

## Biologia computacional molecular e suas aplicações na agricultura

José Gilberto Jardine  
Izabella Agostinho Pena Neshich  
Ivan Mazoni  
Inácio Henrique Yano  
Fábio Rogério de Moraes  
Jose Augusto Salim  
Luiz Borro  
Leticia Sayuri Nishimura  
Goran Neshich

### 1 Biologia computacional, uma nova ciência aplicada

Uma definição abrangente para Biologia Computacional envolve qualquer técnica computacional aplicada que busca solucionar problemas da Biologia. A partir dessa definição podemos dizer que a Biologia Computacional Molecular é um ramo da biotecnologia calcada nos avanços da Biologia, da Química, da Bioquímica e Biofísica que busca, a partir de ferramentas oferecidas pela Ciência da Computação, Matemática aplicada e Estatística, oferecer uma percepção transdisciplinar de aspectos relacionados a sequências de nucleotídeos e aminoácidos, a estrutura e dinâmica de proteínas e a interação proteína-proteína, proteína-DNA, proteína-ligante.

Dentre as áreas de abrangência da Biologia Computacional, tais como bioinformática que abriga bancos de dados sobre sequências de DNA, RNA e proteínas; biomodelagem computacional, um campo da biocibernética para a construção de modelos computacionais de sistemas biológicos; simulação molecular, que lida com métodos teóricos e técnicas computacionais para modelar o comportamento de biomoléculas; biologia sistêmica, que modela redes de interação biológica, destaca-se a área que trata especificamente do planejamento e desenho de novas drogas, fármacos e agroquímicos ou agro defensivos in silico.

A agricultura dos países do Cone Sul que desempenha papel de destaque na economia mundial, sendo responsável por grande parte da renda desses países, está faminta por soluções inovadoras que propiciem aumento da produção e, principalmente, da produtividade agrícola. A geração de um inventário de novos agroquímicos e fármacos pode ser um dos mais importantes desafios perante o agronegócio desses países, representando um dos fatores fundamentais de conquista de vantagens competitivas no mercado internacional. Para ingressar nesta área é necessário lembrar que fornecer anotação funcional para a vasta quantidade de dados de sequências e estruturas e as funções das proteínas, particularmente das enzimas, geradas por tecnologias de alto desempenho em larga escala, é uma das principais tarefas da era pós genômica. Caminhando na direção do planejamento de fármacos, é fundamental a identificação dos resíduos catalíticos de uma enzima, sendo esse um passo importante no entendimento do seu papel biológico e suas aplicações,

principalmente devido ao fato de que um pequeno número de resíduos situado no sítio catalítico é responsável pela função enzimática, e participa diretamente do processo de catálise. Os arranjos espaciais, bem como as propriedades físicas, químicas e físico-químicas destes resíduos determinam a reação química catalisada pela enzima.

Os fatores que determinam a funcionalidade do sítio ativo de uma proteína são muito complexos e dependem de sua estrutura tridimensional, além das propriedades bioquímicas e biofísicas. Sítios funcionais podem ser entendidos como uma redondeza física, química e físico-química que acompanha uma função, em contraposição a um grupo de resíduos fixos que não participam diretamente, porém, providenciam o arcabouço estrutural. Um melhor entendimento deste complexo processo tem um significativo impacto no desenvolvimento de novas drogas, bem como na identificação de desordens genéticas e na engenharia de proteínas com novas funções.

Diversos métodos matemáticos e estatísticos têm sido propostos para calcular informações sobre as propriedades estruturais, físicas, químicas e físico-químicas dos aminoácidos (e.g. energia de contatos, potencial eletrostático, hidrofobicidade, densidade). Estas propriedades são denominadas descritores estruturais de proteínas, e tais propriedades permitem realizar diversas análises segundo diferenças e semelhanças entre aminoácidos de várias regiões de uma proteína. Além de descritores físicos, químicos e físico-químicos, existem descritores baseados na conservação da estrutura primária da proteína, ou seja, baseiam-se em evidências evolutivas obtidas a partir da comparação entre sequências de aminoácidos das proteínas homólogas, determinando quais aminoácidos são mais frequentes em certas posições da sequência, o que sugere sua importância na manutenção da funcionalidade da proteína. Outro descritor utilizado em sistema de predição de aminoácidos posicionados em determinados distritos proteicos é a propensão (*propensity*, em inglês), definida como sendo a porcentagem entre o número de aminoácidos de interesse e aqueles que não têm preferência para determinada posição, para cada um dos vinte tipos de aminoácidos.

O planejamento e desenvolvimento de novos fármacos e agroquímicos, modificadores de função biológica das macromoléculas cruciais para a patogenicidade de microrganismos, por exemplo, que servem para proteção das plantas, animais e humanos, é um processo bastante complexo e longo e demanda um investimento contínuo para que se possa chegar até o resultado/produto com potencial para ser oferecido ao mercado (KUNTZ, 1982).

## 2 Desenho racional de drogas, fármacos e agroquímicos

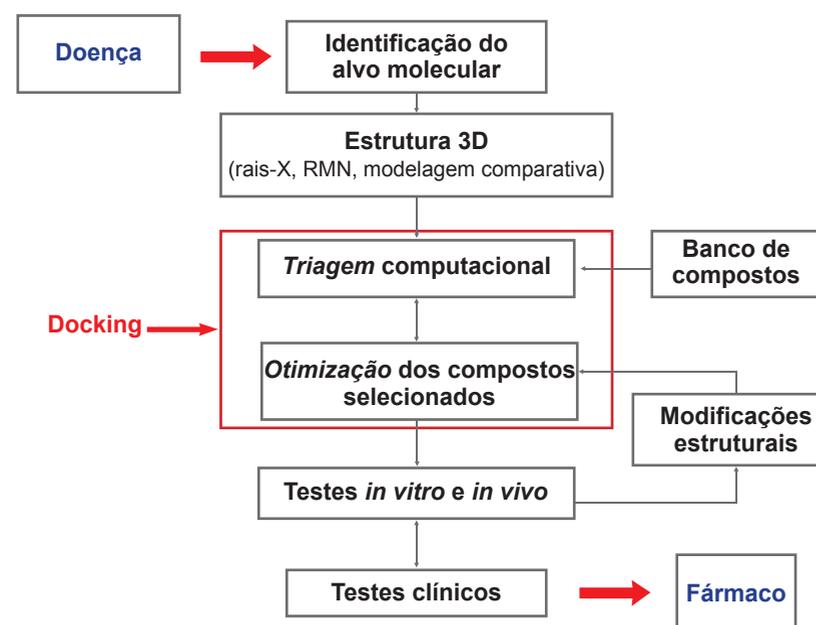
O processo para se obter novos fármacos tem mudado através dos anos. Há bem pouco tempo, novos fármacos eram resultado de testes aleatórios chamados de triagem cega (*screening*) em células, animais, plantas ou modelos destes. Esse método tradicional consiste em testar aleatoriamente várias micromoléculas sem nenhum conhecimento dos mecanismos de interação da micromolécula com a molécula ligante. Esse método é ineficiente em função de que a probabilidade é cada vez menor de se identificar um novo fármaco. A necessidade de se testar em bancada de laboratório *in vitro* e posteriormente *in vivo* milhões de compostos com, conseqüentemente, custo e tempo elevados, apesar de ter sido o método pelo qual a maioria dos medicamentos hoje disponíveis foram desenvolvidos, tornou-se num grande problema (ANDERSON, 2003).

