado no Potato

virus X (PVX). Este vetor viral, baseado no vetor pGR107 (Chapman S, Kavanagh T and Baucombe D (1992) Potato virus X as a vector for gene expression in plants. Plant J 2:549-557), apresenta uma duplicação do promotor da capa protéica e sítios de restrição para a clonagem de genes exógenos. A sequência correspondente ao PVX está clonada em um vetor binário, sob controle do promotor CaMV35S, permitindo a inoculação através da *Agrobacterium* (agroinoculação). O vetor original, pGR107, foi modificado pela adição do cassete de conversão do sistema Gateway™ de clonagem (Invitrogen), no sítio de SmaI. O vetor resultante, PVXGW, portanto, é compatível com o sistema de clonagem Gateway™ (Invitrogen), o que permitiu a clonagem de genes de forma muito mais rápida e eficiente, comparada aos métodos usuais de restrição e ligação. Linhagens de *Agrobacterium tumefaciens* com vetores binários ou vetores virais contendo os genes de aranha foram inoculadas em folhas de *Nicotiana benthamiana* e a expressão transiente foi avaliada após 4-6 dias.

Exemplo 8: Construção de vetores de expressão contendo genes das proteínas das sedas de aranha em vetores bacterianos e produção dessas proteínas no sistema de expressão bacteriano

[222] Os vetores de expressão em bactérias utilizados na presente invenção foram construídos utilizando-se o sistema pET (Novagen). Módulos repetitivos das sequências codificadoras das proteínas de seda de aranha selecionadas do grupo SEQ ID NO 1-19 foram sintetizados e clonados no vetor pUC19 associados a sítios de restrição que permitiram a multiplicação das unidades repetitivas N vezes, segundo a estratégia descrita por Lewis et al. (Lewis RV, Hinman M, Kothakota S, Fournier MJ. Expression and purification of a spider silk protein: A new strategy for producing repetitive proteins. PROTEIN EXPRESSION AND PURIFICATION 7 (4): 400-406 JUN 1996). Os cassetes contendo os módulos repetitivos foram transferidos para o vetor pET19b (Novagen) (Figura 4) sob o controle do promotor T7 e fusionados a uma cauda de histidinas N-terminal. Esses vetores resultantes foram utilizados para transformar células competentes

de *E. coli* (linhagens BL21(DE3) e BL21(DE3)pLysS) por choque térmico. As bactérias recombinantes contendo o vetor de expressão foram inoculadas em meio de cultura contendo o antibiótico apropriado e crescidas até uma OD₆₀₀ entre 0,8 e 0,9. A produção das proteínas foi induzida com IPTG na concentração de 1mM durante 3h a 37°C sob agitação. A cultura foi então centrifugada a 1500g por 15 min e ressuspensa em tampão de lise, como descrito pelo fabricante do sistema. Nessas condições experimentais, as células BL21 (DE3) e BL21(DE3)pLysS induziram a expressão dos genes citados nesse sistema heterólogo, e o extrato lisado foi usado para se proceder à purificação em coluna de cromatografia utilizando-se a resina Ni-NTA His-Bind Resin (Qiagen) carregada com 50mM de NiSO₄ e 5mM de imidazole. A fração contendo a proteína recombinante foi eluída utilizando-se um tampão de eluição contendo 100mM de imidazole, dialisada em água destilada durante 2 dias e liofilizada. A recuperação da proteína purificada se manteve entre 0,2mg/g e 10mg/g do peso seco de células.

Exemplo 9: Construção de vetores de expressão contendo genes das proteínas das sedas de aranha em vetor de células de mamíferos e produção dessas proteínas no sistema de expressão de mamíferos

[223] Pelo menos uma das seqüências foi utilizada para clonagem no vetor pCMV-script® (Stratagene). Esse é um vetor para expressão em células de mamíferos em cultura, CHO (células de ovário de hamster chinês), que usa o promotor de Citomegalovírus e o sítio de poliadenilação SV40, que permitem altas taxas de expressão constitutiva. Os produtos são expressos, purificados do sobrenadante de cultura e testados.

[224] A seqüência dos genes codificadores das proteínas da seda selecionada do grupo SEQ ID NO 1-19, assim como versões manipuladas modularmente dessas seqüências, foi inserida em um vetor baseado no promotor early de Citomegalovirus. Os vetores resultantes (Figura 5) foram utilizados na co-transfecção de células de ovário de hamster (Chinese Hamster Ovary - CHO) com uso de lipofectina e fostato de cálcio, juntamente

com os vetores repórteres pCMV-Gal (promega) e pGFP/NEO (promega). A integridade das proteínas de seda sob condições redutoras ou não redutoras foram avaliadas através de Western Blotting. Essa técnica visa a detecção de proteínas após a separação através de eletroforese em gel de poliacrilamida e transferência para membranas de nitrocelulose ou nylon. A detecção foi feita através de anticorpos, que reagem especificamente com epítomos da proteína de interesse, seguida por reações coloriméricas ou radiográficas .

