



República Federativa do Brasil
Ministério do Desenvolvimento, Indústria
e do Comércio Exterior
Instituto Nacional de Propriedade Industrial

(21) PI 1104122-6 A2



(22) Data de Depósito: 04/08/2011

(43) Data da Publicação: 25/08/2015
(RPI 2329)

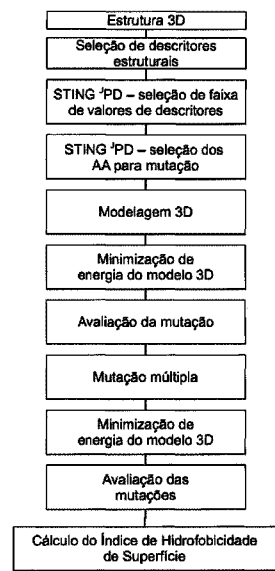
(54) **Título:** METODO PARA PREVISÃO DE MUTANTES QUE AUMENTEM O ÍNDICE DE HIDROFOBICIDADE DA SUPERFÍCIE DE PROTEÍNAS

(51) **Int.Cl.:** G06F19/12; C12N9/20

(73) **Titular(es):** Empresa Brasileira de Pesquisa Agropecuária - Embrapa

(72) **Inventor(es):** Goran Nesic, Ivan Manzoni, Izabella Agostinho Pena Neshich, José Augusto Salim, José Gilberto Jardine

(57) **Resumo:** MÉTODO PARA PREVISÃO DE MUTANTES QUE AUMENTEM O ÍNDICE DE HIDROFOBICIDADE DA SUPERFÍCIE DE PROTEÍNAS. A presente invenção refere-se a um método de engenharia de proteínas, no qual são identificadas regiões da superfície de enzimas que seriam passíveis de mutações que influenciem de maneira reduzida as propriedades físico-químicas e estruturais dos aminoácidos do sítio catalítico, com o objetivo de criar mutantes que apresentem a superfície macromolecular mais hidrofóbica. Esta tecnologia pode ser aplicada para o desenho racional de enzimas com maior solubilidade em meios hidrofóbicos como os óleos vegetais.



MÉTODO PARA PREVISÃO DE MUTANTES QUE AUMENTEM O ÍNDICE DE HIDROFOBICIDADE DA SUPERFÍCIE DE PROTEÍNAS

CAMPO DA INVENÇÃO

A presente invenção refere-se a um método de engenharia de proteínas, no qual são
5 identificadas regiões da superfície de enzimas que seriam passíveis de mutações que
influenciem de maneira reduzida as propriedades físico-químicas e estruturais dos
aminoácidos do sítio catalítico, com o objetivo de criar mutantes que apresentem a
superfície macromolecular mais hidrofóbica. Notadamente, esta tecnologia pode ser
aplicada para o desenho racional de enzimas com maior solubilidade em meios
10 hidrofóbicos (óleos vegetais, por exemplo).

FUNDAMENTOS DA INVENÇÃO

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
15 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).

Biodiesel é definido como um éster de ácido graxo obtido a partir da reação química
catalisada preferencialmente em meio básico de gorduras de origem vegetal ou animal com
20 um álcool primário (etílico ou metílico). Quando obtido dessa maneira o processo é
denominado de transesterificação. O biodiesel pode ser obtido também utilizando-se ácidos
graxos livres pelos processos de esterificação em meio preferencialmente ácido e pelo
processo de craqueamento. Óleos vegetais compostos por ácidos graxos de cadeias curtas
(ácido láurico, por exemplo) garantem melhor rendimento ao processo, pois a interação
25 com o agente transesterificante e o catalisador é mais eficaz (RICACZESKI, C.C.;
ZANCANARO, D.; ALZANI, A.; FERREIRA, E. F. Biodiesel, um combustível em
expansão. Synergismus científica UTFPR, Pato Branco, 01 (1,2,3,4) : 1-778. 2006).

A reação de transesterificação é de caráter reversível, sendo necessário um excesso

de álcool na reação (1:6 molar) para aumentar o rendimento de alquil ésteres e permitir a formação de uma fase separada de glicerol. 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 (RICACZESKI, C.C.; ZANCANARO, D.; ALZANI, A.; FERREIRA, E. F. Biodiesel, um combustível em expansão. Synergismus scyentifica UTFPR, Pato Branco, 01 (1,2,3,4) : 1-778. 2006).

O catalisador mais utilizado é o hidróxido de sódio (NaOH), amplamente conhecido como soda cáustica. Também pode ser utilizado o hidróxido de potássio (KOH). Cerca de 0,5% em relação ao peso do óleo é o indicado. Catalisadores básicos, como os citados acima, aceleram a reação em torno de 4000 vezes a mais que catalisadores ácidos, como o ácido clorídrico (HCl), além de serem mais viáveis economicamente. No entanto, a utilização de catalisadores básicos promove um maior nível de saponificação no processo: o catalisador reage com os ácidos graxos livres do óleo, formando sabão. A cada 1% em peso de soda cáustica usada como catalisador, cerca de 7% em peso de sabão será originado. Portanto, para que o processo de transesterificação seja satisfatório, os óleos vegetais devem conter no máximo 3% de ácido graxo livre.

Com isso, 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, para que a reação não seja prejudicada. 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 nem por isso menos importante - a quantidade de álcool de cadeia curta a ser usado deve ser muito maior que a razão molar da reação, e a evaporação/refluxo do álcool em excesso conduz a um aumento do consumo de energia (Documento de Patente PI 0418062-3A).

Estas desvantagens, somadas à demanda mundial por processos limpos, renováveis e seletivos colocam em destaque a possibilidade do uso de catalisadores biológicos como alternativa para a síntese do Biodiesel.

5 As lipases são enzimas que catalisam a hidrólise e síntese de acilgliceróis em ácidos graxos de cadeia longa, utilizando triacilglicerol com o substrato quando em meio aquoso. 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
10 através do uso de detergentes, situação inviável para a produção de biodiesel, pois, adiciona passos ao processo e que aumentam o custo e o tempo de produção.

Estas enzimas possuem baixa atividade catalítica em meio predominantemente apolar (hidrofóbico), contendo triglicerídios e álcool como solventes. Como resultado, estabelece-se a necessidade do uso de grandes quantidades de catalisador e longos tempos
15 de reação, o que torna o processo ainda mais oneroso. Esta baixa capacidade catalítica se deve ao fato da maioria das enzimas ser inibida por metanol. O metanol se associa à superfície polar da enzima formando uma espécie de selo que impede a interação desta com o triglicerídio (Ma, F., Hanna, M. A. (1999). Biodiesel production: a review. Bioresources Technology, 70, 1– 15.).

20 O uso de Lipases como bio-catalisadores permite a fácil recuperação de glicerol sem a necessidade de purificação ou produção de lixo químico. Apesar das vantagens mencionadas, as enzimas possuem um baixo rendimento no processo quando comparadas aos catalisadores inorgânicos.

Para criar enzimas mais adequadas para tal processo, é hipotetizado no presente
25 proposta que uma enzima Lipase com superfície mais hidrofóbica reagiria 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.

Da literatura científica pode-se perceber que existem vários tipos de lipases, sendo que algumas delas são mais passíveis de uso na síntese do biodiesel. Souza e colaboradores
30 (2008) testaram diferentes lipases comerciais imobilizadas (Lipozyme RM-IM, Lipozyme

TL-IM e Novozyme 435) com esta finalidade e encontraram que na conversão do óleo de soja desodorizado e destilado, a Novozym 435 teve o melhor desempenho, com 83.5% de conversão em reação a 50°C, usando 3 wt% da enzima e 2 g de etanol ($T_0 = 1$ g; $T_{30\text{ min}} = 1$ g) (Souza MS, Aguieiras EC, da Silva MA, Langone MA. Biodiesel synthesis via esterification of feedstock with high content of free fatty acids. Appl Biochem Biotechnol. 2009 May;154(1-3):74-88. Epub 2008 Dec 9).

Devido a esta maior eficácia na formação do biodiesel, a lipase Novozyme 435 foi selecionada para o presente estudo. Esta enzima, comercializada na forma imobilizada pela Novozymes (<http://www.novozymes.com>), é a lipase B do fungo *Candida antarctica* (NCBI gi: 576300), sendo uma proteína monomérica, pertencente à família das hidrolases com dobramento do tipo α/β . Sua estrutura foi resolvida por cristalografia e difração de raios X e depositada no PDB (código 1TCB) e apresenta uma tríade catalítica similar à encontrada nas serino-proteases, composta por Ser105, Asp187 e His224 (Uppenberg J, Patkar S, Bergfors T, Jones TA. Crystallization and preliminary X-ray studies of lipase B from *Candida antarctica*. J Mol Biol. 1994 Jan 14;235(2):790-2.).

A partir da estrutura da Lipase B da *Candida antarctica* depositada no PDB, foram usadas técnicas de Biologia Computacional Estrutural para identificar sítios passíveis de modificação com o intuito de criar uma enzima com maior solubilidade em ambientes hidrofóbicos (sendo, portanto, mais hidrofóbica em sua superfície) e que mantenha a função nativa da proteína base. Para tal, objetivou-se identificar regiões com aminoácidos hidrofílicos (ou aminoácidos pouco hidrofóbicos) localizados na superfície protéica da lipase, que não se encontrem próximos ao sítio catalítico e nem possuem contatos internos da forma que, na sua substituição, influenciarão minimamente a estrutura (como também a função) em si. Em seguida foi feita a modelagem in silico de lipases mutantes que possuam estes resíduos substituídos por resíduos de alta hidrofobicidade, como é o caso da Valina. Outro fator importante considerado foi de não usar para estas trocas, resíduos de importância na catálise, como os indicados pela “Lipase Engineering Database”: Gly39 e Trp104 (importantes na ligação com o álcool), Thr40, Gln106, Asp134, Thr138 e Gln157 (regiões de ligação e coordenação ao ácido graxo), Leu73 (um resíduo de ancoragem), Cisteínas 22, 64, 216 e 258 (envolvidas em pontes dissulfeto) e Leu144, Val149, Val154 e

Iso285 (resíduos envolvidos em sítios de ligação hidrofóbica na formação de um túnel) (Fischer M, Pleiss J. The Lipase Engineering Database: a navigation and analysis tool for protein families. *Nucleic Acids Res.* 2003 Jan 1;31(1):319-21.). O protocolo descrito acima permitiu a criação de lipases mutantes in silico com aumentada hidrofobicidade na superfície.

Existem vários trabalhos com reconhecimento científico comprovado que usaram técnicas similares às usadas neste trabalho, com o intuito de primeiro compreender melhor e posteriormente, mudar as estruturas de proteínas ou usá-las para a geração mais eficaz de biodiesel. A seguir expõe-se uma série de citações de alguns destes trabalhos.