Esses fatos implicaram a necessidade de desenvolvimento de uma metodologia sistemática objetivando o desenho racional de fármacos baseado na estrutura tridimensional do receptor das

pequenas moléculas (ligantes ou inibidoras). O desenho racional de fármacos fundamenta-se no estudo de estruturas moleculares tridimensionais da molécula receptora para o desenho de compostos protótipos, tomando como base as informações estruturais e as interações envolvidas no processo de reconhecimento molecular receptor-ligante (YANG, 2010).

Metodologias de atracamento proteína-ligante (*docking* proteína-ligante) são amplamente utilizadas tanto para a descoberta de novas substâncias bioativas, técnica conhecida como triagem virtual (*virtual screening*), quanto para o aperfeiçoamento de compostos bioativos já identificados (GUEDES et al., 2014). A metodologia de desenho racional de fármacos baseado em estrutura tridimensional consiste de seis etapas básicas, como pode ser visualizado esquematicamente na Figura 1, que são:

- 1) Escolha adequada do alvo terapêutico (proteína ou enzima) relacionada à doença cuja função deva ser bloqueada ou ativada.
- 2) Obtenção da estrutura tridimensional do biorreceptor através de técnicas experimentais, tais como difração de raios-X, ressonância magnética nuclear ou então por modelagem por homologia. Hoje existem bancos de dados de estruturas tridimensionais de moléculas de acesso livre, como o Protein Data Bank (PDB)<sup>1</sup> e bancos de dados de estruturas moleculares de moléculas ligantes, como o Maybridge e a Cambridge Structural Database (CSD).
- 3) Triagem *in silico* com *virtual screening docking* (esse procedimento não é diferente da metodologia tradicional de triagem às cegas). A diferença é que agora a busca de micromoléculas



**Figura 1.** Fluxograma com as etapas básicas do desenho racional de fármacos baseado em estrutura. Fonte: Magalhães (2006).

<sup>1</sup> Disponível em: < <http://www.rcsb.org/pdb/home/home.do> >.

candidatas é feita com os recursos computacionais, sem gasto de materiais e a tempo reduzidíssimo.

- 4) Identificada uma molécula promissora, o próximo passo consiste na análise por um profissional especialista para identificar as alterações que serão necessárias na molécula para se obter a resposta biológica desejada. Modificações subsequentes são realizadas para aprimorar a molécula ligante e torná-la mais específica para um determinado alvo e ajustes farmacocinéticos como absorção, distribuição, metabolismo e eliminação. Nessa fase são utilizadas metodologias de *docking* mais acuradas, para a identificação da conformação de ligação das moléculas selecionadas e a otimização dos compostos modificados estruturalmente por proposição do especialista.
- 5) O penúltimo passo consiste na síntese laboratorial da molécula candidata e teste *in vitro*.
- 6) Finalmente, são requeridos testes *in vivo* para a análise do comportamento do fármaco e sua toxicidade (MAGALHÃES, 2006; MAGALHÃES et al., 2007).

### 3 Biologia computacional na Embrapa

No Laboratório do Grupo de Pesquisa em Biologia Computacional (GPBC) da Embrapa Informática Agropecuária, Unidade temática de pesquisa da Empresa Brasileira de Pesquisa Agropecuária (Embrapa), que tem como missão institucional viabilizar soluções de pesquisa, desenvolvimento e inovação em tecnologia da informação para a sustentabilidade da agricultura, foi desenvolvido o software Sting e o maior banco de dados de descritores físicos, químicos, físico-químicos, estruturais e biológicos sobre estruturas proteicas, o Sting\_DB (Neshich et al., 2003), disponível na sua versão on-line<sup>2</sup>.

Blue Star Sting é uma suíte de programas com ferramentas para a visualização e análise estrutural de proteínas. Estes programas (módulos) estão concentrados em um único pacote que visa oferecer um instrumento para estudos das macromoléculas, suas estruturas e as relações estrutura-função. Informações como posição dos aminoácidos na sequência e na estrutura, busca por padrões, identificação de vizinhança, ligações de hidrogênio, ângulos e distância entre átomos, são facilmente obtidas, além de dados sobre a natureza e volume dos contatos atômicos inter e intra-cadeias, conservação e relação entre os contatos intra-cadeia e parâmetros funcionais (Neshich et al., 2006). Principalmente pelo fato do Blue Star Sting oferecer fácil acesso a um rico repositório de características da proteína, a plataforma Sting (Neshich et al., 2004, 2005, 2006; MANCINI et al., 2004) já foi utilizada para prever classe de enzimas (BORRO et al., 2006) em análise de proteínas ligante (FERNANDEZ et al., 2003; FREITAS et al., 1997), análise de proteína mutantes (MARCELLINO et al., 1996; SIMÕES et al., 2007), análise de padrões de interação proteína-proteína, bem como em pesquisas ligadas a alguns problemas biológicos específicos (BRAGHINI et al.; DIAS-LOPES et al., 2013).

A partir de um conjunto de enzimas com seus resíduos catalíticos devidamente identificados (e.g. através de experimentos de mutagênese) como estudo de caso, um modelo pode ser construído

usando métodos de aprendizagem de máquina, para caracterizar o nano-ambiente destes resíduos segundo suas propriedades físicas, químicas e físico-químicas através de descritores de proteínas extraídos do banco de dados Sting\_DB, e assim avaliar se tais modelos possuem capacidade para predição dos resíduos catalíticos de enzimas ainda não anotadas e também de novas enzimas.

Apesar de métodos baseados em estrutura limitarem-se às proteínas com estrutura conhecida, ou a modelos estruturais de boa qualidade, tais métodos são importantes ao entendimento de como as enzimas realizam sua função, uma vez que a função das proteínas é mais conservada na estrutura terciária do que na estrutura primária.

O Grupo de Pesquisa em Biologia Computacional (GPBC) da Embrapa Informática Agropecuária executa projetos de pesquisa em bioinformática estrutural. Decorrente do trabalho realizado nos últimos cinco anos dedicado à finalização da plataforma Sting, que serve para a análise teórica das estruturas proteicas e seus complexos e para construção dos fármacos e agro defensivos/agrotóxicos embasado no conceito de *rational drug design*, obteve várias soluções inovadoras para problemas de fundamental interesse para a agricultura brasileira e demais países que se beneficiam igualmente da agricultura.

O procedimento adotado pelo GPBC consiste, inicialmente, na escolha dos alvos terapêuticos e proteínas. A escolha dos alvos proteicos e, mais especificamente, dos resíduos a serem usados como alvos em um procedimento de desenho computacional de fármacos baseado em estruturas proteicas foi o passo inicial para o desenho computacional de fármacos que resultou em quatro pedidos de patentes já registradas no Instituto Nacional da Propriedade Industrial (INPI) e que mais adiante serão utilizados para ilustrar os processos de desenho de novos fármacos. Mais importante que a definição da proteína-alvo por si só é a definição dos resíduos que serão usados como alvo para o procedimento de SBDD (do inglês Structure Based Drug Design). Usando o banco de dados Catalytic Site Atlas (CSA)<sup>3</sup> (PORTER et al., 2004) são selecionados os resíduos do sítio ativo de cada proteína alvo, caso possuam atividade enzimática e/ou, usando o software SurfV são selecionados resíduos de interface envolvidos em interação proteína-proteína. A identificação dos resíduos formadores de interface (IFR) se dá pelo cálculo da diferença entre a área acessível ao solvente (ASA) da proteína isolada e complexada usando o programa SurfV (SRIDHARAN et al., 1992).