[225] A produção de camundongos transgênicos contendo vetores de expressão com as seqüências dos genes das proteínas da teia de aranha foi realizada através da técnica da microinjeção pronuclear. Esta técnica é utilizada para a geração de animais transgênicos por adição. A técnica permite ainda a introdução de seqüências longas de DNA de diferentes espécies no genoma de mamíferos produzindo altos níveis de expressão e integração do transgene nas células germinativas. Utilizando-se um micromanipulador acoplado a um microscópio de alta resolução, cópias dos vetores de expressão contendo seqüências dos genes das proteínas da teia de aranha selecionadas do grupo SEQ ID NO 1-19 foram injetadas diretamente em um dos pronúcleos de embriões recém fertilizados coletados do oviduto de fêmeas doadoras superovuladas. Os pronúcleos são os núcleos materno e paterno provindos do óvulo e do espermatozóide, respectivamente, antes que se unam para formar um único núcleo contendo o genoma do novo indivíduo. Após a microinjeção os embriões foram transferidos para o oviduto de uma fêmea receptora pseudográvida previamente acasalada com um macho vasectomizado que leva a termo o nascimento dos possíveis transgênicos posteriormente genotipados quanto à presença do gene exógeno. A integração do transgene por microinjeção pronuclear ocorre de forma aleatória no genoma e todas as células do animal são geneticamente modificadas, inclusive as germinativas, de modo que esta alteração seja transmitida aos seus descendentes. Todo o animal transgênico positivo originado de um embrião microinjetado foi classificado como um fundador de uma linhagem transgênica única que difere de outro fundador quanto ao local de inserção e número de cópias do transgene no genoma. O protocolo detalhado de manipulação dos animais para a adição de um gene está bem descrito no estado da técnica e pode ser encontrado em manuais e revisões da literatura (Hogan, B., Beddington, R., Costantini, F., Lacy, E. (1994). Manipulating the

mouse embryo: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor; Godard, A.L.B., Guénet, J. (1999) Genética de Camundongos. Modelos animais de doenças humanas. Biotecnologia, Ciência & Desenvolvimento 9:96-100).

[226] A produção de bovinos transgênicos foi utilizada para a efetiva produção de proteínas das sedas de aranhas no leite. Foram utilizados vetores de expressão derivados do pBC1 (Invitrogen) para expressão de proteínas recombinantes no leite de animais transgênicos. Genes das proteínas da teia de aranha selecionadas do grupo SEQ ID NO 1-19, foram clonados no sítio de XhoI destes vetores (Figura 6), que apresentam em sua constituição promotores que dirigem a expressão das proteínas para o leite, tais como o promotor da beta-caseína e constitutivo (Iguma L.T., Lisauskas S. F.C., Melo E. O., Franco M. M., Pivato I., Vianna G. R., Sousa R.V., Dode M.A.N., Aragão F.J.L., Rech E.L., Rumpf R. (2005) Development of bovine embryos reconstructed by nuclear transfer of transfected and non-transfected adult fibroblast cells. Genet. Mol. Res. 4: 55-66; Oliveira R. R., Carvalho D. M. de, Lisauskas S., Mello E., Vianna G.R., Dode M. A.N., Rumpf R., Aragão F.J.L., Rech E. L. (2005) Effectiveness of liposomes to transfect livestock fibroblasts. Genet. Mol. Res. 4:185-196).

Exemplo 10: Síntese das proteínas da teia de aranha

[227] As proteínas da teia de aranha da presente invenção foram sintetizadas utilizando a técnica síntese automática em fase sólida F-moc em sintetizador Pioneer da Perseptive Biosystems. As proteínas foram purificadas em cromatógrafos líquidos Shimadzu Class VP e Akta Explorer utilizando colunas de fase reversa. Proteínas e peptídeos foram analisados por espectrometria de massa MALDI-TOF e seqüenciados por MS/MS utilizando espectrômetros Voyager DE STR, 4700 Proteomics Analyser e Q-TOF, respectivamente. Os lipossomos de composição fosfolipídica variável foram preparados seguindo instruções de Kits comercialmente disponíveis da GIBCO BRL (USA).

Exemplo 11: Análise das fibras da teia de aranha através de microscopia eletrônica de varredura

[228] Para se determinar detalhes ultraestruturais (espessura, comprimento, ornamentação externa, disposição de camadas protéicas, etc.) das fibras da teia de aranha foi realizado um estudo de microscopia eletrônica de varredura. Para tanto, amostras das fibras foram fixadas, desidratadas em ponto crítico, montadas em stub e metalizadas com diferentes ligas metálicas e observadas utilizando um microscópio Zeiss DSM 962 de acordo com metodologia adaptada de Bozzola & Russel (Bozzola, J.J.; Russel, L.D. (1999). Electron microscopy: principles and techniques for biologists. Jones and Bartlett Publishers, Inc. (eds). London, UK. pp.670-678).