Como mencionado, Souza et al.,2008 (Souza MS, Aguiéiras EC, da Silva MA, Langone MA. Biodiesel synthesis via esterification of feedstock with high content of free fatty acids. *Appl Biochem Biotechnol.* 2009 May;154(1-3):74-88. Epub 2008 Dec 9) estudaram a síntese de ésteres etílicos via esterificação de óleo de soja destilado e desodorizado através de lipases imobilizadas em um sistema livre de solventes, indicando a lipase Novozyme 435 como a que proporcionou a reação mais rápida e com maior eficácia.

Patkar et al., 1998 (Patkar S, Vind J, Kelstrup E, Christensen MW, Svendsen A, Borch K, Kirk O. Effect of mutations in *Candida antarctica* B lipase. *Chem Phys Lipids.* 1998 Jun;93(1-2):95-101.) testaram os efeitos de mutações na lipase B da *C. antarctica*, através de mutações de resíduos próximos ao sítio ativo e encontraram que a mutação T103G que introduziu a sequência consenso G-X-S-X-G (encontrada na maioria das lipases) levou a um aumento da termo estabilidade porém reduziu para a metade da atividade específica da enzima na formação dos ésteres. Outra mutação estudada, a W104H afetou drasticamente estas duas propriedades (termo estabilidade e atividade), reduzindo-as.

Existem várias patentes relacionadas à produção de Biodiesel através da catálise biológica usando lipases para tal. No documento US 2010/0047884, foi descrito um método para produção de biodiesel com base em misturas com substancial conteúdo de ácidos graxos livres, através da utilização de enzimas microbianas como biocatalisadoras que sejam efetivas para processos livres de solvente para a produção de ésteres de ácidos graxos e alquil-álcoois C₁-C₃.

O documento US 6.398.707 descreve uma técnica de aumentar a atividade de uma

enzima lipase imobilizada e uma técnica para regeneração de lipase imobilizada desativada, através do uso de um álcool com um número de carbonos não menor que três.

5 O documento PI 0418062-3A descreve um processo para a produção de biodiesel a partir de óleo renovável na presença de catálise por lipase em um sistema de reação em meio orgânico. Neste processo é usado um álcool de cadeia curta ROH como um receptor de acila, sendo usado como meio de reação um solvente orgânico relativamente hidrofílico sem efeito negativo na reatividade da lipase, e matéria-prima de óleo renovável sendo catalisada por uma lipase para a síntese do biodiesel por transesterificação, obtendo produção de biodiesel em 94% e em tempo de reação reduzido.

10 O documento PI 0419166-8A descreve um processo de produção de biodiesel a partir da matéria-prima de sabão, ou seja: qualquer matéria-prima de sabão gerada no processo de refino de álcali que contem de 10 a 60% de água, 0,1 a 2,0% de esteróis, 35 a 85% de derivados graxos incluindo glicerídeos parciais. Este processo ocorre através de: neutralizar e separar os sabões com ácidos fortes, até atingir pH 2-8, seguido por
15 esterificação enzimática usando Lipase com concentração variando de 100ppm a 10% em peso, usando alcanol de C1 a C6 em uma razão de peso de 5 a 100% em relação aos componentes graxos e usando temperatura de 15 a 70°C.

A descrição da invenção contida no presente documento consiste em uma nova abordagem de engenharia protéica para lipases com o intuito de sua otimização em uma
20 forma alternativa para a obtenção de proteínas mais lipossolúveis. A presente metodologia indica que a troca de aminoácidos localizados na superfície protéica da lipase B de C. antarctica levariam a um aumento substancial da hidrofobicidade de superfície (calculada a partir de um índice denominado SHI (“Surface Hydrophobicity Index”) e que postula-se que esta propriedade seja importante na melhoria da eficiência na produção do biodiesel por
25 biocatálise.

SUMÁRIO DA INVENÇÃO

A presente invenção refere-se a um método de engenharia de proteínas, no qual são identificadas regiões da superfície de enzimas que seriam passíveis de mutações que influenciem de maneira reduzida as propriedades físico-químicas e estruturais dos

aminoácidos do sítio catalítico, com o objetivo de criar mutantes que apresentem a superfície macromolecular mais hidrofóbica.

Uma primeira concretização da invenção diz respeito a um método para previsão de mutantes que aumente o índice de hidrofobicidade da superfície de proteínas caracterizado por compreender as seguintes etapas:

- 5 a) Selecionar a proteína de interesse para aumento de hidrofobicidade da sua superfície em banco de dados de estruturas macromoleculares público;
- b) Calcular usando o software e armazenar as características físico-químicas da proteína em banco de dados específico;
- 10 c) Selecionar as características físico-químicas e estruturais da proteína (superfície e sítio catalítico) de acordo com interesse;
- d) Definir valores limites para cada uma das características selecionadas na etapa "c";
- e) Selecionar os aminoácidos cujas características satisfizeram os valores de cada um dos descritores selecionados na etapa "c" através de um software específico;
- 15 f) Realizar uma modelagem por homologia com um software específico com mutações singulares modificando cada um dos aminoácidos selecionados na etapa "e" para o resíduo de interesse;
- g) Realizar, opcionalmente, minimização de energia por dinâmica molecular dos modelos 3D gerados na etapa "f";
- 20 h) Monitorar a variação das propriedades físico-químicas e/ou estruturais dos aminoácidos que compõe o sítio catalítico através da comparação dos descritores gerados pelo servidor do software específico;
- i) Construir modelos de proteínas com mutações múltiplas baseadas nos melhores valores de V_4 para as mutações singulares estudadas, utilizando o software de modelagem;
- 25 j) Realizar, opcionalmente, minimização de energia por dinâmica molecular dos modelos 3D gerados na etapa "i";
- k) Avaliar os modelos com mutações múltiplas pelo escore V_4 , comparando com a estrutura nativa;

Uma segunda concretização da invenção diz respeito a proteínas mutantes obtidas pelo referido método.

BREVE DESCRIÇÃO DAS FIGURAS

Figura 1 – Etapas do processo de geração de mutantes in silico de estruturas protéicas com área de superfície mais hidrofóbica. Etapas: A) Seleção de estrutura protéica cuja superfície deve ser modificada. B) Seleção dos descritores e C) suas faixas de valores. D) Seleção dos aminoácidos que apresentem as características escolhidas utilizando a ferramenta ¹PD do BluStar STING. E) Construção de modelos tridimensionais por modelagem por homologia seguida de F) refinamento por minimização de energia por dinâmica molecular. G) As modelos finais são classificados de acordo com o escore V4 para a menor interferência no sítio catalítico da enzima *wt*. H) As melhores mutações singulares de acordo com o escore V4 são escolhidas para compor as enzimas mutante com mutações múltiplas que, I) novamente são avaliadas em termos do escore V4 e pelo J) cálculo do Índice de Hidrofobicidade.

Figura 2 – Estrutura tridimensional em forma “cartoon” da proteína Lipase B de *Candida antarctica* (1TCB.pdb) mostrando sua estrutura secundária composta por três folhas betas e seis alfa-hélices (a). A superfície proteica colorida de acordo com potencial eletrostático evidencia a quantidade de aminoácidos polares (cinza e cinza escuro) presentes na superfície da enzima estudada – a área em branco apresenta a área ocupada por aminoácidos hidrofóbicos (b).

Figura 3 – Comparação entre os parâmetros calculados para o aminoácido Ser-105 (um dos três amino ácidos membros de tríade catalítica de lipase) em proteína contendo mutações e em estrutura nativa da enzima Lipase B de *Candida antarctica*. A figura está demonstrando para Ser-105 a variação dos valores dos parâmetros listados na tabela 3, encontrados em modelos a onde amino ácidos citados na tabela 1 foram substituídos por Val. Os descritores “acessibilidade ao solvente” e “Cross Presence Order” se apresentam como principais fatores introdutórios de variabilidade entre os modelos estudados. Os parâmetros pertinentes a estrutura nativa e aos diferentes modelos se alteram em cor, iniciando com proteína nativa (em preto) e mutante “Ser_3” (em cinza escuro).

Figura 4 – Comparação entre os parâmetros calculados para o aminoácido Asn-187 (um dos três amino ácidos membros de tríade catalítica de lipase) em mutantes, relativo aos

valores obtidos para a estrutura nativa da enzima Lipase B de *Candida antarctica*. A figura está demonstrando para Asn-187 a variação dos valores dos parâmetros listados na tabela 3, encontrados em modelos a onde amino ácidos citados na tabela 1 foram substituídos por Val. O descritor “acessibilidade ao solvente” se apresenta como principal fator introdutor de variabilidade entre os modelos estudados. Os parâmetros pertinentes a estrutura nativa e aos diferentes modelos se alteram em cor, iniciando com proteína nativa (em preto) e mutante “Ser_3” (em cinza escuro).

Figura 5 - Comparação entre os parâmetros calculados do aminoácido His-224 (um dos três amino ácidos membros de tríade catalítica de lípase) em mutantes, relativo aos valores obtidos para a estrutura nativa da enzima Lipase B de *Candida antarctica*. A figura está demonstrando para His-224 a variação dos valores dos parâmetros listados na tabela 3, encontrados em modelos a onde amino ácidos citados na tabela 1 foram substituídos por Val. O descritor “acessibilidade ao solvente” se apresenta como principal fator introdutor de variabilidade entre os modelos estudados. Os parâmetros pertinentes à estrutura nativa e aos diferentes modelos se alteram em cor, iniciando com proteína nativa (em preto) e mutante “Ser_3” (em cinza escuro).

Figura 6 – Comparação entre os parâmetros calculados para o aminoácido Ser-105(um dos três amino ácidos membros de tríade catalítica de lípase) em mutantes, após de etapa adicional de minimização de energia por dinâmica molecular nos modelos obtidos na saída de Modeller, relativo aos valores obtidos para a estrutura nativa da enzima Lipase B de *Candida antarctica*. A figura está demonstrando para Ser-105 a variação dos valores dos parâmetros listados na tabela 3, encontrados em modelos a onde amino ácidos citados na tabela 1 foram substituídos por Val. Os descritores “acessibilidade ao solvente” e “Cross Presence Order” se apresentam como principais fatores introdutores de variabilidade entre os modelos estudados. Os parâmetros pertinentes a estrutura nativa e aos diferentes modelos se alteram em cor, iniciando com proteína nativa (em preto) e mutante “Ser_3” (em cinza escuro).

Figura 7 – Comparação entre os parâmetros calculados para o aminoácido Asn-187 (um dos três amino ácidos membros de tríade catalítica de lípase) em mutantes, após de etapa adicional de minimização de energia por dinâmica molecular nos modelos obtidos na saída

de Modeller, relativo aos valores obtidos para a estrutura nativa da enzima Lipase B de *Candida antarctica*. A figura está demonstrando para Asn-187 a variação dos valores dos parâmetros listados na tabela 3, encontrados em modelos a onde amino ácidos citados na tabela 1 foram substituídos por Val. O descritor “acessibilidade ao solvente” se apresenta

5 como principal fator introdutor de variabilidade entre os modelos estudados. Os parâmetros pertinentes à estrutura nativa e aos diferentes modelos se alteram em cor, iniciando com proteína nativa (em preto) e mutante “Ser_3” (em cinza escuro).