Através do pacote de programas e bancos de dados Sting e do Java Protein Dossier é feito um refinamento de conhecimento sobre o alvo, por exemplo, buscando por cavidades próximas aos resíduos de interesse e que possam alocar um fármaco inibitório. Com o pacote de programas Sting é possível calcular o número, o tipo e, portanto, a energia total dos contatos estabelecidos entre os resíduos que se localizam proximamente aos *pockets* (bolsos) e cavidades selecionadas, bem como nas interfaces entre cadeias adjacentes, contatos esses que possivelmente deverão ser vencidos pela interação com um ligante potencial. A realização da seleção dos resíduos a serem usados como alvos terapêuticos com base em características físicas, químicas, físico-químicas, valores altos de energia de contatos, polaridade e características estruturais como área exposta ao solvente, presença em *pocket*, dentre outras de interesse como ocorrência de sítio catalítico ou próximo a ele, é efetuada usando o módulo Select do Sting Java Protein Dossier. Os programas PyMol (PyMOL MOLECULAR GRAPHICS SYSTEM, 2014) e Accelrys Discovery Studio

<sup>2</sup> Disponível em: <[www.cbi.cnptia.embrapa.br/SMS](http://www.cbi.cnptia.embrapa.br/SMS)>.

<sup>3</sup> Disponível em: <<http://www.ebi.ac.uk/thornton-srv/databases/CSA/>>.

Visualizer (ACCELRY'S SOFTWARE INC., 2014) são utilizados em conjunto ao Sting para visualização e geração de imagens moleculares.

O grupo de resíduos-alvos selecionado de acordo com as características de interesse para cada proteína-alvo deve ser analisado detalhadamente em relação à sua ocorrência e conservação entre outros organismos. São considerados alvos preferenciais aqueles que ocorrerem, na mesma posição estrutural, em organismos patogênicos, mas não em organismos não patogênicos de vida livre. Esta ausência em tais organismos deve minimizar o impacto ecológico, por exemplo, em populações bacterianas que não causam danos para o hospedeiro. Esta abordagem é efetuada através de alinhamento entre a estrutura primária de proteínas homólogas à proteína de interesse através dos programas ClustalW 2.0 (LARKIN et al., 2007), Muscle (EDGAR, 2004), T-Coffee (Tree-based Consistency Objective Function For alignment Evaluation) (NOTREDAME et al., 2000), evidenciando as similaridades e diferenças entre as proteínas e buscando as correspondências.

Selecionados os resíduos, elabora-se um pipeline de desenho de fármacos baseado na estrutura proteica (SBDD). Técnicas de SBDD podem ser aplicadas com base em possíveis alvos terapêuticos preditos.

A seguir é apresentada uma exemplificação de um dos vários caminhos (pipeline) adotados pelo GPBC para o desenho de novos fármacos.

O procedimento tem início com o mapeamento de posições favoráveis de interações para grupos funcionais (em que posições podem ser desenhadas grupos hidroxila, amina, hidrofóbicos, cíclicos, ou mesmo pequenos fragmentos de moléculas). Novos compostos podem, então, ser desenhados de modo que relevantes grupos funcionais se localizem em posições que determinem uma correta relação/posição espacial com o sítio-alvo. Após o desenho deve ocorrer a modelagem de sua estrutura tridimensional, ensaios de *docking*, escolha dos melhores ligantes, predição das bases moleculares/estruturais de sua ligação. Existe uma série de softwares úteis para o desenho de fármacos e abordagens de *screening*, como o Sprout<sup>4</sup>, muito usado para o desenho baseado em fragmentos. Este programa inclui módulos para identificar e selecionar grupos funcionais e posições nos sítios-alvo para formar fragmentos iniciais de compostos para geração de estrutura (módulo EleFANT) e, conforme estes são selecionados, são gerados esqueletos que satisfaçam as restrições estéricas do *pocket*-alvo através de crescimento de fragmentos espaçadores e conectando-os aos fragmentos iniciais (módulo SPIDeR). Por fim, é feita a substituição de átomos no esqueleto até gerar moléculas que sejam compatíveis com as propriedades eletrostáticas do sítio-alvo (módulo Marabou). As soluções podem ser agrupadas e terem escores de ligação calculados usando o módulo ALLigaTOR.

Além do desenho de compostos, pode-se também realizar a busca por compostos que se liguem nos sítios-alvo preditos por meio de *virtual screening* em larga escala (virtual High Throughput Screening- vHTS), através do uso de bancos de dados de estruturas tridimensionais de pequenas moléculas. Um exemplo de banco de dados de estruturas de pequenas moléculas é o ChEMBL<sup>5</sup>, manualmente curado. A versão ChEMBL\_19 (setembro, 2014), possui cerca de 1,64 milhão de compostos que podem ser usados em *screening* contra um sítio-alvo. A simulação da interação

<sup>4</sup> Disponível em: <<http://www.simbiosys.ca/sprout/>>.

<sup>5</sup> Disponível em: <<https://www.ebi.ac.uk/chembl/>>.

proteína-ligante com o intuito de avaliar a capacidade de um ligante em formar interações fortes e de encaixar-se a um sítio-alvo é chamada *docking*. Algoritmos de *docking* tais como Dock; FlexX; AutoDock; Mol-Dock e Gold podem ser usados para realizar o *docking* na região de interesse para encontrar as moléculas que estabelecem interações mais favoráveis e que se adaptem melhor ao sítio-alvo, sendo escolhidas por meio do ranqueamento do escore de *docking*. É provável que compostos que inibam a função das proteínas envolvidas sejam encontrados dentre os melhores compostos preditos *in silico*.

Através da plataforma Discovery Studio da Accelrys, recentemente adquirida pelo GPBC, o processo de avaliação de candidatos a novos fármacos pode ser expandido para a análise da sua absorção potencial, distribuição, metabolismo e excreção pelo organismo hospedeiro e, finalmente, sua toxicidade ao organismo: *absorption, distribution, metabolism, excretion, toxicity* (ADMET). Uma análise completa, que atribui um fator favorável de ADMET ao ligante em potencial, possibilitará uma seleção mais eficaz dos compostos *leads*, ainda na fase inicial de descoberta e planejamento das novas drogas.

Em adição, pode-se valer do módulo do programa Accelrys que realiza estudos das relações quantitativas entre estrutura química e a função dos compostos escolhidos (Quantitative Structure-Activity Relationship - QSAR). Nesta abordagem estuda-se o processo pelo qual a estrutura e o arranjo de átomos-chave dos compostos *leads* se relacionam de forma quantitativa a um processo biológico como a sua ligação a um receptor ou sua reatividade química. Através da integração entre os processos de descoberta e detalhamento de alvos, estruturação de bibliotecas de compostos químicos, uso de descritores Sting de nano-ambiente que definem o *pocket*-alvo, estudos de ADMET e QSAR, acredita-se que os caminhos traçados podem convergir à descoberta e planejamento de novos fármacos.

Dentre os resultados obtidos pelo GPBC da Embrapa Informática Agropecuária, decorrentes de sua produção técnico-científica, merecem destaque quatro processos patenteados no Instituto Nacional da Propriedade Industrial (INPI):

**1) “Identificação de alvos terapêuticos para desenho computacional de drogas contra bactérias dotadas da proteína PilT”: Controle da *Xylella fastidiosa* em citros, uva e café (BR n. INPI: 020100089068).**

A *Xylella fastidiosa* é uma bactéria gram-negativa e não-flagelada que provoca várias doenças em plantas, tais como a Clorose Variegada dos Citrus (CVC), popularmente conhecida como “amarelinho” e a doença de Pierce, que afetam a citricultura e viticultura, respectivamente. A previsão de produção de laranja no Brasil, safra 2014/2015, está estimada em 289,9 milhões de caixas (de 40,9 quilos) setembro, 2014 (CITRUS BR, 2014)<sup>6</sup>. Desde a sua constatação em 1987, a CVC, ou amarelinho, passou a ser a doença mais importante da cultura dos citros, causando um prejuízo anual de 100 milhões de dólares.