Exemplo 12: Isolamento e caracterização das proteínas da teia de aranha por cromatografia líquida de alta eficiência (HPLC) e por cromatografia líquida capilar acoplada a espectrometria de massa (CapLC/Q-TOF/MS/MS; TOF-TOF/MS/MS)

[229] Os extratos protéicos de interesse foram fracionados em sistema HPLC e as frações obtidas foram digeridas enzimaticamente submetidas a cromatografia líquida capilar acoplada à espectrometria de massa. Dessa forma foram obtidas seqüências internas das proteínas isoladas, as quais foram identificadas, caracterizadas físico-quimicamente e utilizadas para buscas em bancos de dados (Pegah R., Dass J.C. (2004). Proteome analysis in the bovine adrenal medulla using liquid chromatography with tandem mass spectrometry. Rapid Commun. Mass Spectrom. 18:1877–1884).

Exemplo 13: Avaliação e quantificação da pureza das proteínas da seda de aranha

[230] A pureza das proteínas foi avaliada utilizando HPLC e composição de aminoácidos. A identidade das proteínas foi confirmada através de Western Blot.

[231] O material purificado foi quantificado utilizando o método e coeficiente de extinção (a 280 nm) e ELISA, com anticorpos policlonais.

Exemplo 14: Preparação e polimerização dos biopolímeros

[232] Amostras de proteínas da glândula de teia de aranha expressas em *E. coli*, purificadas, dialisadas e liofilizadas foram preparadas para polimerização e extrusão. Entre 60-90 mg de proteína foram solubilizadas em 200 μ l de hexafluoroisopropanol (HFIP), isopropanol ou etanol e mantidas overnight sob constante agitação à temperatura ambiente. O material insolúvel foi removido através de centrifugação e o sobrenadante foi colocado em uma seringa adaptada a um tubo usado para HPLC de diâmetro 4-300 μ m. Após a retirada de todo o ar do sistema da seringa, foi feita a extrusão em um banho de coagulação em recipiente contendo isopropanol/ metanol ou etanol e outros solventes orgânicos, e houve a formação de fibras semelhantes às fibras naturais (ver um exemplo anexo 1: seda produzida in vitro). Esse processo foi adaptado da metodologia descrita por Seidel (Seidel A. (1998) Artificial spinning of spider silk. *Macromolecules* 31:6733-6736).

Exemplo 15: Testes dos biopolímeros

[233] Um teste de mecânica estática quasi foi realizado no equipamento Instron 55R4201, a 23°C e 50% de umidade relativa. As propriedades mecânicas foram determinadas utilizando o software Instron serie IX para teste de materiais. Para testes de alta resistência, o equipamento Hopkinson Tension tem sido identificado como apropriado para as análises de sensibilidade de diferentes materiais (Shim V.P.W. (2001) Dynamic mechanical properties of fabric armor. *Int. J. of Impact Eng.* 25:1-15; Huh H., Kang W.J., Han S.S. (2002). A tension split hopkinson bar for investigating dynamic behavior of sheets metals. *Exp. Mechanics* 42:8-17). Os filamentos foram alinhados e pré-tensionados para assegurar a uniformidade da distribuição durante as avaliações.

Exemplo 16: Proteínas recombinantes da glândula produtora de seda de aranha utilizadas como defensinas, peptídeos antimicrobianos e microbicidas

[234] Proteínas recombinantes de sedas de aranhas foram utilizadas para inibir a replicação de diferentes vírus e outros microrganismos patogênicos de vegetais e animais. Isso pode acontecer por meio de vários mecanismos como a ligação de proteínas recombinantes da glândula da seda carregadas negativamente que se ligam a cargas da capa protéica de diferentes vírus, inibindo sua replicação, funcionando assim como microbicidas (Scordi-Bello I. A., Mosoian A., He C., Chen Y., Jarvis, Marla G. A., Keller J., Hogarty K., Waller D. P., Profy A. T., Herold B. C., Klotman M. E. 2005. Candidate Sulfonated and Sulfated Topical Microbicides: Comparison of Anti-Human immunodeficiency Virus activities and Mechanisms of Action. *ANTIMICROBIAL AGENTS AND CHEMOTHERAPY*, 49: 3607–3615), defensinas (Thevissen K., Francois I.E.J.A., Alberts A.M., Cammue B.P.A. (2005) Fungal sphingolipids as targets for the development of selective antifungal therapeutics. *Current Drugs Targ.* 6:923-928) e peptídeos ou polipeptídeos antimicrobianos (Prates M. V., Sforça M. L., Regis W.C. B., Leite J.R. S. A., Silva L. P., Pertinhez H. A., Araujo A. L. T., Azevedo R.B., Spisni A., Bloch C. Jr. (2004) The NMR-derived Solution Structure of a New Cationic Antimicrobial Peptide from the Skin Secretion of the Anuran *Hyla punctata*. *J. Biol. Chemistry* 279:13018-13026). Por isso, as proteínas da glândula produtora de seda foram expressas e testadas quanto à esse tipo de atividade, mostrando resultados positivos para inibição do crescimento de vírus, fungos e bactérias.