Figura 8 – Comparação entre os parâmetros calculados do aminoácido His-224 (um dos três amino ácidos membros de tríade catalítica de lípase) em mutantes, após de etapa

10 adicional de minimização de energia por dinâmica molecular nos modelos obtidos na saída de Modeller, relativo aos valores obtidos para a estrutura nativa da enzima Lipase B de *Candida antarctica*. A figura está demonstrando para His-224 a variação dos valores dos parâmetros listados na tabela 3, encontrados em modelos a onde amino ácidos citados na tabela 1 foram substituídos por Val. O descritor “acessibilidade ao solvente” se apresenta

15 como principal fator introdutor de variabilidade entre os modelos estudados. Os parâmetros pertinentes à estrutura nativa e aos diferentes modelos se alteram em cor, iniciando com proteína nativa (em preto) e mutante “Ser_3” (em cinza escuro).

Figura 9 – Valores do índice de hidrofobicidade de superfície, o SHI (“Surface Hydrophobicity Index”), para a proteína nativa (1tcb.pdb) e para os dez mutantes

20 modelados com o auxílio do software Modeller (representados em preto). DmSHI (representados em cinza) indica a variação do SHI após mutação. Todos os mutantes apresentam DmSHI positivo, ou seja, todas as mutações estudadas aumentaram a região de superfície composta por aminoácidos hidrofóbicas. Entre os mutantes, os modelos Comb1 e Comb2 (ambos com mutações múltiplas) se destacam entre os modelos que mais elevaram

25 o SHI.

Figura 10 – Indicação de posição tridimensional dos amino ácidos substituídos por Valina na superfície de proteína nativa (1TCB.pdb).

DESCRIÇÃO DETALHADA DA INVENÇÃO

A invenção refere-se à identificação de aminoácidos com propriedades específicas

em superfícies protéicas cuja mutação não altere propriedades físico-químicas e estruturais dos aminoácidos considerados catalíticos. É demonstrada a aplicabilidade da técnica com a enzima Lipase B de *Candida antarctica*, importante alternativa na produção de biodiesel sem a adição de catalizadores inorgânicos, trazendo outros benefícios ambientais, como economia de energia utilizada em relação ao método atual de produção com a reação de transesterificação, e também sem a necessidade de purificação de glicerol, um dos subprodutos do método atual.

A presente invenção diz respeito ainda a um método para previsão de mutantes que aumente o índice de hidrofobicidade da superfície de proteínas e as proteínas obtidas por esse método.

O método da presente invenção é caracterizado por compreender as seguintes etapas:

a) Selecionar a proteína de interesse para aumento de hidrofobicidade da sua superfície em bancos de dados de estruturas macromoleculares;

b) Calcular e armazenar as características físico-químicas da proteína em um software e banco de dados como o Blue Star STING (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., Mancini, A.L., 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. Genet Mol Res. 2006 Dec 1;5(4):717-22) e no Java Protein Dossier (Neshich, G., Rocchia, W., Mancini, A.L., Yamagishi, M.E., Kuser, P.R., Fileto, R., Baudet, C., Pinto, I.P., Montagner, A.J., Palandrani, J.F., Krauchenco, J.N., Torres, R.C., Souza, S., Togawa, R.C., Higa, R.H. 2004. JavaProtein Dossier: a novel web-based data visualization tool for comprehensive analysis of protein structure. Nucleic Acids Res. 2004 Jul 1;32 (Web Server issue):W595-601

c) Selecionar as características físico-químicas e estruturais da proteína (superfície e sítio catalítico) de acordo com interesse;

d) Definir valores limites (característicos para cada estrutura escolhida para análise e posterior aumento de hidrofobicidade na respectiva superfície) para cada uma das características selecionadas na etapa "c".

- e) Selecionar os aminoácidos cujas características satisfizeram os valores de cada um dos descritores selecionados na etapa “c” através de um módulo específico no software Blue Star STING;
 - f) Realizar uma modelagem por homologia com um software específico de mutantes com mutações singulares modificando cada um dos aminoácidos selecionados na etapa “e” para o resíduo de interesse;
 - g) Monitorar a variação das propriedades físico-químicas e/ou estruturais dos aminoácidos que compõe o sítio catalítico através da comparação dos descritores gerados pelo servidor do software
 - h) Construir modelos de proteínas com mutações múltiplas baseadas nos melhores valores de V_4 , desenvolvido para avaliar a discrepância entre os modelos gerados e a proteína nativa, para as mutações singulares estudadas, utilizando o software de modelagem.
 - i) Avaliar os modelos com mutações múltiplas pelo escore V_4 , comparando com a estrutura nativa.
 - j) Medir a variação da área de superfície hidrofóbica e do seu parâmetro SHI (Surface Hydrophobicity Index) em relação à estrutura nativa da enzima selecionada o que proporciona mais um avaliador de discrepância entre a proteína original e mutada, especificamente no sentido de atingir o maior indicador de hidrofobicidade na superfície .
- Os bancos de dados para selecionar as proteínas podem ser vários. Mais especificamente a presente invenção utiliza o banco de dado público chamado Protein Data Bank (PDB, Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. “The Protein Data Bank.” Nucleic Acids Res. 2000 Jan 1; 28(1):235-42).
- As características da proteína a serem selecionadas podem ser, mas não estão limitadas à, distância do sítio catalítico, densidade de contatos internos, hidrofobicidade, potencial eletrostático, conservação de posição sequencial de aminoácidos, *cross link order*, *cross presence order*, densidade, esponjicidade, fator temperatura e outros de uma lista de 711 parâmetros físico-químicos e estruturais armazenados na STING_DB, devidamente listados e descritos por Neshich G, Rocchia W, Mancini AL, Yamagishi MEB, Kuser PR,

Fileto R, Baudet C, Pinto I, Montagner AJ, Palandrani JF, Krauchenco JN, Torres RC, Souza S, Togawa RC e Higa RH “JavaProtein Dossier: a novel web-based data visualization tool for comprehensive analysis of protein structure.” Nucleic Acids Res. 2004 Jul 1;32(Web Server issue):W595-601.

5 A seleção dos aminoácidos pode ser feita através de um módulo específico do software utilizado, como por exemplo, mas não limitado ao módulo ¹PD do Blue Star STING.

A modelagem dos aminoácidos pode ser realizada por softwares de modelagem como por exemplo, mas não limitado aos softwares SwissModel (Arnold K., Bordoli L., Kopp J., and Schwede T. (2006). The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. Bioinformatics, 22,195-201) e Modeller (A. Sali & T.L. Blundell. “Comparative protein modelling by satisfaction of spatial restraints” J. Mol. Biol. 234, 779-815, 1993). Preferencialmente para a presente invenção foi utilizado o software Modeller.

15 Na presente invenção foi utilizado como exemplo a proteína lipase e, para este caso, os aminoácidos selecionados foram substituídos pelos resíduos de Alanina, uma vez que este aminoácido mantém as características estruturais da cadeia principal (formação de estrutura secundária) ou preferencialmente por uma Valina, que possui o maior índice de hidropatia (em inglês: “hydropathy index”) que a Alanina (ambas hidrofóbicas).

20 O monitoramento dos aminoácidos modelados pode ser realizado em softwares como, mas não limitado a, Blue Star STING e armazenados em arquivos com formato TGZ (formato exclusivo desenhado para otimizar a performance de STING_server e do software STING). A comparação realizada no software pode envolver vários descritores que podem ser mas não estão limitados a: Energia de contatos não utilizados; Contatos não utilizados; 25 Densidade; Área de acessibilidade ao solvente; Esponja; Cross Link Order) (grão de interconexão entre aminoácidos); Cross Presence Order (grão de co-localização entre os aminoácidos); Hidrofobicidade; Curvatura local (escolha pode ser entre os 711 parâmetros disponíveis em STING_DB).

30 Os mutantes são monitorados com base no escore V_4 desenvolvido e explicado no exemplo com a enzima Lipase B de *Candida antarctica*. Quanto menor o valor do escore,

menor a influência dos resíduos substituídos sobre os aminoácidos do sítio catalítico.

- No caminho alternativo há duas etapas adicionais opcionais: entre as etapas “f” e “g” e entre as etapas “h” e “i”, do método. Após a geração do modelo inicial pelo Modeller a estrutura gerada é refinada via minimização de energia por dinâmica molecular. No exemplo estudado mostramos como essas duas etapas de minimização de energia podem ser feitas com o pacote de software de dinâmica molecular Gromacs. No entanto outros algoritmos podem ser utilizados nessas etapas tais como Yasara (<http://www.yasara.org/yamber.htm>) ou NAMD (Scalable molecular dynamics with NAMD. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, Schulten K. *J Comput Chem.* 2005 Dec;26(16):1781-802.). Dessa forma, uma outra concretização da presente invenção é um método contendo as seguintes etapas:
- a) Selecionar a proteína de interesse para aumento de hidrofobicidade da sua superfície em bancos de dados de estruturas macromoleculares;
 - b) Calcular usando o software Blue Star STING e armazenar as características físico-químicas da proteína em banco de dados específico (MySQL);
 - c) Selecionar as características físico-químicas e estruturais da proteína (superfície e sítio catalítico) de acordo com interesse;
 - d) Definir valores limites para cada uma das características selecionadas na etapa “c”;
 - e) Selecionar os aminoácidos cujas características satisfizeram os valores de cada um dos descritores selecionados na etapa “c” através de um software específico ^JPD;
 - f) Realizar uma modelagem por homologia com um software específico (Modeller) com mutações singulares modificando cada um dos aminoácidos selecionados na etapa “e” para o resíduo de interesse;
 - g) Realizar a minimização de energia por dinâmica molecular dos modelos 3D gerados na etapa “f”;
 - h) Monitorar a variação das propriedades físico-químicas e/ou estruturais dos aminoácidos que compõe o sítio catalítico através da comparação dos descritores gerados pelo servidor do software específico;
 - i) Construir modelos de proteínas com mutações múltiplas baseadas nos melhores valores de V_4 (um indicador de abrangência geral porém, neste caso com aplicação específica para

os descritores escolhidos na etapa “c” acima) para as mutações singulares estudadas, utilizando o software de modelagem;

j) Realizar a minimização de energia por dinâmica molecular dos modelos 3D gerados na etapa “i”;

5 k) Avaliar os modelos com mutações múltiplas pelo escore V_4 , comparando com a estrutura nativa;

l) Medir a variação da área de superfície hidrofóbica e do seu parâmetro SHI (Surface Hydrophobicity Index) em relação à estrutura nativa da enzima selecionada