O microrganismo *Xylella fastidiosa* é limitado a persistir apenas colonizando os vasos do xilema, os vasos condutores de água e sais em plantas, e no trato digestório anterior de alguns insetos, como os popularmente conhecidos como cigarrinhas, que se alimentam da seiva das plantas e servem como vetores para carreamento e inserção da bactéria nas plantas (HOPKINS; PURCELL,

<sup>6</sup> Disponível em: <<http://www.citrusbr.com/>>.

2002). Há muitos anos tem-se testado o uso de inseticida contra os insetos-vetores desta bactéria, porém não houve suficiente eficácia.

Os mecanismos de virulência da *Xylella fastidiosa* e a forma como esta interage com plantas hospedeiras não são totalmente compreendidos. A explicação mais plausível é a formação de agregados como biofilmes que, ao colonizar vasos do xilema, causa um bloqueio ao fluxo de seiva e culmina nos sintomas da doença. Hopkins e Purcell (2002) sugeriram que a colonização e a patogenicidade da bactéria *Xylella fastidiosa* que levam à doença de Pierce estão estritamente relacionadas com a sua capacidade de se movimentar dentro dos elementos de vaso do xilema, o que permite a colonização de outras regiões da planta.

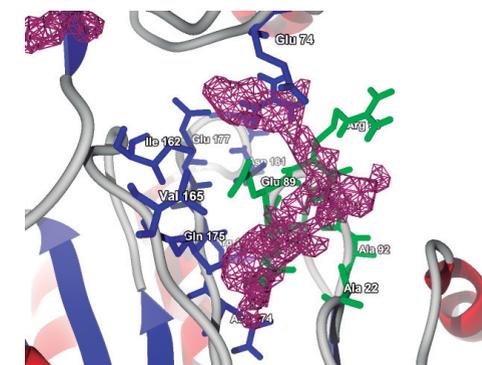
O projeto genoma da *Xylella fastidiosa* (SIMPSON et al., 2000) revelou a presença de genes que codificam proteínas envolvidas na biogênese e função do *type IV pili* (T4P). O *Twitching motility* é uma forma de movimento associado à superfície pelo qual as bactérias puxam-se rapidamente ao longo das superfícies através de ciclos de polimerização e despolimerização do Pilus (T4P). A energia necessária para o movimento é fornecida por meio de hidrólise do ATP por proteínas chamadas PilB e PilT para montagem e desmontagem, respectivamente, do Pilus. A perda de função da proteína PilT ou da PilB resulta na ausência deste tipo de motilidade, associada à privação da extensão ou retração do pilus.

Essa invenção se refere a um método para identificar regiões-alvo existentes na interface de monômeros constituintes da proteína PilT, com o objetivo de desenhar moléculas potencialmente aplicáveis no comprometimento da atividade desta proteína, controlando, assim, processos infecciosos. O método é caracterizado pela:

- 1) Seleção de, pelo menos, uma sequência de aminoácidos constituidora de monômero de PilT.
- 2) Desenvolvimento de um modelo computacional tridimensional da estrutura homo-hexamérica da PilT.
- 3) Análise computacional para determinar os resíduos de aminoácidos formadores de interface (IFR) e suas características físicas, químicas, físico-químicas e estruturais para todas as cadeias dos modelos de complexos hexaméricos gerados.
- 4) Seleção das regiões a serem usadas como alvos terapêuticos e alvos terapêuticos preferenciais na interface entre os monômeros baseada na intensidade dos seguintes parâmetros:
  - Energia de contatos interfaciais.
  - Exposição, em área em complexo.
  - Propensidade do aminoácido-alvo em estabelecer pontes de hidrogênio (como doador ou acceptor).
  - Propensidade do aminoácido-alvo em estabelecer contatos de cunho eletrostático.
  - Preferencialmente, presença em *pockets* (em isolamento ou em complexo).
  - Preferencialmente, que estejam presentes somente em organismos-alvo em relação a organismos não patogênicos.
- 5) Modelagem computacional para o desenho de moléculas potencialmente capazes de efetuar ligações e/ou interações entre regiões-alvo dos monômeros (EMBRAPA INFORMÁTICA AGROPECUÁRIA, 2010).

A pesquisa resultou na patente da metodologia para identificação de alvos terapêuticos específicos para o desenho de novos fármacos em estruturas modeladas da proteína PilT da *Xylella*

*fastidiosa* e que tem possível aplicação para organismos patogênicos que também possuem esta proteína com alto grau de similaridade de sequência e correspondência dos resíduos indicados como alvos em alinhamento de estrutura primária. Resultou também na proposição de uma lista de 54 resíduos-alvos a serem usados em processos de desenho de drogas baseado em estrutura. A Figura 2 representa a estrutural de um dos alvos terapêuticos, o resíduo Glu 89 na cadeia A do complexo XfAa1 (sobreposição estrutural do modelo da PilT de *X. fastidiosa*), mostrando os resíduos adjacentes que formam o nanoambiente no qual o fármaco a ser desenhado se ligará. Este alvo foi escolhido para representar visualmente, pois é de extrema importância por existir unicamente na XfPilT, ausente em microrganismos não patogênicos, contendo pocket adjacente, sendo essa característica muito importante ao desenho de fármaco. A estrutura coberta por uma teia nesta figura (roxo) identifica o volume e posição do ligante que interage com o alvo selecionado.



**Figura 2.** Estrutura de um alvo terapêutico, o resíduo 89 na cadeia A do complexo XfAa1 (sobreposição estrutural do modelo da PilT de *X. fastidiosa*).

Fonte: Embrapa (2010).

## 2) “Método para sugestão de mutantes que aumentam o índice de hidrofobicidade da superfície de proteínas mantendo parâmetros físico-químicos minimamente alterados no sítio catalítico”: Otimização de lipases para produção de biodiesel: (BR n. INPI: 012110000604).

Toda gordura de origem vegetal ou animal é composta, principalmente, de triglicerídeos (uma molécula de glicerol - um tri-álcool - esterificada com três moléculas de ácido graxo) e ácidos graxos livres (AGL). No processo de transesterificação para obtenção de biodiesel, os triglicerídeos presentes no óleo são transformados em moléculas menores de ésteres de ácido graxo (biodiesel) a partir de um agente transesterificante (álcool primário) e um catalisador (base ou ácido). O biodiesel pode ser obtido também usando ácidos graxos livres pelos processos de esterificação em meio preferencialmente ácido e pelo processo de craqueamento.

O álcool mais utilizado na obtenção do biodiesel é o metanol, que promove melhores rendimentos. Considerando que o Brasil é um dos maiores produtores de álcool etílico (etanol) no mundo, há um estímulo para a substituição do metanol pelo etanol, gerando um combustível agrícola totalmente independente do petróleo. A dificuldade na utilização do etanol consiste no fato de que a água é um dos agentes causadores de reações paralelas de saponificação, consumindo o catalisador e reduzindo a eficiência da reação de transesterificação. A utilização de álcool anidro é uma forma de diminuir a formação de sabões, porém, eleva em demasia os custos de produção inviabilizando esse procedimento.