Exemplo 17: Modelagem molecular das proteínas da teia de aranha

[235] A predição de estruturas tridimensionais das proteínas de glândulas de teia de aranha foi feita pela modelagem por homologia ("comparative protein modeling"), que baseia-se na observação de que a homologia entre seqüências de aminoácidos implica em semelhança estrutural e funcional e de que proteínas homólogas apresentam regiões

internas conservadas (principalmente constituídas de elementos de estrutura secundária α -hélices e folhas β).

[236] A modelagem dessas proteínas por homologia seguiu basicamente quatro passos sucessivos:

[237] - Identificação e seleção de proteínas-molde com estrutura tridimensional conhecida diretamente do PDB (Protein Data Bank). Foi feita uma busca sistemática por BLAST³⁶ de um ou mais moldes adequados utilizando a sequência primária das proteínas como isca em um banco de dados de estruturas primárias derivadas de proteínas com estruturas terciárias resolvidas depositadas no PDB.

[238] - Alinhamento das sequências de resíduos de aminoácidos. O objetivo do alinhamento é alinhar resíduos estruturalmente equivalentes levando em conta características estruturais comuns, tais como elementos de estrutura secundária. Com isso é possível reconhecer regiões estruturalmente conservadas e regiões variáveis.

[239] - Construção do modelo. Esta etapa passa pela modelagem das regiões estruturalmente conservadas, pela modelagem das regiões de alças e pela modelagem das cadeias laterais, e foi feita utilizando-se o programa Blue Star Sting (Goran Neshich, Ivan Mazoni, Stanley R. M. Oliveira, Michel E. B. Yamagishi, Paula R. Kuser-Falcão, Luiz C. Borro, Douglas U. Morita, Kassius R. R. Souza, Gustavo V. Almeida, Diego N. Rodrigues, José G. Jardine, Roberto C. Togawa, Adauro L. Mancini, Roberto H. Higa, Sérgio A. B. Cruz, Fábio D. Vieira, Edgard H. dos Santos, Raquel C. de Melo and Marcelo M. Santoro. The Star STING server: A multiplatform environment for protein structure analysis. Genet. Mol. Res. 5 (2) 2006).

[240] - Validação do modelo. Uma proteína adequadamente modelada deve ter uma estrutura terciária satisfatória. Sua qualidade depende da proteína escolhida como molde e do alinhamento calculado. É importante verificar se existem grandes diferenças conformacionais não explicadas entre os elementos de estrutura secundária (regiões conservadas) das estruturas-molde e da estrutura-modelada. A validação deste exemplo foi feita utilizando-se o programa PROCHECK (Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M.; J. Appl. Crystallogr. 1993, 26, 283).

[241] Esses passos foram seguidos a partir da SEQ ID NO 1 (Produto do Gene: **Nephilengys cruentata** – NCFlag), e o modelo obtido segue na Figura 7.

[242] Não havendo estrutura atômica ainda conhecida para as proteínas das sedas e das estruturas modulares, serão construídas as geometrias ab initio (predição da estrutura tridimensional de proteínas a partir de sua estrutura primária), baseando-se nos conteúdos das estruturas secundárias e nas hipóteses relativas ao dobramento das proteínas. Porções como o N- terminal e C-terminal que não contém os blocos de seqüências modulares terão sua estrutura determinada através de cristalografia de Raio-X. As proteínas purificadas serão cristalizadas através da otimização dos resultados das matrizes. Possuindo as características de novas proteínas, as estruturas serão determinadas utilizando: múltiplas substituições das isoformas (MIR) após tratamento dos átomos ou através de Dispersão anômala dos cristais contendo selenometionina.

REIVINDICAÇÕES

1. Gene quimérico caracterizado por compreender:
 - a) um promotor opcionalmente ligado a uma sequência líder e operacionalmente ligado a;
 - b) uma sequência codificadora selecionada do grupo SEQ ID NO 1 a 19.
2. Gene quimérico de acordo com a reivindicação 1 caracterizado pelo fato de o promotor ser selecionado do grupo consistindo de constitutivos, induzíveis e tecido-específicos.
3. Gene quimérico de acordo com a reivindicação 2 caracterizado pelo fato de o promotor tecido-específico ser selecionado entre promotores de genes de fibra de algodão.
4. Gene quimérico de acordo com a reivindicação 3 caracterizado pelo fato de os promotores de genes de fibra de algodão serem selecionados do grupo consistindo de E6, H6S, Rac13, LTP, ACP, Expansina, CAP, Anexina, FbL2A e actina 2.
5. Gene quimérico de acordo com a reivindicação 1 caracterizado pelo fato de o promotor conter elementos potencializadores.
6. Gene quimérico de acordo com a reivindicação 1 caracterizado pelo fato de o promotor poder ser expresso em plantas, animais, bactérias, fungos ou insetos.
7. Gene quimérico de acordo com a reivindicação 1 caracterizado pelo fato de a sequência líder ser obtida do mesmo gene que o promotor utilizado para dirigir a transcrição da molécula isolada de ácido nucleico.
8. Vetor de expressão caracterizado por conter um gene quimérico de acordo com qualquer uma das reivindicações 1 a 7.
9. Vetor de expressão caracterizado por compreender:
 - a) um promotor opcionalmente ligado a uma sequência líder e operacionalmente ligado a;
 - b) uma sequência codificadora selecionada do grupo SEQ ID NO 1 a 19 operacionalmente ligada a;
 - c) um sinal de terminação;