O método da presente invenção utiliza o software Blue Star STING (Neshich, G.,
 10 Togawa, R., Mancini, A. L., Kuser, P. R., Yamagishi, M. E. B., Pappas Jr., G., Torres, W. V., Campos, T. F., Ferreira, L. L., Luna, F. M., Oliveira, A. G., Miura, R. T., Inoue, M. K., Horita, L. G., de Souza, D. F., Dominiquini, F., Álvaro, A., Lima, C. S., Ogawa, F. O., Gomes, B. G., Palandrani, J. C. F., dos Santos, G. F., de Freitas, E. M., Mattiuz, A. R., Costa, I. C., de Almeida, C. L., Souza, S., Baudet, C. and Higa, R. H. 2003. STING
 15 Millennium: a Web based suite of programs for comprehensive and simultaneous analysis of protein structure and sequence. Nucleic Acids Research, 31:13, 3386-3392) que por sua vez calcula mais de 700 descritores físico-químicos e estruturais para cada aminoácido de determinada cadeia proteica depositada no repositório das estruturas macromoleculares (PDB) ou então, gera as informações correspondentes para os modelos estruturais e
 20 encapsula os mesmos em um arquivo de saída em formato .TGZ. A partir de filtros bem definidos, foram selecionados aminoácidos hidrofílicos, encontrados na superfície de proteína, para serem mutados *in silico* por aminoácidos hidrofóbicos. Um conjunto de características foi selecionado para monitorar como as mutações singulares afetam os descritores dos aminoácidos do sítio catalítico. Foi demonstrado que as mutações geraram
 25 mínimas variações dessas características inclusive nos modelos contendo mutações múltiplas, resultando dessa forma em modelos com área de superfície constituída por um número maior de aminoácidos hidrofóbicos, aumentando assim a solubilidade em meios hidrofóbicos, como por exemplo, em meios de óleos vegetais.

A presente invenção diz respeito também às proteínas mutantes obtidas através do
 30 método de predição de mutantes descrito no presente relatório.

A invenção será agora descrita em maiores detalhes por meio dos exemplos a seguir, os quais não devem ser interpretados como limitadores do escopo da invenção.

EXEMPLO

Exemplo 1 – Identificação da proteína e seleção dos aminoácidos

5 A primeira etapa no desenvolvimento de possíveis mutantes da proteína Lipase é a escolha de uma estrutura tridimensional base (template). Através do PDB, podemos escolher a enzima Lipase B de *Candida antarctica* depositada sob o código 1TCB.pdb (figura 1).

10 Na primeira etapa são escolhidos aminoácidos que se encontram na superfície protéica, e que, portanto, promovem interações com o solvente. Como o objetivo é desenvolver uma proteína com maior solubilidade em ambientes hidrofóbicos, aminoácidos polares serão alvos de possíveis mutações na superfície protéica. Outra opção utilizada foi a substituição de aminoácidos apolares de baixa valor na escala de hidropatia por resíduos com maiores valores na mesma escala. Embora não seja sempre válida, a hipótese que
15 resíduos distantes estruturalmente do sítio catalítico não interferem em nano-ambiente deste mesmo sítio catalítico será utilizada.

Exemplo 2 – Modelagem

Após a seleção dos aminoácidos alvos, foi utilizado o software Modeller (A. Sali & T.L. Blundell. “Comparative protein modelling by satisfaction of spatial restraints” J. Mol.
20 Biol. 234, 779-815, 1993) para a geração de modelos mutantes em relação a estrutura base pelo método de modelagem por homologia. O software Modeller realiza a geração de modelos utilizando a estrutura template como geradora de restrições espaciais aplicadas a estrutura a ser modelada.

O arquivo de entrada consiste em alinhamento entre a sequência de aminoácidos do
25 template (presente no PDB) com a sequência de aminoácidos que deseja modelar, como ilustrado no quadro 1 abaixo. Os amino ácidos sublinhados representam aqueles resíduos que substituíram os originais de template. Neste caso trata se de mutante múltiplo que nomeamos: “combo1” e que contem todos os resíduos listados na tabela 1.

Tabela 1 - Aminoácidos de superfície proteica, polares e pouco hidrofóbicos, selecionados como alvos para mutação com base em sua distância do sítio catalítico da enzima Lipase B de *Candida antarctica* (1TCB.pdb) e no mesmo tempo contendo nenhum contato interno com outros aminoácidos. Os aminoácidos listados em destaque (fundo cinza) são levemente hidrofóbicos e por isto são considerados como candidatos apropriados para substituição por Valina que é mais hidrofóbica.

Ser-3	Gly-4	Ser-5	Gly-24
Thr-244	Ala-276	Arg-309	Thr-316

Quadro 1 - Alinhamento entre a sequência de aminoácidos do template (presente no PDB) com a sequência de aminoácidos que deseja modelar

```
>P1;1tc
structureX:1tc:1:A:317:A:::
LPSGSDPAFSQPKSVLDAGLTCQGASPSVSKPILLVPGTGTTGPQSFDNWIPLSTQLGYTPCWISPPPFMLNDTQVNT EYMVNAITALY
AGSGNNKLPVLTWSQGGLVAQWGLTFPFSIRSKVDRLMAFAPDYKGTVLAGPLDALAVSAPSVWQQTGTSALTALRNAGGLTQIVPT
TNLYSATDEIVQPQVSNPLDSSYLFNGKNVQAQAVCGPLFVIDHAGSLTSQFSYVVGSRALRSTTGQARSADYGITDCNPLPANDLTPE
QKVAAAALLAPAAAIVAGPKQNCPELMPYARPFVAVGKRTCSGIVTP*

>P1;lip1
sequence:lip1:::0.00:0.00
LPVVYDPAFSQPKSVLDAGLTCQVASPSVSKPILLVPGTGTTGPQSFDNWIPLSTQLGYTPCWISPPPFMLNDTQVNT EYMVNAITAL
YAGSGNNKLPVLTWSQGGLVAQWGLTFPFSIRSKVDRLMAFAPDYKGTVLAGPLDALAVSAPSVWQQTGTSALTALRNAGGLTQIVP
TTNLYSATDEIVQPQVSNPLDSSYLFNGKNVQAQAVCGPLFVIDHAGSLTSQFSYVVGSRALRSYTGQARSADYGITDCNPLPANDLTPE
EQKVAAAYLLAPAAAIVAGPKQNCPELMPYARPFVAVGYTCSGIVYP*
```

10

ALINHAMENTO MODELLER

O processo foi repetido 8 vezes substituindo os aminoácidos da tabela 1 com o resíduo Valina, uma vez que esse aminoácido possui elevado valor na escala de hidropatia (tabela 2) além de ter tamanho menor quando comparado com os aminoácidos Leucina e Isoleucina.

15

Tabela 2 - Escala de hidropatia de acordo com: “Radzicka, A. & Wolfenden, R. (1988). Comparing the polarities of the amino-acids -- side-chain distribution coefficients between the vapor-phase, cyclohexane, 1-octanol, and neutral aqueous-solution.

Biochemistry 27, 1664-1670.” Amino ácidos em destaque (fundo cinza) são hidrofóbicos.

Hidropatia	Aminoácido	Hidropatia	Aminoácido
-14.92	(Arg)	0.94	(Gly)
-8.72	(Asp)	1.28	(Cys)
-6.81	(Glu)	1.81	(Ala)
-6.64	(Asn)	2.33	(Trp)
-5.55	(Lys)	2.35	(Met)
-5.54	(Gln)	2.98	(Phe)
-4.66	(His)	3.50	(Pro)
-3.40	(Ser)	4.04	(Val)
-2.57	(Thr)	4.92	(Ile)
-0.14	(Tyr)	4.92	(Leu)

Exemplo 3 – Minimização de energia

- 5 Devido ao fato de modelos gerados pelo Modeller da forma descrita não passar por nenhum refinamento da estrutura, foi utilizado o algoritmo de minimização de energia do pacote Gromacs para dinâmica molecular (Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ 2005 "GROMACS: fast, flexible, and free". J Comput Chem **26** (16): 1701–18), gerando dessa forma estruturas que passaram por etapas de relaxamento
- 10 das ligações entre os átomos da estrutura tridimensional da proteína, obtendo assim uma estrutura similar a gerada pelo software Modeller, porém com energia global menor.

O campo de força utilizado foi o Gromos96 (van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberger, P. H.; Krüger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. “Biomolecular Simulation: The GROMOS96 Manual and User Guide” vdf

Hochschulverlag AG an der ETH Zürich and BIOMOS b.v.: Zürich, Groningen, 1996.), com parâmetros existentes para aminoácidos e ácidos nucleicos, além de íons como cálcio, cloro e também para moléculas de água, uma vez que o protocolo utilizado utiliza solvente explícito. O quadro 2 abaixo resume a sequência de comandos utilizados na geração de

5 modelos com essa etapa adicional de refinamento:

Quadro 2 - Sequência de comandos utilizados na geração de modelos utilizando a etapa de minimização de energia

```

pdb2gmx -ff G53a6 -f lip1.pdb -o lip1.gro -p lip1.top -ignh
grompp -v -f minim.mdp -c lip1.gro -p lip1.top -o lip1-EM-vacuum.tpr
mdrun -v -deffnm lip1-EM-vacuum -c lip1-EM-vacuum.pdb

```

MINIMIZAÇÃO GROMACS

10 A primeira linha de comando utiliza o arquivo em formato “pdb” como a entrada e fazendo a conversão para o arquivo de coordenadas do formato gromacs (.gro) e de topologia (.top), utilizando o campo de forças (que contém parâmetros para todos os átomos de sistema) G53a6, que é código para o campo Gromos96. Na segunda linha os arquivos de topologia e coordenadas são unidos gerando o arquivo de saída no formato

15 “tpr” com parâmetros definidos pelo arquivo minim.mdp, que contém informação sobre qual biblioteca e algoritmos devem ser utilizados, além de valores limites sobre até que valor de energia a minimização deve ser rodada. Na última linha de comando todos os arquivos gerados sob o nome lip1-EM-vacuum são utilizados na rodada de dinâmica molecular tendo como arquivo de saída um arquivo no formato “pdb” com as coordenadas

20 dos átomos do sistema com energia global minimizada.

Seguindo esse procedimento foram gerados outros 8 modelos que serão comparados com os 8 modelos iniciais obtidos pelo Modeller, no âmbito de seus valores do escore V_4 . Para que a comparação com a estrutura nativa seja mais próxima do ideal, foi utilizado o mesmo procedimento de minimização de energia para a proteína 1TCB.pdb.

25

Exemplo 4 – Análise dos modelos gerados

Os 17 modelos gerados e a enzima Lipase B de *Candida antarctica* foram analisados pelo software Blue Star STING, que armazena em seu banco de dados (STING_DB) mais de 700 descritores físico-químicos, estruturais e evolutivos. Os valores de cada um dos descritores para a enzima Lipase B de *Candida antarctica* estão pré calculados e armazenados no STING_DB, enquanto que os descritores para os modelos gerados pelo Modeller podem ser calculados pelo servidor do Blue Star STING, gerando arquivos em formato TGZ. Os descritores utilizados na análise comparativa estão listados na tabela 3.