O uso de catalisadores químicos (bases ou ácidos fortes) para a síntese do biodiesel possui algumas desvantagens, como a exigência de baixo teor de ácidos graxos livres e água na matéria-prima do óleo renovável. Além disso, indesejavelmente, ocorre a formação de uma emulsão devido à má solubilidade do álcool nos óleos, o que complica as etapas posteriores do tratamento. Por último, porém não menos importante, a quantidade de álcool a ser usado deve ser muito

maior que a razão molar da reação, e a evaporação/refluxo do álcool em excesso conduzem a um aumento do consumo de energia. Estas desvantagens, somadas à demanda mundial por processos limpos, renováveis e seletivos colocam em destaque a possibilidade do uso de catalisadores biológicos (enzimas) como alternativa para a síntese do biodiesel.

As lipases são enzimas que catalisam a hidrólise de triglicerídeos. Lipases de determinados microrganismos podem catalisar tanto reações de esterificações, bem como reações de transesterificação tendo, como substrato, triglicerídeos e, como reagentes, ácidos graxos de cadeia longa ou álcool primário. A utilização das lipases na produção de biodiesel é relativamente recente, porém tem se mostrado muito promissora devido a uma série de vantagens sobre os catalisadores químicos. As enzimas lipolíticas comerciais em geral foram selecionadas para fins relacionados à indústria de alimentos. Em tais processos, o meio reacional é emulsificado através do uso de detergentes, situação inviável para a produção de biodiesel, pois adiciona passos ao processo que aumentam o custo e o tempo de produção. Há que se considerar ainda que estas enzimas possuem baixa atividade catalítica em meio predominantemente apolar (hidrofóbico), contendo triglicerídeos e álcool como solventes. Como resultado, estabelece-se a necessidade do uso de grandes quantidades de catalisador e longos tempos de reação, o que torna o processo ainda mais oneroso.

Para criar enzimas mais adequadas para tal processo, é hipotetizado que uma enzima lipase com superfície mais hidrofóbica interage melhor com o substrato em um ambiente livre de solventes polares, levando a um rendimento maior na conversão de óleo em biodiesel, quando comparado com lipases naturais.

O método consistiu das seguintes etapas:

- Seleção da proteína de interesse em banco de dados públicos como o Protein Data Bank.
- Seleção de características físico-químicas e estruturais presentes no banco de dados Blue Star Sting.
- Definição de valores limites para cada uma das características selecionadas na etapa dois.
- Uso do módulo JPD do Blue Star Sting para seleção dos aminoácidos cujas características satisfazem os valores de cada um dos descritores selecionados na etapa dois.
- modelagem por homologia com software Modeller de mutantes com mutações singulares, modificando cada um dos aminoácidos selecionados na etapa quatro, por resíduo de Valina (Alanina).
- Monitoramento da variação das propriedades físico-químicas e/ou estruturais dos aminoácidos que compõem o sítio catalítico, através de geração de arquivos em formato TGZ pelo servidor do Blue Star Sting.
- Seleção dos melhores mutantes com base no escore V4.
- Construção de modelos com mutações múltiplas baseadas nos melhores valores de V4 para as mutações singulares estudadas, utilizando o Modeller.
- Avaliação dos modelos com mutações múltiplas pelo escore V4, comparando com a estrutura nativa.
- Medida da variação da área de superfície hidrofóbica em relação à estrutura nativa da enzima selecionada visando a maximização do parâmetro SHI (EMBRAPA INFORMÁTICA AGROPECUÁRIA, 2011).

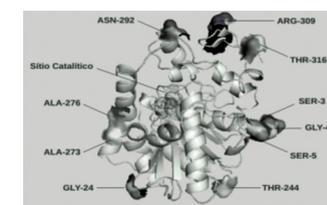
As pesquisas desenvolvidas pelo GPBC possibilitaram que a Embrapa pleiteasse junto ao INPI o registro da patente do método de engenharia de proteínas, no qual são identificadas regiões da superfície de enzimas como mostra a Figura 3, passíveis de mutações que interferem de maneira reduzida nas propriedades físicas, químicas e estruturais dos aminoácidos do sítio catalítico, com o objetivo de criar mutantes que apresentem a superfície macromolecular mais hidrofóbica e, portanto, serem empregadas para a obtenção de biodiesel por catálise biológica.

### 3) “Inibidores das enzimas poligalacturonases de fungos fitopatogênicos”: Fungicidas contra *Fusarium* e outros fungos patogênicos (BR n. INPI: 02012 000 3126).

Grande parte das doenças de plantas que causam prejuízos para a agricultura brasileira e mundial é causada por fungos fitopatogênicos, e a maioria deles de solo. Seria importante a redução do potencial patogênico destes fungos em áreas infestadas. Uma medida que tem sido utilizada é a de incorporação de matéria orgânica no solo, já que a introdução de antagonistas é uma medida de controle biológico. Entretanto, o índice de controle obtido com este método, isoladamente, está abaixo do necessário para impedir danos à cultura. O uso e desenvolvimento de cultivares resistentes seria uma melhor opção de controle destas doenças; todavia, muitos hospedeiros não apresentam resistência a esses patógenos. A funcionalidade nem sempre é possível, devido à inexistência no mercado de cultivares com todas as características desejadas. Existem métodos de controle químico contra os fungos fitopatogênicos, como, por exemplo, o realizado, até há pouco tempo, com um agrotóxico de amplo espectro, o brometo de metila utilizado nos últimos 60 anos como fumigante de solo em pré-plantio. Embora altamente eficaz, rápido, de fácil penetração no solo, amplo espectro e baixa resistência dos fungos, foi comprovado que este confere riscos para o ambiente, para o homem e para a camada de ozônio.

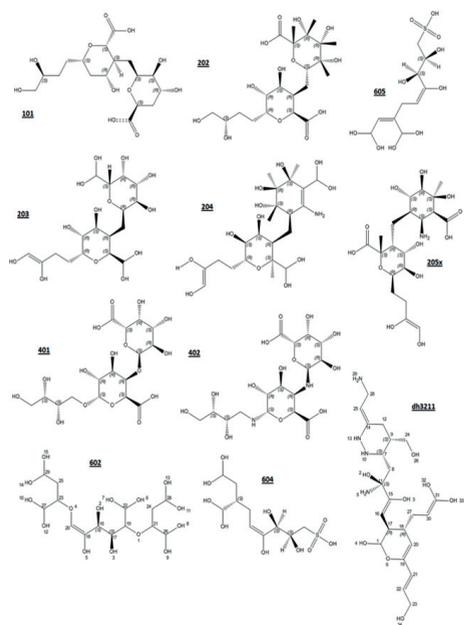
A invenção que descrevemos em seguida para ilustrar o trabalho do GPBC da Embrapa nessa área refere-se ao desenho computacional de novos compostos com potencial inibitório para enzimas endopoligalacturonases (PG) de fungos fitopatogênicos, com o intuito de evitar ou diminuir a colonização desses microrganismos nos tecidos vegetais. A PG integra um grupo de enzimas secretadas por microrganismos fitopatogênicos durante o processo de invasão dos tecidos vegetais, participando na catálise da hidrólise da pectina, culminando na desestruturação do arcabouço da parede celular, o que favorece a invasão de hifas dos fungos. Uma gama de microrganismos fitopatogênicos utiliza essas enzimas como fatores de patogenicidade que levam a doenças em uma grande variedade de plantas de interesse econômico como o trigo, a cevada, o tomate, o morango, a manga, o arroz, a cana-de-açúcar, dentre outros.

Com o objetivo de minimizar as perdas causadas por estes patógenos, foram desenhadas pequenas moléculas planejadas (Figura 4) para se ligarem com alta afinidade aos resíduos do sítio de ligação ao substrato desses patógenos. O ligante dh3211 posicionado no sítio-alvo da 1HG8 (PDB) (Figura 5), é o ligante que apresentou melhor resultado de interação entre os resíduos específicos de PG fúngicas: D194, H188 e G305, tanto na 1HG8 quanto na 2IQ7. Possui grupos químicos que conferem polaridades diferentes para cada extremidade da estrutura desta molécula o que o faz encaixar na região do sítio ligante e do catalítico através de complementaridade eletrostática.