- d) uma origem de replicação;
 - e) um marcador seletivo; e
 - f) um sítio de clonagem
10. Vetor de expressão de acordo com a reivindicação 9 caracterizado pelo fato de o promotor ser selecionado do grupo consistindo de constitutivos, induzíveis e tecido-específicos.
 11. Vetor de expressão de acordo com a reivindicação 10 caracterizado pelo fato de o promotor tecido-específico ser selecionado entre promotores de genes de fibra de algodão.
 12. Vetor de expressão de acordo com a reivindicação 11 caracterizado pelo fato de os promotores de genes de fibra de algodão serem selecionados do grupo consistindo de E6, H6S, Rac13, LTP, ACP, Expansina, CAP, Anexina, FbL2A e actina 2.
 13. Vetor de expressão de acordo com qualquer uma das reivindicações 8 a 12 caracterizado pelo fato de o promotor conter elementos potencializadores.
 14. Vetor de expressão de acordo com a reivindicação 9 caracterizado pelo fato de a sequência líder ser obtida do mesmo gene que o promotor utilizado para dirigir a transcrição da molécula isolada de ácido nucleico.
 15. Vetor de expressão de acordo com a reivindicação 14 caracterizado pelo fato de a sequência líder ser obtida do mesmo gene que o promotor utilizado para dirigir a transcrição da molécula de ácido nucleico.
 16. Vetor de expressão de acordo com a reivindicação 9 caracterizado pelo fato de o promotor poder ser expresso em plantas, animais, bactérias, fungos ou insetos.
 17. Vetor de expressão de acordo com a reivindicação 9 caracterizado pelo fato de o sinal de terminação ser selecionado do grupo sinal de terminação de SV40, sinal de adenilação de HSV TK, sinal de terminação do gene da nopalina sintetase de *Agrobacterium tumefaciens* (NOS), sinal de terminação do gene da octopina sintetase, sinal de terminação do gene 19S e 35S do CaMV, sinal de terminação do gene da álcool desidrogenase de milho, sinal de terminação do gene da manopina sintetase, sinal de terminação do gene da beta-faseolina, sinal de terminação do gene da ssRUBISCO, sinal de terminação do gene da sucrose sintetase, sinal de

terminação do vírus que ataca o *Trifolium subterranean* (SCSV), sinal de terminação do gene *trpC* de *Aspergillus nidulans* e outros semelhantes.

18. Vetor de expressão de acordo com a reivindicação 9 caracterizado pelo fato de o marcador seletivo ser selecionado entre seqüências que conferem resistência a antibióticos ou marcadores visuais.
19. Vetor de expressão de acordo com a reivindicação 18 caracterizado pelo fato de o marcador seletivo ser selecionado entre as seqüências codificadoras dos genes canamicina, neomicina, ampicilina, cloranfenicol, estreptomicina, higromicina, geneticina, fosfinotricina, glifosato, glufosinato de amônio, AHAS, BAR e GUS.
20. Célula transformada oriunda de qualquer um dos grupos consistindo de bactérias ou fungos caracterizada pelo fato de conter um gene quimérico de acordo com qualquer uma das reivindicações 1 a 7.
21. Célula transformada oriunda de qualquer um dos grupos consistindo de bactérias ou fungos caracterizada pelo fato de conter um vetor de expressão de acordo com qualquer uma das reivindicações 9 a 20.
22. Célula transformada oriunda de qualquer um dos grupos consistindo de bactérias ou fungos caracterizada pelo fato de conter uma molécula de proteína recombinante compreendendo qualquer uma das seqüências selecionadas do grupo identificado como SEQ ID NO 20 a 38.
23. Microrganismo ou uma parte do mesmo caracterizado por compreender um gene quimérico de acordo com qualquer uma das reivindicações 1 a 8.
24. Microrganismo ou uma parte do mesmo caracterizado por compreender um vetor de expressão de acordo com qualquer uma das reivindicações 8 a 19.
25. Método para produzir um organismo geneticamente modificado caracterizado pelo fato de compreender as seguintes etapas:
 - a) transformar uma célula, tecido, órgão ou embrião com um gene quimérico de acordo com qualquer uma das reivindicações 1 a 7 ou um vetor de expressão de acordo com qualquer uma das reivindicações 8 a 19;
 - b) selecionar células transformadas, calos de células, embriões ou sementes;