Tabela 3 – Descritores físico-químico e estruturais usados na análise comparativa entre a estrutura nativa da enzima Lipase B de *Candida antarctica* (1TCB.pdb) e os modelos mutantes gerados por homologia.

1. Energia de contatos não utilizados;	2. Contatos não utilizados;	3. Densidade;
4. Área de acessibilidade ao solvente;	5. Esponja;	6. Cross Link Order;
7. Cross Presence Order;	8. Hidrofobicidade;	9. Curvatura local;

Como o objetivo é criar mutantes que mantenham a atividade da enzima nativa, comparamos os valores das características e descritores presentes no Blue Star STING para os três aminoácidos do sítio catalítico (Uppenberg J, Ohrner N, Norin M, Hult K, Kleywegt GJ, Patkar S, Waagen V, Anthonsen T, Jones TA. Crystallographic and molecular-modeling studies of lipase B from *Candida antarctica* reveal a stereospecificity pocket for secondary alcohols. Biochemistry. 1995 Dec 26;34(51):16838-51). Pequenas mudanças estruturais devem mudar ligeiramente as características do sítio catalítico, levando a uma possível (neste caso inclusive, a desejável) manutenção da atividade catalítica.

Após monitorar as variações e garantir que as mutações singulares influenciavam minimamente nas características físico-químicas e estruturais da tríade catalítica, foram geradas estruturas mutantes com mutações combinadas em um total de 4 modelos dos quais dois foram criados utilizando o Modeller e dois com uma sequencia de Modeller e Minimização de Energia executada pelo Gromacs, e finalmente seguida pela análise comparativa com a estrutura nativa da Lipase B de *Candida antarctica*, uma vez que mutações combinadas devem influenciar mais a solubilidade em meio hidrofóbico da

enzima em questão do que as mutações simples.

Com a geração dos modelos tridimensionais dos mutantes da proteína Lipase, foram atingidas as condições para cálculo dos descritores para os modelos o que foi realizado e os arquivos em formato “TGZ” do software Blue Star STING foram gerados para cada um dos
5 modelos contendo as mutações descritas acima.

O valor de “Energia de contato não utilizados” e “Contatos não utilizados” apresentaram valores da ordem do dobro dos valores observados em estrutura nativa, sendo consequência de um dos artefatos da modelagem por homologia com o Modeller. O mesmo foi observado para a estrutura nativa após o processo de minimização de energia por
10 dinâmica molecular. Dessa forma o parâmetro “Energia de contato não utilizados” e “Contatos não utilizados” não foram utilizados na análise comparativa entre os modelos gerados.

As figuras 3 a 5 mostram os resultados de cada um dos parâmetros da tabela 3 (com exceção dos parâmetros “Energia de contato não utilizados” e “Contatos não utilizados”) de
15 forma comparativa com os valores apresentados pela estrutura nativa 1TCB.pdb, para os três aminoácidos do sítio catalítico: Ser-105, Asp-187 e His-224, respectivamente.

Para quantificar quais mutantes perturbaram menos os aminoácidos do sítio catalítico foi desenvolvido o “escore de variação” que mede a variação dos diversos parâmetros em relação aos da estrutura nativa. Para cada parâmetro presente nos gráficos
20 das figuras 3 a 5 foi calculado o módulo da variação dos valores em relação ao da estrutura 1TCB.pdb:

$$V_j = \sum_{i=1}^7 |P_{1TCB,i,j} - P_{Mut,i,j}| \quad (1)$$

onde V_j é o escore de variação para o j-ésimo aminoácido do sítio catalítico (j=1 para Ser-105, j=2 para Asn-187 e j=3 para His-224), $P_{1TCB,i,j}$ é o valor calculado para o i-ésimo
25 parâmetro da estrutura 1TCB.pdb (i=1 para Acessibilidade, i=2 para Hidrofobicidade, i=3 para Curvatura, i=4 para Densidade com sonda esférica centrada no Carbono-alfa, i=5 para Fator Esponja com sonda esférica centrada no Carbono-alfa, i=6 para Cross Link Order centrado no Carbono-alfa, i=7 para Cross Link Order centrado no Carbono-beta, i=8 para Cross Link Order centrado no LHA, i=9 para Cross Presence Order centrado no Carbono-

alfa, $i=7$ para Cross Presence Order centrado no Carbono-beta, $i=8$ para Cross Presence Order centrado no LHA, $i=9$) do j -ésimo aminoácido, da mesma forma $P_{Mul,i,j}$ é o valor calculado para o i -ésimo parâmetro do j -ésimo resíduo de cada mutante construído. Dessa forma V_1 representa a variação dos parâmetros do resíduo Ser-105 de cada mutante em
 5 relação aos parâmetros do resíduo Ser-105 da proteína nativa 1TCB.pdb. V_2 e V_3 representam o mesmo escore de variação para os resíduos Asn-187 e His-224, respectivamente.

Em seguida, foram somados para cada mutante os valores calculados pela equação (1), originando o escore V_4 que mede a variação de todos os parâmetros de todos os
 10 aminoácidos do sítio catalítico em relação a proteína nativa:

$$V_4 = \sum_{i=1}^3 V_i \quad (2)$$

Para os modelos gerados com a etapa adicional de minimização de energia por dinâmica molecular foi realizada a mesma análise descrita acima, porém com os parâmetros
 15 “Energia de contato não utilizados” e “Contatos não utilizados” sendo levados em consideração, uma vez que o template 1TCB.pdb (também sujeito a etapas de minimização de energia) possui valores próximos aos dos outros mutantes construídos. Os gráficos das figuras 3 a 5 apresentam os parâmetros calculados para os aminoácidos do sítio catalítico, Ser-105, Asp-187 e His-224, respectivamente, de todos os mutantes obtidos.

20 Os escores de variação foram calculados, gerando uma lista de melhores mutações, no sentido de menor influência nos parâmetros da tríade catalítica indicado por menores valores de V_4 , mostrados na tabela 4. A tabela 4 demonstra que a variação dos parâmetros monitorados para todos os mutantes foi mínima e por isto, todas as substituições poderiam ser utilizadas para geração do mutante múltiplo.

25 Esse procedimento foi repetido para as mesmas mutações, mas agora para os modelos gerados com a etapa adicional de minimização de energia por dinâmica molecular. Os gráficos das figuras 6 a 8 indicam as variações dos parâmetros desses modelos, e a tabela 5, os valores calculados para os escores de variação, indicando os melhores

candidatos para a construção de mutantes com múltiplas mutações.

Tabela 4 – Valores de escore de variação V_4 para cada mutante. Os valores menores indicam menor influência dos resíduos substituídos sobre os aminoácidos do sítio catalítico.

O V_4 é definido como $V_4 = \sum_{i=1}^3 V_j$ a onde $V_j = \sum_{i=1}^7 |P_{ITCB,i,j} - P_{Mut,i,j}|$

	1tc6	3ser	4gly	5ser	24gly	244thr	276ala	609arg	326thr	Combo 1	Combo 2
V1	0	3,3	2,01	2,39	5,83	3,5	0,69	3,32	5,24	2,22	5,69
V2	0	3,13	2,72	1,12	0,64	0,37	1,11	0,78	0,37	0,54	0,38
V3	0	0,7	1,99	1,45	1,8	0,5	1,46	1,64	0,19	1,96	2,29
V4	0	7,13	6,72	4,96	8,27	4,37	3,26	5,74	5,8	4,72	8,36

5 Como no caso anterior, verificou-se que as variações em parâmetros monitorados dos todos os mutantes foram mínimas.

Para o calculo dos escores de variação dessas amostras, levamos em consideração os parâmetros “Energia de contato não utilizados” e “Contatos não utilizados”, que devido a ordem de grandeza do próprio parâmetro resultou em valores de V_4 bem acima dos cálculos

10 feitos anteriormente para o procedimento sem etapas de minimização de energia.

No total, com base nos resultados apresentados nas tabelas 4 e 5, geramos 2 mutantes com mutações múltiplas - o combo1, composto de todas as mutações estudadas (lista dos amino ácidos substituídos esta demonstrada na tabela 1) e combo2, composto pelas mesmas mutações acrescido de outras duas mutações de aminoácidos: Ala_273 e

15 Asn_292, ambos com apenas dois contatos internos do tipo hidrofóbico. Estes dois amino ácidos foram selecionados após procurar qualquer amino acido na superfície proteica, hidrofóbico ou hidrofílico, que possui um índice de hidrofobicidade perto de linha de separação entre as categorias hidrofóbica e hidrofílica, e também tem o mínimo de contatos internos (tanto no sentido numérico quanto no sentido de valor energético). Este último

20 mutante múltiplo foi concebido na base de demanda de aumentar ainda mais o SHI das lípases, substituindo resíduos de Ala_273 e Asn_292 com Val.

Tabela 5 – Valores de escore de variação V_4 para cada mutante para os modelos gerados com o software Modeller e com a etapa adicional de minimização de energia por dinâmica molecular, onde menores valores indicam menor influência dos resíduos

substituídos sobre os aminoácidos do sítio catalítico. O V_4 é definido como $V_4 = \sum_{i=1}^3 V_j$ a

onde
$$V_j = \sum_{i=1}^7 |P_{1TCB,i,j} - P_{Mut,i,j}|$$

	1tcB	3ser-MD	4gly-MD	5ser-MD	24gly-MD	24th-MD	276ala-MD	309arg-MD	326th-MD	Combo 1-MD	Combo 2-MD
V1	0	1,16	1,39	2,86	2,7	0,41	0,59	1,34	1,2	3,41	9,05
V2	0	2,3	7,32	0,44	5,62	5,42	4,51	0,22	6,56	2,47	0,96
V3	0	3,88	1,4	18,99	21,01	7,47	3,53	9,14	9,78	1,63	8,68
V4	0	7,34	10,11	22,29	29,33	13,3	8,63	10,7	17,54	7,51	18,69

Utilizando o escore V_4 para quantificar as variações que as mutações causaram nos parâmetros da tríade catalítica, foi aferido que o mutante “Combo 1” apresenta o menor valor de V_4 porém, o combo 2 possui o maior valor de DmSHI. Entre estes dois modelos deveria ser encontrado aquele que tem as características ideais tanto para solubilidade em meio hidrofóbico, quanto para ter a sua funcionalidade minimamente alterada devido as descritas substituições.