**Figura 3.** Indicação de posição tridimensional dos aminoácidos substituídos por Valina na superfície de proteína nativa (1TCB.pdb).

Fonte: Embrapa Informática Agropecuária (2011).



**Figura 4.** Representação estrutural das moléculas desenhadas computacionalmente que obtiveram melhores resultados nos *dockings* com as PG 1HG8 e 2IQ7 (PDB) e que, provavelmente, possuirão caráter inibitório para as enzimas PG.

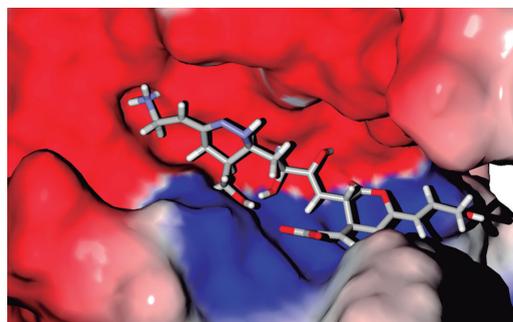
Fonte: Embrapa Informática Agropecuária (2012b).

fruto, formação das células do xilema e crescimento do tubo polínico (EMBRAPA, 2012).

#### 4) “Inibidores das enzimas alfa amilases de insetos”: Controle de insetos que prejudicam armazenamento de sementes (BR n. INPI: 02012 000 3145).

Os grãos estão entre os principais constituintes da base alimentar dos povos. Já sua produção é sazonal e em muitos lugares as colheitas ocorrem apenas uma vez ao ano. Para que seja possível alimentar todas as pessoas do mundo é necessário que a maior parte da produção dos principais grãos (arroz, trigo, sorgo, milho, soja e painço) seja estocada por até mais que um ano.

Para controle e minimização das perdas de grãos estocados, já existem várias estratégias, tais como desinfestação térmica, baixa umidade, baixo teor de oxigênio, resfriamento e armazenamento hermético, por exemplo, que são delineadas segundo o clima da região, tipo de grão e tipo de inseto. Ainda assim, há grandes perdas de produtos estocados causadas por pragas, que perfazem cerca de 10% da produção total, e pelo menos 50% dessa perda devem-se a insetos. Tomando o trigo como exemplo, cuja produção é estimada para a safra 2013-2014 em 711,42 milhões de toneladas, das quais 182,78 milhões de toneladas deverão ser estocadas, e considerando que o preço da tonelada seja de US\$ 496,21 em 16 de setembro de 2014 (CUSTÓDIO, 2014), com perdas de 10% desse produto em estoque alcança-se um prejuízo em torno de US\$ 9,07 bilhões (US\$ 4,13 bilhões na safra 2009-2010 por ocasião da solicitação da patente), sendo metade disso devido a insetos. Se considerarmos a crescente demanda anual de estoque de grãos



**Figura 5.** Ligante dh3211 posicionado no sítio-alvo da 1HG8 (PDB).

Fonte: Embrapa Informática Agropecuária (2012b).

A região catalítica é predominantemente negativa e na imagem aparece na cor vermelha; a região ligante, majoritariamente positiva, aparece na imagem como azul; as regiões em rosa claro são apolares.

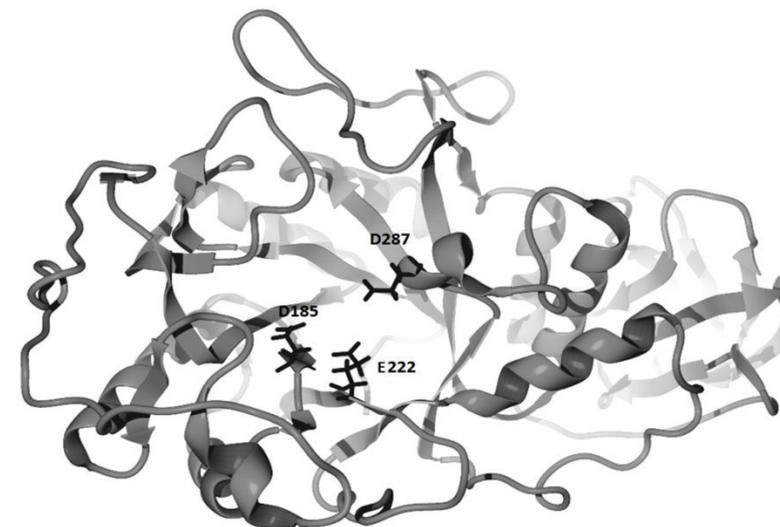
Além disso, outras funções das PG vegetais poderão ser manipuladas usando os compostos desenvolvidos neste trabalho, caso seja confirmada sua ação inibitória sobre as PG de plantas, tais como: processos de separação de células, germinação, abscisão de órgãos, deiscência das anteras, maturação do grão de pólen, amadurecimento do

em resposta à crescente produção agrícola e somarmos os prejuízos de todos os outros grãos, pode-se chegar a um valor exorbitante.

Logo, o interesse em reduzir perdas de grãos durante o período de estocagem é, sem dúvida nenhuma, tanto de países produtores como daqueles que importam e estocam.

A presente invenção que ilustra o trabalho do GPBC da Embrapa nesta área específica refere-se ao desenho computacional de novos compostos com potencial inibitório para enzimas alfa-amilases de insetos, com o intuito de diminuir os danos causados por eles, principalmente em produtos estocados.

Com o objetivo de diminuir as perdas agrícolas estocadas causadas pela ação de insetos, foram desenhadas estruturas que se ligam à alfa-amilase com alta afinidade teórica. Na Figura 6 podemos identificar a tríade catalítica da alfa-amilase do *Tenebrio molitor* (TMA, 1JAE.pdb) com os resíduos de Aspartato (D185 e D287) e Glutamato (E222) destacados em preto onde deve se ligar as estruturas inibidoras da enzima. Dessa forma, os insetos cuja alfa-amilase foi inibida não conseguem obter os resíduos de açúcar necessários para obtenção de energia e, em consequência, morrem de privação energética (EMBRAPA, 2012).



**Figura 6.** Visão da tríade catalítica da alfa-amilase do *Tenebrio molitor* (TMA, 1JAE.pdb) com os resíduos de Aspartato (D185 e D287) e Glutamato (E222) destacados em preto. Imagem gerada pelo Molegro Virtual Docker.

Fonte: Embrapa Informática Agropecuária (2012a).

### 3.1 Infraestrutura computacional do GPBC

O GPBC dispõe de uma infraestrutura computacional constituída de vários servidores convencionais e um servidor HPC, SGI UV200 com 128 cores, 1Tb de RAM e 72 Tb de HD que agiliza o uso das técnicas de High throughput in silico screening, drug design, ADMET, QSAR, Ludi e também possibilita buscas mais rápidas (em memória) em banco de dados Sting\_RDB expandindo a capacidade de identificação dos novos alvos para planejamento dos agroquímicos e drogas/fármacos.