- c) regenerar plantas maduras, embriões maduros ou microorganismos de células transformadas, calos de células, embriões ou sementes selecionados na etapa (b);
 - d) selecionar plantas maduras, embriões maduros ou células de microorganismos da etapa (c) contendo o gene quimérico ou vetor de expressão com as seqüências de nucleotídeos que codificam a proteína da seda da aranha.
26. Método para produção de proteína recombinante caracterizado pelo fato de compreender as seguintes etapas:
- a) transformar uma célula, tecido, órgão ou embrião com um vetor de expressão de acordo com qualquer uma das reivindicações 8 a 19;
 - b) selecionar células transformadas, calos de células, embriões ou sementes;
 - c) regenerar plantas maduras, embriões maduros ou microorganismos de células transformadas, calos de células, embriões ou sementes selecionados na etapa (b);
 - d) selecionar plantas maduras, embriões maduros ou células de microorganismos da etapa (c) contendo vetor de expressão com as seqüências de nucleotídeos que codificam a proteína da seda da aranha;
 - e) Fazer a extração da proteína recombinante da seda de aranha produzida nos organismos selecionados na etapa (d).
27. Composição dermatológica caracterizada pelo fato de compreender:
- a) Proteína recombinante obtida pelo método descrito na reivindicação 26;
 - b) Veículo farmaceuticamente aceitável.
28. Composição dermatológica de acordo com a reivindicação 27 caracterizada pelo fato de o veículo farmaceuticamente aceitável ser selecionado do grupo consistindo de glicerol, água, solução salina, etano, solução com fosfatos e sais de ácidos orgânicos.
29. Composição microbicide caracterizada pelo fato de compreender:
- a) Proteína recombinante obtida pelo método descrito na reivindicação 26 com atividade microbicide;

b) Veículo agriculturalmente aceitável e, opcionalmente,

c) Aditivos

30. Composição microbicida de acordo com a reivindicação 29 caracterizada pelo fato de o veículo agriculturalmente aceitável ser selecionado do grupo consistindo de água, solventes orgânicos, umectantes, preservativos, espessantes, agentes antimicrobianos, antioxidantes, emulsificadores, polímeros formadores de filme e misturas destes.
31. Composição microbicida de acordo com a reivindicação 30 caracterizada pelo fato de os aditivos serem selecionados do grupo consistindo de goma rosina, látex, pirrolidona polivinil, álcool polivinil, cloreto polivinil, polietileno, acetato de polivinil, metil, metalcrilato, e misturas destes.
32. Biopolímeros produzidos a partir da proteína recombinante obtida de acordo com o método descrito na reivindicação 26.

[illegible]