Ao final, é usado um índice desenvolvido pelo grupo, o SHI (“Surface Hydrophobicity Index”) que mede a hidrofobicidade de superfície de uma dada proteína em isolamento ou em complexo (Neshich IAP, De-Moraes FR, Salim JA, Mazoni I, Mancini A, Jardine JG, Neshich G: Surface Hydrophobicity Index (SHI): Insight Into The Mechanisms Of Protein-Protein Associations. X-Meeting Eletronic Abstracts Book 2009). O SHI é definido como uma associação entre a informação contida na área dos resíduos e a hidrofobicidade do aminoácido baseada em escalas conhecidas. A área acessível ao solvente é calculada usando o programa SurfV (Sridharan, S., Nicholls, A. and Honig, B. (1992) A new vertex algorithm to calculate solvent accessible surface areas. Biophys. J., 61, A174.) em termos da área de superfície de Van der Waals. Scripts locais em Perl executam o programa SurfV para as cadeias de lipases “em isolamento”, onde cada dada cadeia é considerada uma molécula solitária e sua área acessível à superfície é calculada. Os arquivos de saída são editados e formatados por scripts em Perl e inseridos em bancos de dados MySQL (usando MySQL server 5.0.67 (<http://www.mysql.com/>)).

Para definir o valor de SHI, é essencial definir o valor de ANHI (“Aminoacid Normalized Hydrophobicity Index”), que é uma medida da hidrofobicidade do aminoácido (extraída das escalas de hidropatia convencionais) normalizada pela área relativa de exposição de um dado resíduo “i” de uma cadeia de um arquivo PDB, de acordo com a fórmula a seguir:

$$ANHI_i = \frac{\text{Acessibilidade de "i"}}{\text{Acessibilidade máxima de "i"}} \times \text{Índice de hidrofobicidade de "i"}$$

A Acessibilidade de “i” representa o valor de área acessível ao solvente do resíduo “i” calculada pelo SurfV, Acessibilidade máxima de “i” representa o valor da acessibilidade máxima para o aminoácido “i”, extraídos dos valores de Área de Superfície para aminoácidos isolados do STING (Neshich G, Rocchia W, Mancini AL, Yamagishi ME, Kuser PR, Fileto R, Baudet C, Pinto IP, Montagner AJ, Palandrani JF, Krauchenco JN, Torres RC, Souza S, Togawa RC, Higa RH. 2004. JavaProtein Dossier: a novel web-based data visualization tool for comprehensive analysis of protein structure. Nucleic Acids Res. 2004 Jul 1;32(Web Server issue):W595-601). O valor de Índice de hidrofobicidade de “i” é extraído de escalas convencionais, sendo que no presente trabalho usou-se a escala Radicka (Radzicka, A. & Wolfenden, R. (1988). Comparing the polarities of the amino-acids -- side-chain distribution coefficients between the vapor-phase, cyclohexane, 1-octanol, and neutral aqueous-solution. Biochemistry 27, 1664-1670).

Finalmente, o SHI da cadeia isolada é calculado como a soma de todos os ANHIs de resíduos hidrofóbicos (HB) da cadeia em questão dividido como a soma de todos os ANHIs de resíduos hidrofílicos (HL) desta mesma cadeia, como mostrado a seguir:

$$SHI_{\text{cadeia isolada}} = \frac{\sum ANHI_{HB \text{ cadeia}}}{\sum ANHI_{HL \text{ cadeia}}}$$

Sendo assim, este índice reflete uma medida da hidrofobicidade de superfície, sendo esta maior, maior será o SHI. Se a superfície contiver maior quantidade de aminoácidos hidrofílicos, menor será o SHI. Foi então calculado o SHI, para todos os mutantes e estes foram comparados com o SHI da Lipase B usada como template (1TCB). Para compará-

los, foi definido o “DmSHI” (diferença dos mutantes em relação ao SHI da proteína nativa), sendo o valor do SHI da 1tcb.pdb menos o SHI do mutante. Na figura 9 são apresentados os valores referentes aos mutantes simples e também os valores encontrados para os dois mutantes múltiplos. Na figura 10 ilustramos as mutações escolhidos na estrutura tridimensional da proteína, assim como a região do sítio catalítico.

5

REIVINDICAÇÕES

1. Método para previsão de mutantes que aumente o índice de hidrofobicidade da
5 superfície de proteínas caracterizado por compreender as seguintes etapas:
 - a) Selecionar a proteína de interesse para aumento de hidrofobicidade da sua superfície em bancos de dados de estruturas macromoleculares;
 - b) Calcular as características da proteína usando um software específico e armazenar as características físico-químicas da proteína em banco de dados específico (MySQL);
 - 10 c) Selecionar as características físico-químicas e estruturais da proteína (superfície e sítio catalítico) de acordo com interesse;
 - d) Definir valores limites para cada uma das características selecionadas na etapa "c";
 - e) Selecionar os aminoácidos cujas características satisfizeram os valores de cada um dos descritores selecionados na etapa "c" através de um software específico;
 - 15 f) Realizar uma modelagem por homologia com um software específico com mutações singulares modificando cada um dos aminoácidos selecionados na etapa "e" para o resíduo de interesse;
 - g) Realizar opcionalmente a minimização de energia por dinâmica molecular dos modelos 3D gerados na etapa "f";
 - 20 h) Monitorar a variação das propriedades físico-químicas e/ou estruturais dos aminoácidos que compõe o sítio catalítico através da comparação dos descritores gerados pelo servidor do software específico;
 - i) Construir modelos de proteínas com mutações múltiplas baseadas nos melhores valores de V_4 para as mutações singulares estudadas, utilizando o software de modelagem;
 - 25 j) Realizar opcionalmente a minimização de energia por dinâmica molecular dos modelos 3D gerados na etapa "i";
 - k) Avaliar os modelos com mutações múltiplas pelo escore V_4 , comparando com a estrutura nativa;
 - l) Medir a variação da área de superfície hidrofóbica e do seu parâmetro SHI (Surface
30 Hydrophobicity Index) em relação à estrutura nativa da enzima selecionada

2. Método de acordo com a reivindicação 1 caracterizado pelo fato do banco de dados para selecionar a proteína ser o Protein Data Bank-PDB
3. Método de acordo com a reivindicação 1 caracterizado pelo fato do software utilizado para calcular e armazenar as características físico-químicas da proteína selecionada ser o
5 Blue Star STING.
4. Método de acordo com a reivindicação 1 caracterizado pelo fato do módulo específico do software utilizado para selecionar os aminoácidos cujas características satisfizeram os valores de cada um dos descritores ser o módulo ¹PD do software Blue Star STING.
5. Método de acordo com a reivindicação 1 caracterizado pelo fato do software utilizado
10 para modelagem por homologia ser um similar as funcionalidades de Modeller.
6. Método de acordo com a reivindicação 1 caracterizado pelo fato do software utilizado para monitorar a variação das propriedades físico-químicas e/ou estruturais dos aminoácidos ser o Blue Star STING.
7. Método de acordo com a reivindicação 5 caracterizado pelo fato dos descritores dos
15 aminoácidos estarem calculados e armazenados em arquivos com formato STING_TGZ.
8. Método de acordo com a reivindicação 1 caracterizado pelo fato do pacote de dinâmica molecular ser um software com funcionalidades similares ao Gromacs.
9. Proteínas mutantes com o índice de hidrofobicidade aumentado na superfície caracterizadas por serem obtidas através do método de qualquer uma das reivindicações de
20 1 a 8.