Os softwares disponíveis para as pesquisas em Biologia Computacional pelo Grupo de Pesquisa são:

- Sting (Sting na versão: Blue Star Sting) - plataforma consolidada para análise de relação entre as estruturas macromoleculares e suas funcionalidades.
- Molegro - plataforma para planejamento e desenho de novos fármacos/drogas/agroquímicos que também tem forte vertente na linha de mineração de dados estruturais e é usado no laboratório de GPBC para *virtual screening* e *docking* das moléculas com alta precisão.
- Accelrys Discovery Studio - plataforma completa para análise estrutural das proteínas com módulos para varredura de alto desempenho embasada no processo de atracamento dos compostos químicos nas proteínas-alvos como também o planejamento de fármacos, análise in silico da sua potencial toxicidade, absorção, distribuição, metabolismo e excreção (ADMET)<sup>7</sup>.
- Softwares: Yassara, NAMD, VMD, Gromacs etc.

## 4 Considerações finais

O planejamento e desenvolvimento de novos fármacos e agroquímicos, que servem para proteção de plantas, animais e humanos, é um processo bastante complexo e longo, demandando investimento contínuo para que se possa chegar ao resultado/produto com potencial para ser oferecido ao mercado.

Convicta da importância de o Brasil diminuir a dependência de importação de matérias ativas para a formulação de agro defensivo e passar a produzir no País esses princípios ativos, o Grupo de Pesquisa em Biologia Computacional da Embrapa já deu os primeiros passos neste sentido e patenteou algumas tecnologias com as quais se pode desenhar os compostos químicos identificados com elevado potencial de atividade in silico contra alvos proteicos encontrados em patógenos.

Agora, é preciso avançar, na direção da síntese desses compostos desenhados, acompanhado de ensaios in vitro e in vivo, devendo, necessariamente, ocorrer uma interação permanente e retroalimentar entre o desenvolvimento in vitro e in vivo com a análise teórica (in silico) no sentido de aperfeiçoar o desenho do agro defensivo.

O produto desse esforço será um inventário de novos agroquímicos e fármacos que pode ser um dos mais importantes desafios perante o agronegócio do país, representando um dos fatores fundamentais de conquista de vantagem competitiva no mercado internacional em futuro próximo.

No final deste caminho, será necessário investir esforços no preparo do processo de produção dos novos compostos químicos em escala piloto e, posteriormente, em larga escala industrial.

## 5 Referências

- ACCELRYS SOFTWARE INC. **Discovery Studio Visualizer version 4.0**. San Diego, CA, 2014.
- ANDERSON, A. C. The process of structure-based drug design. **Chemistry & Biology**, Cambridge, Mass., v. 10, n. 9, p. 787-797, Sept. 2003. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1074552103001947>>. Acesso em: 07 out. 2014.
- BORRO, L. C.; OLIVEIRA, S. R. M.; YAMAGISHI, M. E.; MANCINI, A. L.; JARDINE, J. G.; MAZONI, I.; SANTOS, E. H. dos; HIGA, R. H.; KUSER, P. R.; NESHICH, G. Predicting enzyme class from protein structure using Bayesian classification. **Genetics and Molecular Research**, Ribeirão Preto, v. 5, n. 1, p. 193-202, Mar. 2006. Disponível em: <[http://www.funpecrp.com.br/gmr/year2006/vol1-5/xm0009\\_full\\_text.htm](http://www.funpecrp.com.br/gmr/year2006/vol1-5/xm0009_full_text.htm)>. Acesso em: 07 out. 2014.
- BRAGHINI, C. A.; NESHICH, I. A. P.; NESHICH, G.; SOARDI, F. C.; MELLO, M. P. de; COSTA, V. P.; VASCONCELLOS, J. P. C. de; MELO, M. B. de. New mutation in the *myocilin* gene segregates with juvenile-onset open-angle glaucoma in a Brazilian family. **Gene**, Amsterdam, v. 523, n. 1, p. 50-57, July, 2013. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S037811913003788>>. Acesso em: 07 out. 2014.
- CITRUS BR. **Estimativa total de safra 2014/2015 e projeção de estoques - maio/2014**. Disponível em: <<http://www.citrusbr.com>>. Acesso em: 07 out. 2014.
- CUSTÓDIO, F. Trigo: USDA reporta aumento na produção mundial e preços recuam na CBOT. **Notícias Agrícolas**, Campinas, 11 jun. 2014. Disponível em: <<http://www.noticiasagricolas.com.br/noticias/usda/140724-trigo-usda-reporta-aumento-na-producao-mundial-e-precos-recuam-na-cbot.html#.VERSaflDUaB>>. Acesso em: 07 out. 2014.
- DIAS-LOPES, C.; NESHICH, I. A. P.; NESHICH, G.; ORTEGA, J. M.; GRANIER, C.; CHÁVEZ-OLORTEGUI, C.; MOLINA, F.; FELICORI, L. Identification of new Sphingomyelinases D in pathogenic fungi and other pathogenic organisms. **Plos One**, San Francisco, v. 8, n. 11, Nov. 2013. Disponível em: <<http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0079240>>. Acesso em: 07 out. 2014.
- EDGAR, R. C. Muscle: multiple sequence alignment with high accuracy and high throughput. **Nucleic Acids Research**, Oxford, v. 32, n. 5, p. 1792-1797, Mar. 2004. Disponível em: <<http://nar.oxfordjournals.org/content/32/5/1792.full.pdf+html>>. Acesso em: 07 out. 2014.
- EMBRAPA INFORMÁTICA AGROPECUÁRIA. G. Neshich; I. A. Neshich; I. Mazoni; J. G. Jardine; J. A. Salim; F. R. de Moraes. **Método para previsão de mutantes que aumentem o índice de hidrofobicidade da superfície de proteínas**. BR n. PI1211604, 04 ago. 2011.
- EMBRAPA INFORMÁTICA AGROPECUÁRIA G. Neshich; I. A. P. Neshich; L. Nishimura; J. A. Salim; I. Mazoni; J. G. Jardine. **Identificação de alvos terapêuticos para desenho computacional de drogas contra bactérias dotadas da proteína PilT**. 2010, Brasil. BR n. PI 1089068, 28 out. 2010.
- EMBRAPA INFORMÁTICA AGROPECUÁRIA. G. Neshich; J. G. Jardine; I. Mazoni; I. A. P. Neshich; L. Nishimura. **Desenho computacional para novos inibidores de alfa-amilases**. BR n. PI 0003126, 09 maio 2012a.
- EMBRAPA INFORMÁTICA AGROPECUÁRIA. G. Neshich; J. G. Jardine; I. Mazoni; I. A. P. Neshich; L. Nishimura. **Inibidores das enzimas poligalacturanases de fungos fitopatogênicos**. BR n. PI 0003145, 02 fev. 2012b.
- FERNANDEZ, J. H.; HAYASHI, M. A.; CAMARGO, A. C.; NESHICH, G. Structural basis of the lisinopril-binding specificity in N- and C-domains of human somatic ACE. **Biochemical and Biophysical Research Communications**, Orlando, v. 308, n. 2, p. 219-226, Aug. 2003. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/12901857>>. Acesso em: 07 out. 2014.
- FREITAS, S. M. de; MELO, L. V. de; SILVA, M. C. M. da; VRIEND, G.; NESHICH, G.; VENTURA, M. M. Analysis of the black-eyed pea trypsin and chymotrypsin inhibitor alpha-chymotrypsin complex. **FEBS Letters**, Amsterdam, v. 409, n. 2, p. 121-127, June, 1997.

<sup>7</sup> Disponível em: <<http://accelrys.com/products/discovery-studio/admet.html>>.

GUEDES, I. A.; MAGALHÃES, C. S. de; DARDENNE, L. E. Receptor-ligand molecular docking. **Biophysical Reviews**, Heidelberg, 2014. v. 6, n. 1, p. 75-87, Mar. 2014.