FIGURA 1

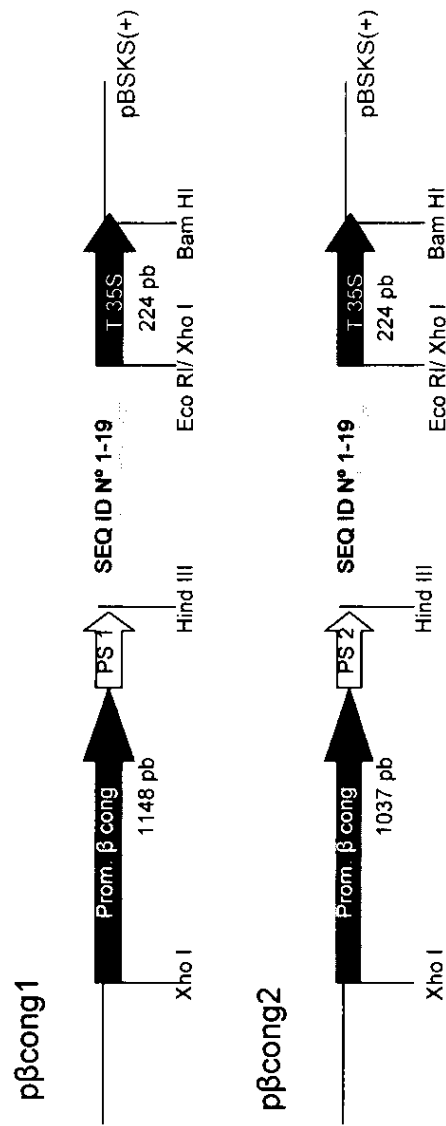


FIGURA 2

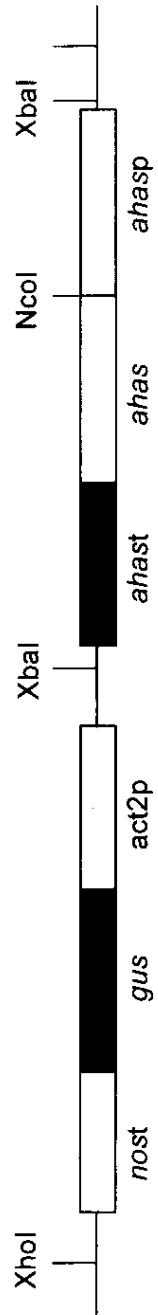


FIGURA 3

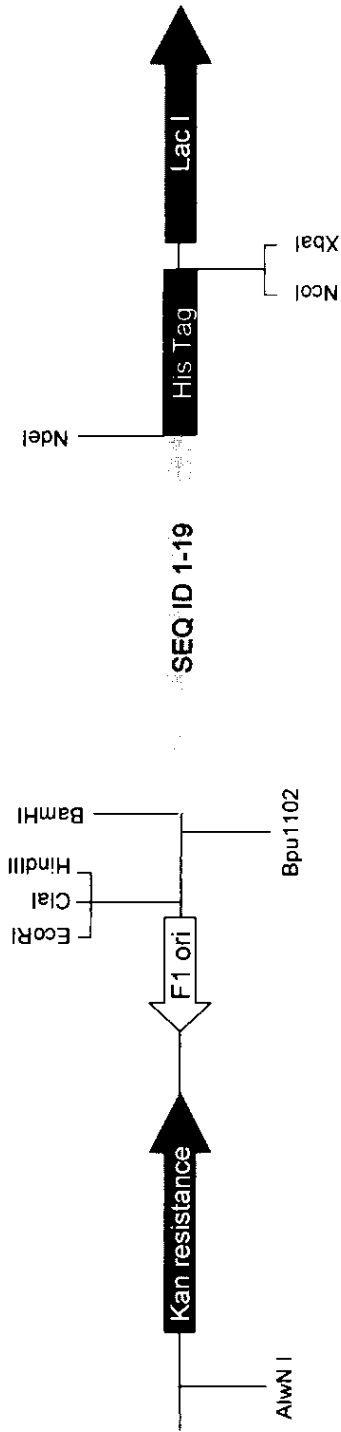