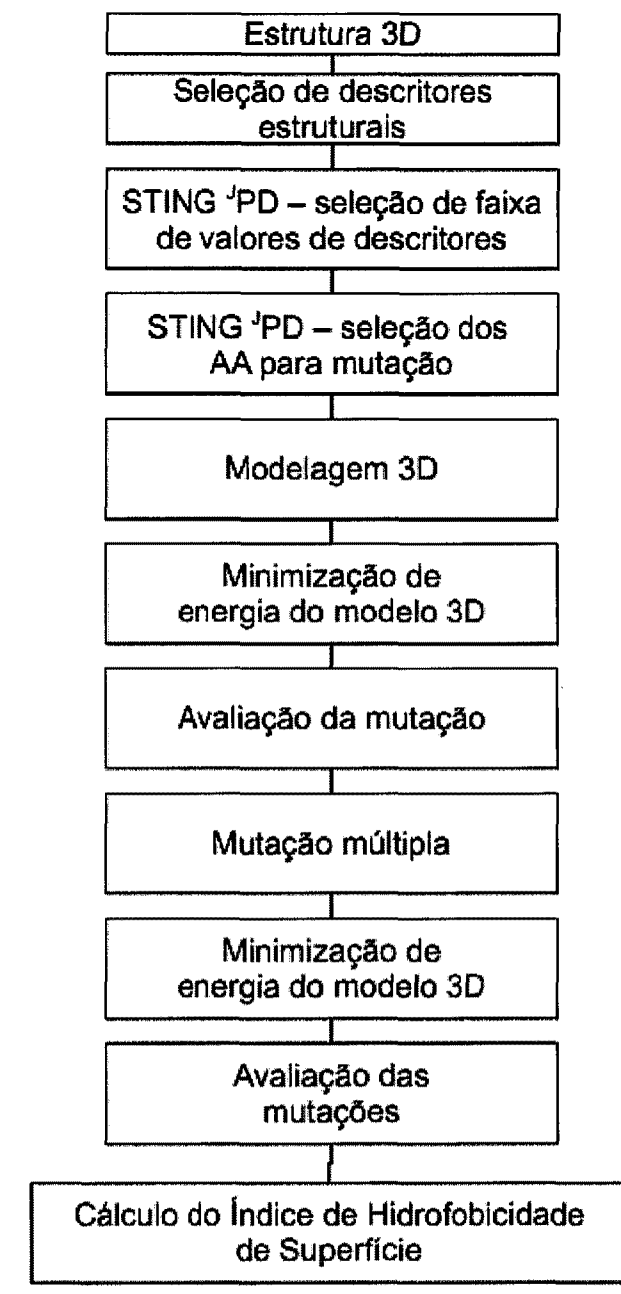


FIGURA 1

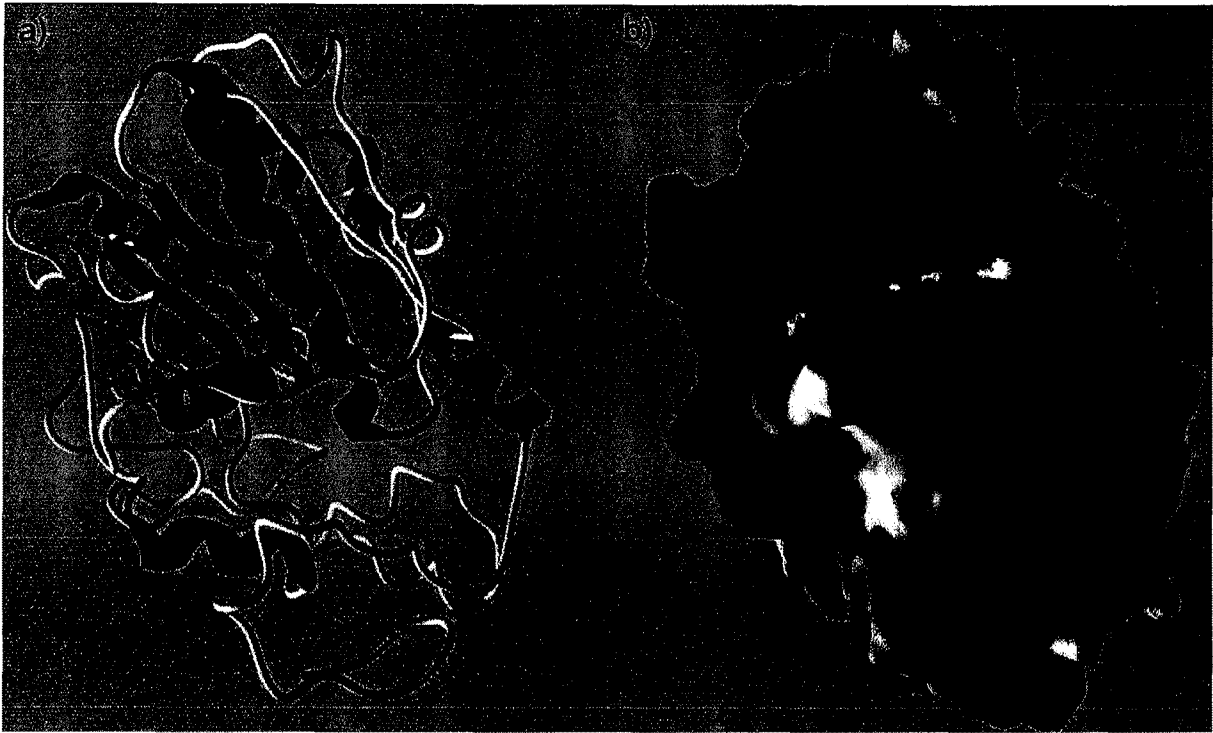


FIGURA 2

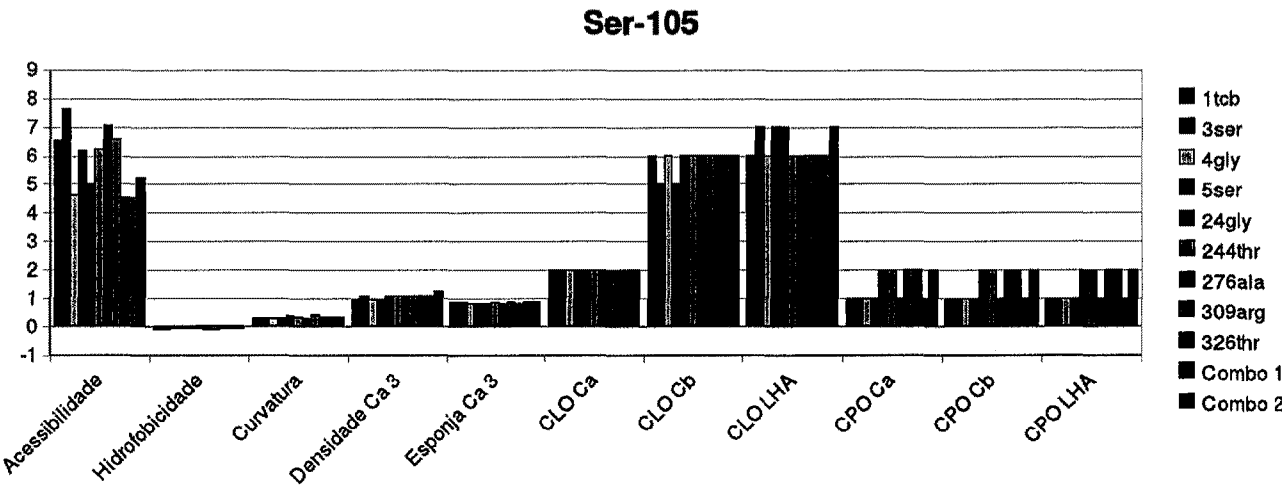
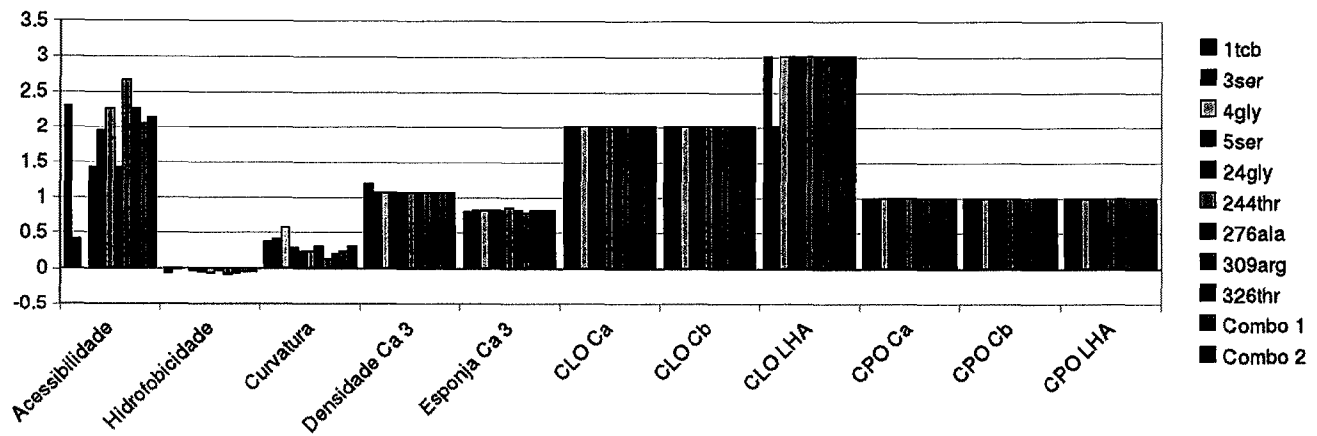
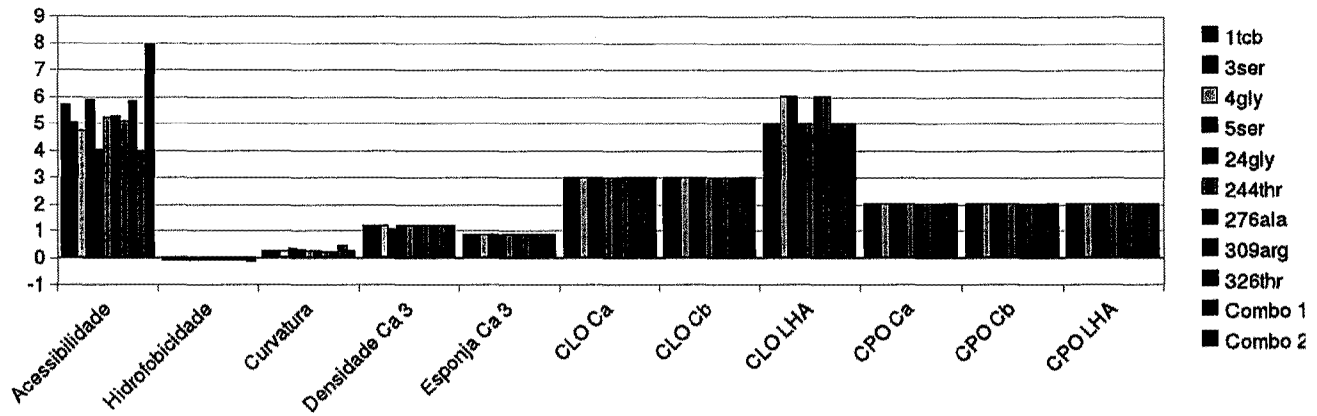
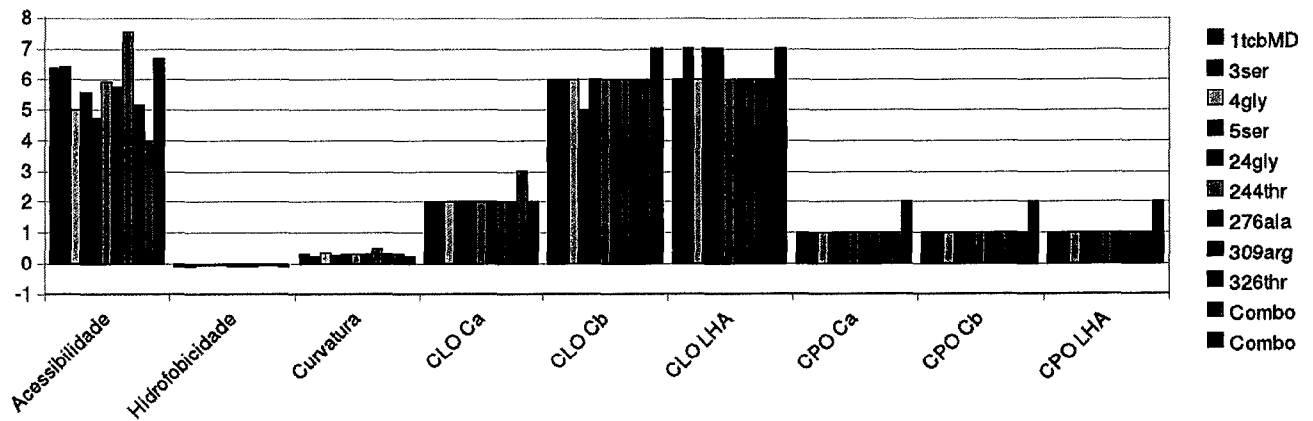