HOPKINS, D. L.; PURCELL, A. H. *Xylella fastidiosa*: cause of Pierce's disease of grapevine and other emergent diseases. **Plant Disease**, Saint Paul, v. 86, n. 10, p. 1056-1066, Oct. 2002. Disponível em: <<http://apsjournals.apsnet.org/doi/pdf/10.1094/PDIS.2002.86.10.1056>>. Acesso em: 07 out. 2014.

KUNTZ, I. D.; BLANEY, J. M.; OATLEY, S. J.; LANGRIDGE, R.; FERRIN, T. A geometric approach to macromolecular-ligand interactions. **Journal of Molecular Biology**, London, v. 161, n. 2, p. 269-288, Oct. 1982.

LARKIN, M. A.; BLACKSHIELDS, G.; BROWN, N. P.; CHENNA, R.; McGETTIGAN, P. A.; McWILLIAN, H.; VALENTIN, F.; WALLACE, I. M.; WILM, A.; LOPEZ, R.; THOMPSON, J. D. ; GIBSON, T. J.; HIGGINS, D. G. Clustal W and Clustal X version 2.0. **Bioinformatics**, Oxford, v. 23, n. 21, p. 2947-2948, Nov. 2007.

MAGALHÃES, C. S. da. **Algoritmos genéticos para o problema de docking proteína-ligante**. 2006. 261 f. Tese (Doutorado em Modelagem Computacional) – Laboratório Nacional de Computação Científica - LNCC, Petrópolis. Disponível em: <[http://www.gmmsb.lncc.br/pdf/T\\_Magalhaes2006.pdf](http://www.gmmsb.lncc.br/pdf/T_Magalhaes2006.pdf)>. Acesso em: 07 out. 2014.

MAGALHÃES, C. S. de; BARBOSA, H. J. C.; DARDENNE, L. E. Métodos de docking receptor-ligante para o desenho racional de compostos bioativos. In: MORGON, N. H.; COUTINHO, K. (Org.). **Métodos de química teórica e modelagem molecular**. São Paulo: Editora Livraria da Física, 2007. p. 489-531.

MANCINI, A. L.; HIGA, R. H.; OLIVEIRA, A.; DOMINQUINI, F.; KUSER, P. R.; YAMAGISHI, M. E.; TOGAWA, R. C.; NESHICH, G. STING Contacts: a web-based application for identification and analysis of amino acid contacts within protein structure and across protein interfaces. **Bioinformatics**, Oxford, v. 20, n. 13, p. 2145-2147, Sep. 2004.

MARCELLINO, L. H.; NESHICH, G.; SÁ, M. F. G. de; KREBBERS, E.; GANDER, E. S. Modified 2S albumins with improved tryptophan content are correctly expressed in transgenic tobacco plants. **FEBS Letters**, Amsterdam, v. 385, n. 3, p. 154-158, May 1996.

NOTREDAME, C.; HIGGINS, D. G.; HERINGA, J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. **Journal of Molecular Biology**, London, v. 302, n. 1, p. 205-217, Sept. 2000.

NESHICH, G.; BORRO, L. C.; HIGA, R. H.; KUSER, P. R.; YAMAGISHI, M. E.; FRANCO, E. H.; KRAUCHENCO, J. N.; FILETO, R.; RIBEIRO, A. A.; BEZERRA, G. B.; VELLUDO, T. M.; JIMENEZ, T. S.; FURUKAWA, N.; TESHIMA, H.; KITAJIMA, K.; BAVA, A.; SARAI, A.; TOGAWA, R. C.; MANCINI, A. L. The Diamond STING server. **Nucleic Acids Research**, Oxford, v. 33, Web server issue, p. W29-W35, July 2005.

NESHICH, G.; MAZONI, I.; OLIVEIRA, S. R.; YAMAGISHI, M. E.; KUSER-FALCÃO, P. R.; BORRO, L. C.; MORITA, D. U.; SOUZA, K. R.; ALMEIDA, G. V.; RODRIGUES, D. N.; JARDINE, J. G.; TOGAWA, R. C.; HIGA, R. H.; CRUZ, S. A.; VIEIRA, F. D.; SANTOS, E. H.; MELO, R. C.; SANTORO, M. M. The Star STING Server: A multiplatform environment for protein structure analysis. **Genetics and Molecular Research**, Ribeirão Preto, v. 5, n. 4, p. 717-722, 2006.

NESHICH, G.; ROCCHIA, W.; MANCINI, A. L.; YAMAGISHI, M. E. B.; KUSER, P. R.; FILETO, R.; BAUDET, C.; PINTO, I. P.; MONTAGNER, A. J.; PALADRANI, J. F.; KRAUCHENCO, J. N.; TORRES, R. C.; SOUZA, S.; TOGAWA, R. C.; HIGA, R. H. Java Protein Dossier: a novel web-based data visualization tool for comprehensive analysis of protein structure. **Nucleic Acids Research**, Oxford, v. 32, Web server issue, p. W595-W601, 2004.

NESHICH, G.; TOGAWA, R. C.; MANCINI, A. L.; KUSER, P. R.; YAMAGISHI, M. E.; PAPPAS JUNIOR, G.; TORRES, W. V.; CAMPOS, T. F. e; FERREIRA, L. L.; LUNA, F. M.; OLIVEIRA, A. G.; MIURA, R. T.; INOUE, M. K.; HORITA, L. G.; SOUZA, D. F. de; DOMINQUINI, F.; ALVARO, A.; LIMA, C. S.; OGAWA, F. O.; GOMES, G. B.; PALANDRANI, J. F.; SANTOS, G. F. dos; FREITAS, E. M. de; MATTIUIZ, A. R.; COSTA, I. C.; ALMEIDA, C. L. de; SOUZA, S.; BAUDET, C.; HIGA, R. H. STING Millennium: a Web based suite of programs for comprehensive and simultaneous analysis of protein structure and sequence. **Nucleic Acids Research**, Oxford, v. 31, n. 13, p. 3386-92, July 2003.

PORTER, C. T.; BARTLETT, G. J.; THORNTON, J. M. The Catalytic Site Atlas: a resource of catalytic sites and residues identified in enzymes using structural data. **Nucleic Acids Research**, Oxford, v. 32, database issue, D129-133, Jan. 2004.

PyMOL MOLECULAR GRAPHICS SYSTEM. 2014. Disponível em: <<http://sourceforge.net/projects/pymol/>>. Acesso em: 18 set. 2014.

SIMÕES, M.; BAHIA, D.; ZERLOTINI, A.; TORRES, K.; ARTIQUENAVE, F.; NESHICH, G.; KUSER, P.; OLIVEIRA, G. Single nucleotide polymorphisms identification in expressed genes of *Schistosoma mansoni*. **Molecular and Biochemical Parasitology**, Amsterdam, v. 154, n. 2, p. 134-140, Aug. 2007.

SIMPSON, A. J. G. et al. The genome sequence of the plant pathogen *Xylella fastidiosa*. **Nature**, London, v. 406, p. 151-157, July 2000.

SRIDHARAN, S.; NICHOLLS, A.; HONIG, B. A. New vertex algorithm to calculate solvent accessible surface areas. **Biophysical Journal**, New York, v. 61, A174, 1992.

YANG, S. Y. Pharmacophore modeling and applications in drug discovery: challenges and recent advances. **Drug Discovery Today**, Kidlington, v. 15, n. 11/12, p. 444-450, June 2010. Disponível em: <[http://csmres.co.uk/cs.public.upd/article-downloads/Most%20downloaded%20article%20Q2%202010\\_a15584.pdf](http://csmres.co.uk/cs.public.upd/article-downloads/Most%20downloaded%20article%20Q2%202010_a15584.pdf)>. Acesso em: 18 set. 2014.