FIGURA 4

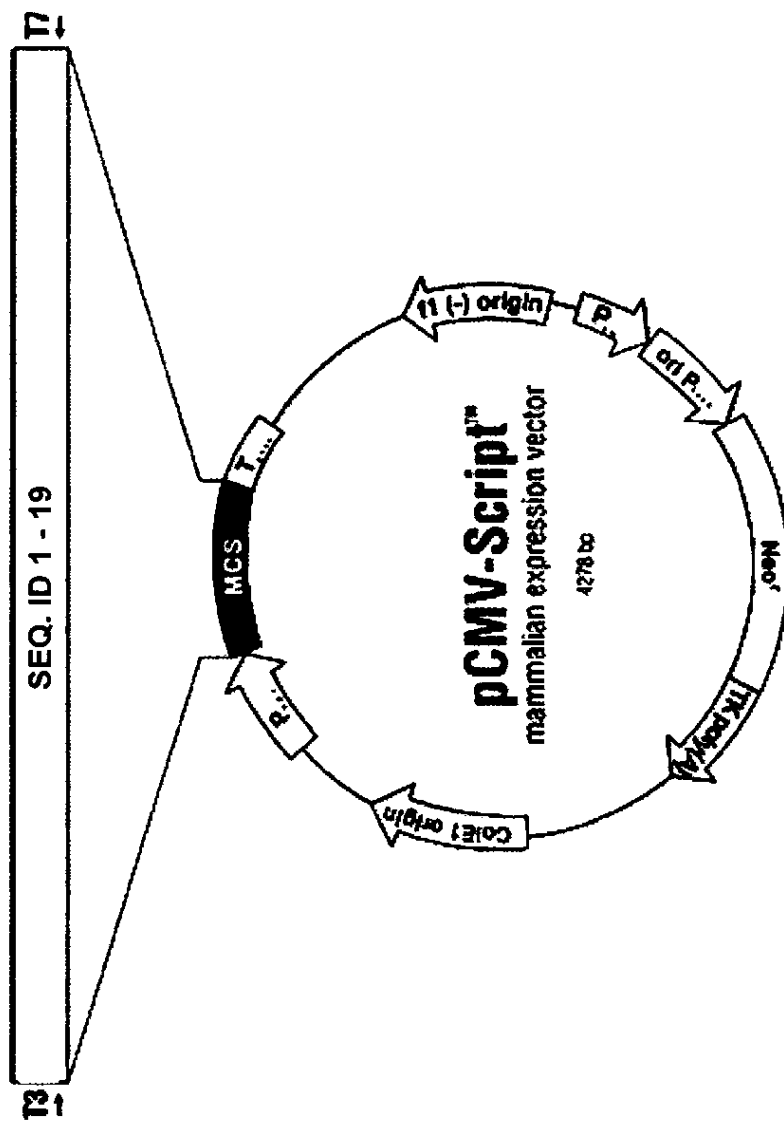
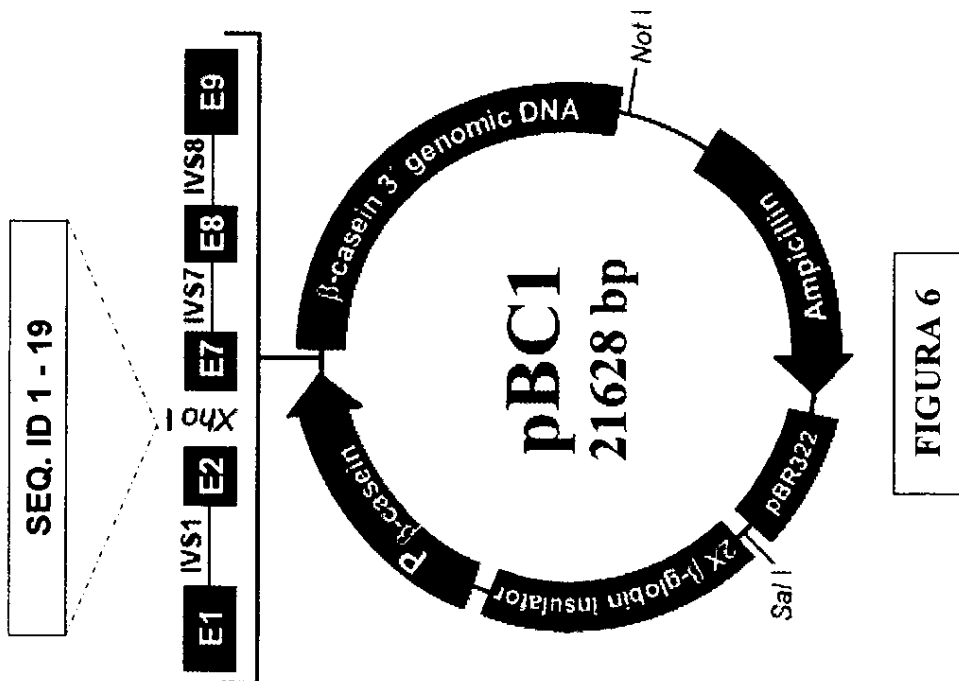
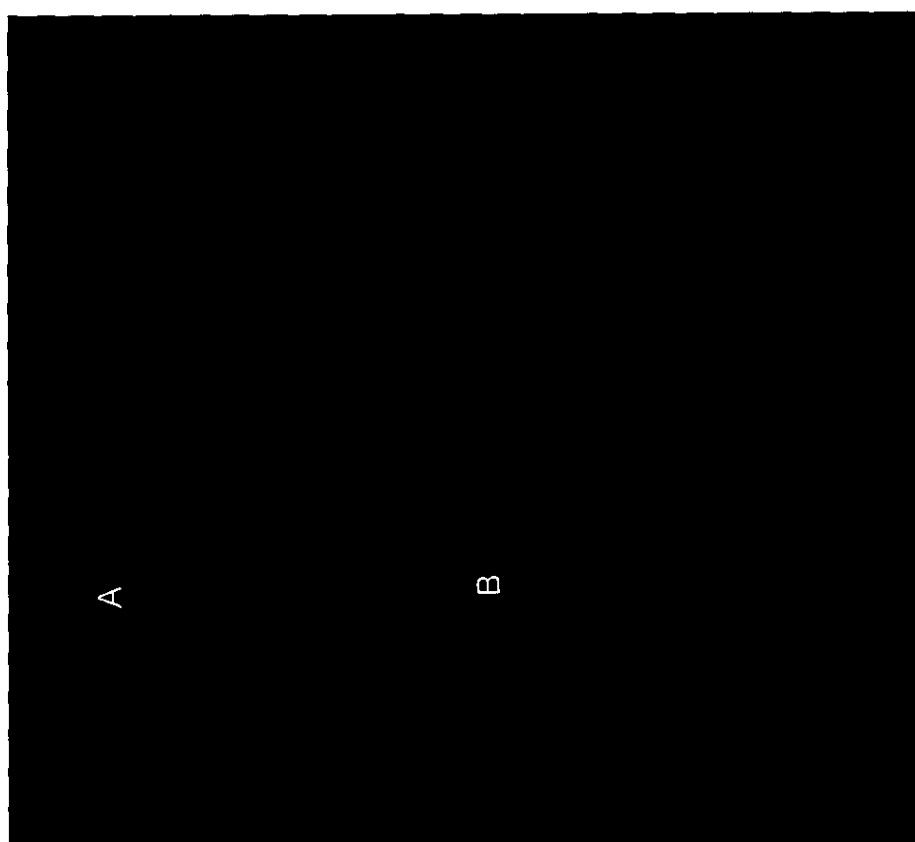


FIGURA 5





A

B

FIGURA 7