FIGURA 3

Asp-187**FIGURA 4****His-224****FIGURA 5****Ser-105-MD****FIGURA 6**

Asp-187-MD

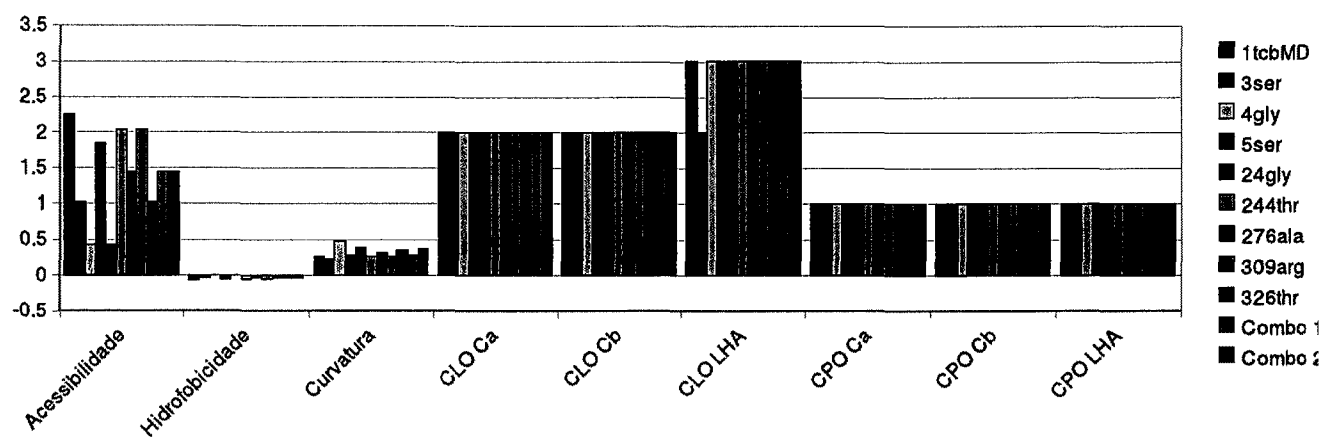


FIGURA 7

His-224-MD

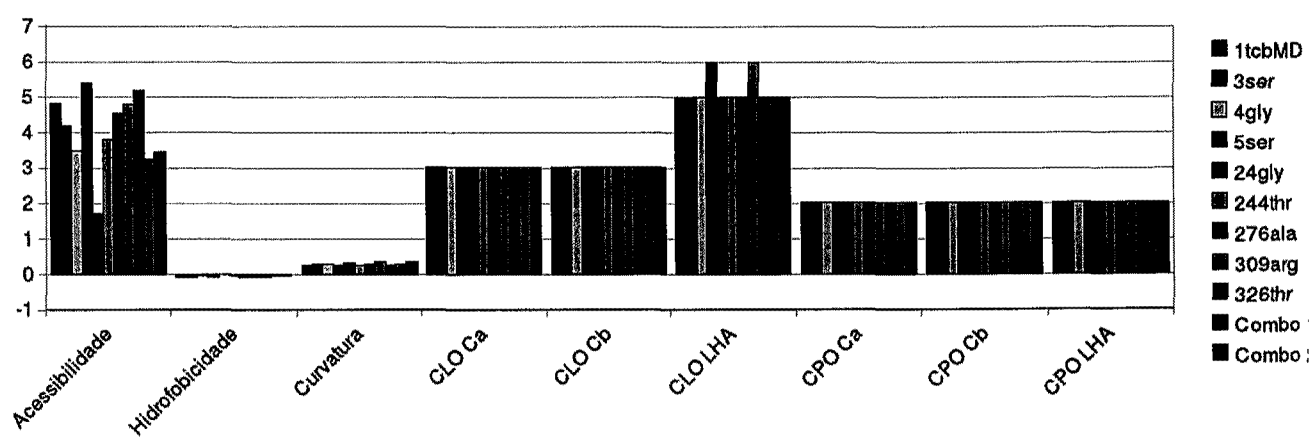


FIGURA 8

Índice de Hidrofobicidade da Superfície

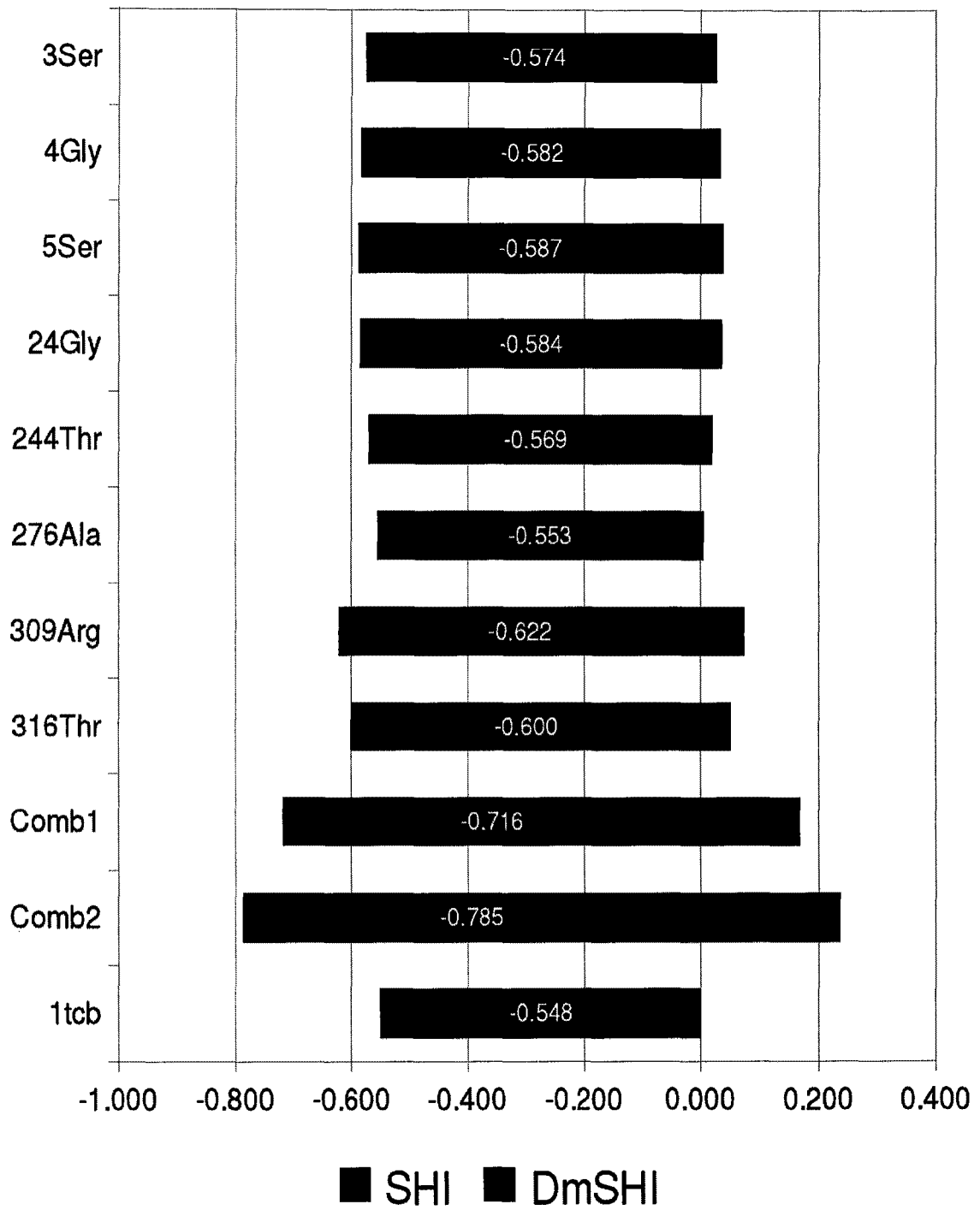


FIGURA 9

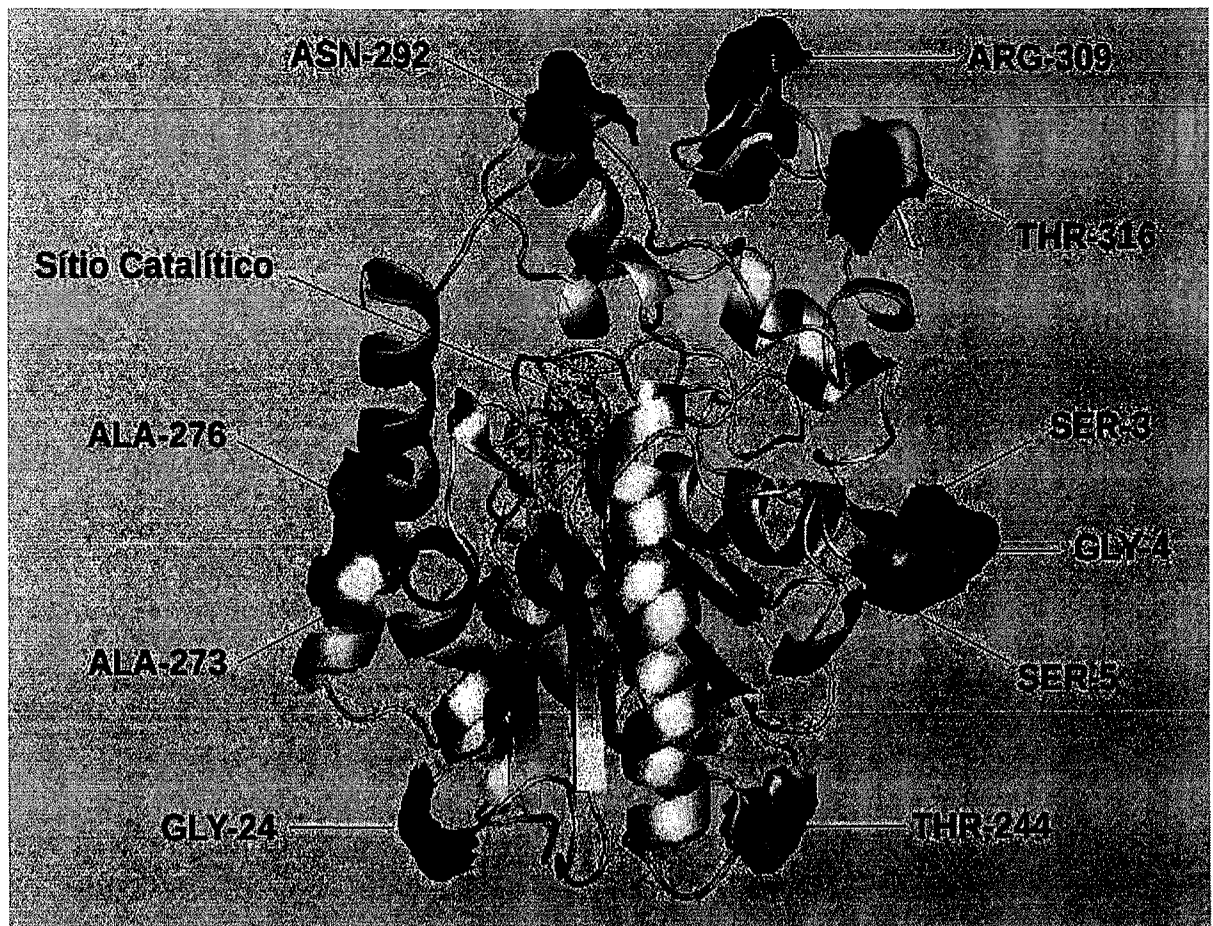


FIGURA 10

MÉTODO PARA PREVISÃO DE MUTANTES QUE AUMENTEM O ÍNDICE DE HIDROFOBICIDADE DA SUPERFÍCIE DE PROTEÍNAS

RESUMO

A presente invenção refere-se a um método de engenharia de proteínas, no qual são
5 identificadas regiões da superfície de enzimas que seriam passíveis de mutações que
influenciem de maneira reduzida as propriedades físico-químicas e estruturais dos
aminoácidos do sítio catalítico, com o objetivo de criar mutantes que apresentem a
superfície macromolecular mais hidrofóbica. Esta tecnologia pode ser aplicada para o
desenho racional de enzimas com maior solubilidade em meios hidrofóbicos como os óleos
10 vegetais.